Computational Investigation of the SARS-CoV-2 Cysteine Proteases

Ho Ting Henry Chan

New College



University of Oxford

A thesis presented for the degree of

Doctor of Philosophy

Trinity 2023

Abstract

The nucleophilic cysteine proteases of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that causes the COVID-19 pandemic include the main protease (M^{pro}) and the papain-like protease (PL^{pro}). Both enzymes are essential in the viral lifecycle by catalysing hydrolytic cleavages in polyproteins, making them promising drug targets.

This Thesis describes the use of computational modelling to gain insights into how M^{pro} and PL^{pro} bind their substrates, which can greatly aid drug discovery. In **Chapter 3,** *in silico* models of M^{pro}-substrate complexes and molecular dynamics (MD) simulations were employed to characterise key interactions. These interactions, also found in small molecule inhibitor-enzyme complexes, provide a basis for the discovery of peptides that competitively inhibit M^{pro}. In **Chapter 4,** dynamical-nonequilibrium MD (D-NEMD) was employed to investigate how M^{pro} adapts to substrate binding, revealing communication from the active site to the dimer interface and to an allosteric inhibition site. Sites susceptible to drug-resistant mutations were identified, particularly in the flexible Val186-Gly195 loop. **Chapter 5** discusses how the binding of each substrate residue induces structural responses in M^{pro}, where the conserved P1 glutamine triggers conformational changes that activate the M^{pro} active site, illustrating the substrate specificity of M^{pro}.

In **Chapter 6,** insights into the contrasting activity of PL^{pro} towards its polyprotein substrates were obtained through modelling of PL^{pro}-substrate complexes. Despite the apparent lack of specificity, a P2′ proline was found to enhance catalysis. The study also revealed the crucial dynamic role of the Gly266-Gly271 loop in PL^{pro}-substrate complexes representative of its deubiquitinating functions (**Chapter 7**).

The computational studies described here offer an understanding of the activity mechanisms of both cysteine proteases, involving dynamic loop movements to enable catalysis. The insights

gained on the relevance of protein dynamics to substrate specificity, promiscuity, and drug resistance will facilitate the development of drugs for future coronavirus variants.

Declaration

I, Ho Ting Henry Chan, declare that the Thesis I am submitting is entirely my own work except where indicated in the text, caption, footnote, or bibliography. This Thesis has not been submitted in whole or in part for any other academic degree or professional qualification.



23 August 2023

Signature *Date*

Acknowledgements

I would like to thank my supervisors, Professor Fernanda Duarte and Professor Chris Schofield, for their supervision, guidance, and support throughout my time here at Oxford, and for giving me the opportunity to carry out a joint research project. The application of chemical knowledge to understand biology and aid drug discovery has always been my biggest motivation to study chemistry, and it has been an honour to participate in the collaborative work to help address the challenges posed by COVID-19. I would also like to thank the EPSRC SBM CDT and its sponsors, the Clarendon Fund, and New College for funding my studies.

I thank my peers and colleagues in the entire Duarte and Schofield groups for the companionship, friendly work environments, and the constant exchange of knowledge and expertise. I would particularly like to thank Tomasz, Ally, Shaoqi, Hafiz, Bernie, Tristan, and Bastien for fruitful discussions in the MD subgroup, and additionally, Matina, Alistair, Tom, Said, Chloe, Hanwen, Veronika, and Aleksy for the always helpful and cheerful conversations in the lab. I am grateful to Tika, Lennart, and Toby for their experimental work and for very kindly sharing experimental results and testing hypotheses. The collaborative work on M^{pro}, carried out entirely via weekly Zoom meetings, would not be possible without all the collaborators involved both within and outside Oxford. I would like to thank all of them, especially Victor (Oxford Chemistry), Garrett and Marc (Oxford Statistics), Martin and David (Diamond), Adrian, Sofia, Becca, Debbie and Richard (Bristol), and Luigi (Grenoble).

I would like to thank my family, especially my parents as well as my aunt, uncle, and cousins here in the UK, for their love and encouragement, no matter how difficult times could get over the past several years. I am grateful to my friends for their continuous support and fellowship. Thanks be to God, my Saviour and Shepherd, whose goodness and love follow me all the days of my life.

List of Publications

This integrated Thesis is based upon the following publications and manuscripts, with my contributions described in the Statement of Authorship following the manuscript:

‡ First author equal. * Corresponding authors.

I. Chan, H. T. H.‡; Moesser, M. A.‡; Walters, R. K.‡; Malla, T. R.‡; Twidale, R. M.; John, T.; Deeks, H. M.; Johnston-Wood, T.; Mikhailov, V.; Sessions, R. B.; Dawson, W.; Salah, E.; Lukacik, P.; Strain-Damerell, C.; Owen, C. D.; Nakajima, T.; Świderek, K.; Lodola, A.; Moliner, V.; Glowacki, D. R.; Spencer, J.; Walsh, M. A.; Schofield, C. J.*; Genovese, L.*; Shoemark, D. K.*; Mulholland, A. J.*; Duarte, F.*; Morris, G. M.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021**, *12* (41), 13686-13703.

II. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.*; Duarte, F.* Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3* (6), 1767-1774.

III. Chan, H. T. H.; Oliveira, A. S. F.; Mulholland, A. J.; Schofield, C. J.; Duarte, F. Substrate specificity and recognition in each SARS-CoV-2 main protease subsite informed by dynamical nonequilibrium molecular dynamics simulations. *Manuscript in preparation* **2023**.

IV. Chan, H. T. H.; Brewitz, L.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A.; Schofield, C. J.*; Duarte, F.* Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2' proline of the viral polyprotein. *bioRxiv* **2023**, DOI: 10.1101/2023.07.11.548309. Submitted to *RSC Chem. Biol.*

V. Brewitz, L.*; Chan, H. T. H.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A.; Duarte, F; Schofield, C. J.* Mass spectrometric assays monitoring the deubiquitinase activity of the SARS-CoV-2 papain-like protease inform on the basis of substrate selectivity and have utility for substrate identification. *Submitted to J. Virol.* **2023**.

During this degree, a contribution was made to the following publication. Part of the results are discussed in this Thesis:

VI. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.; Owen, C. D.; Lukacik, P.; Chan, H. T. H.; Maheswaran, P.; Salah, E.; Duarte, F.; Yang, H.; Rao, Z.; Walsh, M. A.; Schofield, C. J.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57* (12), 1430-1433.

List of Abbreviations

$\angle$	Angle
↓	Cleavage site scissile amide
|	Cleavage site
α	Effective Born radius
γ	Surface tension coefficient / friction constant
Δ	Change
ΔG	Free energy change
ΔG_{bind}	Free energy change of binding
Δt	Timestep
ε	Minimum energy of Lennard-Jones potential / permittivity
θ	Bond angle
κ	Inverse Debye length
λ	Mixing or change variable between two Hamiltonians
μ	Reduced mass
ρ	Charge density
σ	Collision parameter of Lennard-Jones potential / symmetry number
τ	Time constant for thermostat coupling
ϕ	General dihedral angle / the amino acid backbone C_{i-1}-N-Cα-C dihedral angle
χ	The amino acid sidechain N-Cα-Cβ-Xγ dihedral angle, where X is a non-hydrogen atom
ψ	The amino acid backbone N-Cα-C-N_{i+1} dihedral angle
ω	Oscillation angular frequency
$3CL^{pro}$	3 Chymotrypsin-like protease
4EHP	Eukaryotic translation initiation factor 4E homologous protein
$\boldsymbol{a}$	Acceleration
A	Area
Å	Angstrom
AA	Amino acid
ABFE	Absolute binding free energy
Ac	Acetyl
ACE	Acetyl capping group
ACE2	Angiotensin-converting enzyme 2
AD4	AutoDock4
ADCP	AutoDock CrankPep
AM1-BCC	AM1 derived population charges with additive bond charge corrections
AMC	7-Amino-4-methylcoumarin
amu	Atomic mass unit
API	Active pharmaceutical ingredient
ATG7	Ubiquitin-like modifier-activating enzyme, also known as autophagy-related protein 7
b	Bond distance
BAR	Bennett acceptance ratio
BL2	Blocking loop 2
Boc	*tert*-Butyloxycarbonyl
Bz	Benzyl
c	Offset constant in non-polar solvation energy
CAS	Computational alanine-scanning mutagenesis
Ch	Chain
COM	Centre of mass
comp	Protein-ligand complex
conf	Conformational

cont	Contact
CoV	Coronavirus
COVID-19	Coronavirus disease 2019
CT	C-terminal
CV	Collective variable
CYM	Unprotonated, negatively charged cysteine
CYP3A	Cytochrome P450 3A
d	Distance between two atoms
Da	Dalton
D-A	Donor-acceptor
Dap	L-2,3-Diaminopropionic acid
decoup	Decoupling or annihilation
Deprot	Deprotonation
DFT	Density functional theory
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
D-NEMD	Dynamical nonequilibrium molecular dynamics
DUB	Deubiquitinase
E	Energy
E protein	Envelope protein
ECFP	Extended-connectivity fingerprint
E. coli	*Escherichia coli*
eIF4E	Eukaryotic translation initiation factor 4E
eIF4G	Eukaryotic translation initiation factor 4G
elec	Electrostatics
eqm	Equilibrium
ESP	Electrostatic potential
EVB	Empirical valence bond
EXP	Exponential averaging
F	Force
FDA	United States Food and Drug Administration
FEP	Free energy perturbation
FES	Free energy surface
FF	Force field
FlexRes	Flexible residue
FMDV	Foot-and-mouth disease virus
Fmoc	Fluorenylmethyloxycarbonyl
FRET	Fluorescence resonance energy transfer
G	Free energy
GA	Genetic algorithm
GAFF	General Amber force field
GB	Generalised Born
HB	Hydrogen bond
HCoV	Human coronavirus
H-D-A	Hydrogen-donor-acceptor
HDMS	High definition mass spectrometer
HEBT	Secondary structure (α-helix + extended strand + β-bridge + hydrogen bonded turn)
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HF	Hartree-Fock
HID	Nδ-protonated, neutral histidine
HIE	Nε-protonated, neutral histidine
HIP	Nδ and Nε-protonated, positively charged histidine

HIV-1	Human immunodeficiency virus 1
HPLC	High performance liquid chromatography
hTyr	L-Homotyrosine
IC_{50}	Half maximal inhibitory concentration
IFN	Interferon
igb	Implicit generalised Born solvation model
iMD-VR	Interactive molecular dynamics in virtual reality
int	Internal
IP	Ion pair state
IRF3	Interferon regulatory factor 3
ISG15	Interferon stimulated gene 15
k	Force constant
k_B	Boltzmann constant
k_{cat}	Constant describing enzyme catalytic rate
k_{cat}/K_m	Constant describing enzyme catalytic efficiency
KDE	Kernel density estimation
KE	Kinetic energy
K_m	Michaelis constant
KS	Kohn-Sham
LC	Liquid chromatography
LCMS	Liquid chromatography-mass spectrometry
Ldt_{Mt2}	*Mycobacterium tuberculosis* L,D-transpeptidase 2
LGA	Lamarckian genetic algorithm
lig	Ligand
LINCS	Linear constraint solver
LJ	Lennard-Jones
L^{pro}	Leader protease
LYC	N^{ε}-branched lysine
LYN	Neutral lysine
m	Mass
M protein	Membrane protein
Mac	Macrodomain
MALDI	Matrix-assisted laser desorption/ionisation
MB	Maxwell-Boltzmann
MBAR	Multistate Bennett acceptance ratio
MBHA	4-Methylbenzhydrylamine
MC	Monte Carlo
MCS	Maximum common substructure
MD	Molecular dynamics
MDA5	Melanoma differentiation-associated protein 5
MERS-CoV	Middle East respiratory syndrome coronavirus
ML	Machine learning
MM	Molecular mechanics
MM/GBSA	Molecular mechanics/generalised Born surface area
MM/PBSA	Molecular mechanics/Poisson Boltzmann surface area
M^{pro}	Main protease
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
Mtt	4-Methyltrityl
mut	Mutation
m/z	Mass-to-charge ratio
n	Multiplicity of periodic function
N	Neutral state

N	Number of atoms / Size of the dataset
N protein	Nucleocapsid protein
NEDD8	Neural precursor cell expressed, developmentally down-regulated protein 8
NEMO	Nuclear factor-κ-B essential modulator
neq	Nonequilibrium
NF	Nuclear factor
NME	N-methyl capping group
NP	Non-polar solvation term
NPT	Constant number of atoms, pressure, and temperature
nsp	Non-structural protein
NT	N-terminal
Nuc	Nucleophilic addition
NVT	Constant number of atoms, volume, and temperature
OBC2	Generalised Born model II by Onufriev, Bashford and Case in *Proteins* **2004**, *55*, 383, with parameters $\alpha = 1.0$, $\beta = 0.8$, $\gamma = 4.85$.
ODDT	Open Drug Discovery Toolkit
ONIOM	Our own n-layered integrated molecular orbital and molecular mechanics
ORF	Open reading frame
P	Position N-terminal to the scissile amide
P	Pressure
P′	Position C-terminal to the scissile amide
PB	Poisson Boltzmann
PBC	Periodic boundary conditions
PDB	Protein Data Bank
PE	Potential energy
PL^{pro}	Papain-like protease
PME	Particle-mesh Ewald
PMF	Potential of mean force
pp	Polyprotein
PreSaVS	Predictive saturation variation scanning
PROS1	Vitamin K-dependent protein S
prot	Protein
PT	Proton transfer
q	Charge
QM	Quantum mechanics
QM/MM	Hybrid quantum mechanics/molecular mechanics
Q-TOF	Quadruple time-of-flight
r	Distance
$\boldsymbol{r}$	Position
$\boldsymbol{R}$	Random force in stochastic dynamics
RBD	Receptor-binding domain
RBFE	Relative binding free energy
RC	Reaction coordinate
ReaxFF	Reactive force field
Rep	Replica simulation
RESP	Restrained electrostatic potential
rest	Restraint
RF	RapidFire
RMS	Root mean square
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
rpm	Revolutions per minute

S	Protein subsite that binds a substrate residue N-terminal to the scissile amide
S	Entropy / AutoDock4 force field solvation parameter
S′	Protein subsite that binds a substrate residue C-terminal to the scissile amide
S protein	Spike protein
SAM	*S*-adenosyl methionine
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SASA	Solvent accessible surface area
SD	Standard deviation
SEM	Standard error of the mean
sol	Solvent
solv	Solvation
SPE	Solid phase extraction
SPE-MS	Solid phase extraction-mass spectrometry
SPPS	Solid phase peptide synthesis
SuCOS	Shape and pharmacophoric overlap
SUMO1	Small ubiquitin-related modifier 1
t	Time
T	Temperature
T1/T2	Ligand binding stability test 1/2
TFA	Trifluoroacetic acid
TI	Thermodynamic integration
TIPS	Triisopropylsilane
TMEM	Transmembrane protein
TMPRSS2	Transmembrane protease serine 2
tors	Torsion angle
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
TS	Transition state
U	Potential energy
U′	Biased potential energy
Ub	Ubiquitin
UBL	Ubiquitin-like modifier
ULK1	Serine/threonine unc-51-like kinase
URM1	Ubiquitin-related modifier 1
$URM1^{+}$	$IRF3_{189\text{-}197}$-URM1 peptide, modelled with a positively charged P3 histidine (HIP)
$URM1^{N}$	$IRF3_{189\text{-}197}$-URM1 peptide, modelled with a neutral P3 histidine (HIE)
US	Umbrella sampling
USP	Ubiquitin specific peptidase
v	Magnitude of the average displacement vector
v	Velocity
V	Volume
v.	Version
vac	Vacuum
vdW	van der Waals
v/v	Volume/volume ratio
w	Weighting coefficient
W	Biasing potential
WHAM	Weighted histogram analysis method
wt	Wild type
X	Any proteinogenic amino acid
Z	Zwitterionic state

Standard 1-letter and 3-letter abbreviations for amino acids are used in this thesis.

Table of Contents

Abstract i

Declaration iii

Acknowledgements iv

List of Publications v

List of Abbreviations vii

Table of Contents xii

1 Introduction 1

1.1 SARS-CoV-2 drug targets 1

1.2 Main protease (M^{pro}) 4

1.3 Papain-like protease (PL^{pro}) 9

1.4 Aims and thesis overview 13

1.5 References 15

2 Theoretical Background 25

2.1 Molecular mechanics 25

2.2 Molecular dynamics 28

2.3 Umbrella sampling 32

2.4 Dynamical nonequilibrium molecular dynamics 33

2.5 Binding energy calculations 34

2.6 Docking 38

2.7 Quantum mechanics/molecular mechanics 40

2.8 References 43

3 M^{pro}-Substrate Interactions to Guide Inhibitor Design 46

3.1 Overview 46

3.2 Manuscript I 46

3.3 Statement of authorship 93

3.4 β-lactam inhibitors 94
3.5 References: β-lactam inhibitors 99
4 M^{pro} Substrate Binding Dynamics 100
4.1 Overview 100
4.2 Manuscript II 100
4.3 Statement of authorship 119
5 M^{pro} Substrate Recognition in each Subsite 120
5.1 Overview 120
5.2 Manuscript III 120
5.3 Statement of authorship 146
6 PL^{pro} Selectivity on Polyprotein Substrates 147
6.1 Overview 147
6.2 Manuscript IV 147
6.3 Statement of authorship 177
7 PL^{pro} Binding of Ubiquitin-Like Substrates 178
7.1 Overview 178
7.2 Manuscript V 178
7.3 Statement of authorship 219
8 Conclusions 220
Appendix for Chapter
3A Supplementary Information to Manuscript I S1
3B Supplementary Information to Section 3.4: β-lactam inhibitors S54
4 Supplementary Information to Manuscript II S64
5 Supplementary Information to Manuscript III S85
6 Supplementary Information to Manuscript IV S162
7 Supplementary Information to Manuscript V S190

1. Introduction

1.1 SARS-CoV-2 drug targets

In December 2019, a severe pneumonia-causing disease of unknown aetiology was reported, characterised by symptoms such as fever, chest tightness, and cough.[1, 2] The disease, later named COVID-19, was declared a global health emergency by the World Health Organization from January 2020 to May 2023. During this period more than 765 million cases were reported globally with almost 7 million deaths.[3] This period also witnessed a significant loss of economic activity globally, with a series of lockdowns being imposed in multiple countries to curb the spread of the pandemic.

Advances in genome sequencing, leading to the reported sequence (GenBank MN908947) of the Wuhan-Hu-1 isolate,[1] enabled the rapid identification of a coronavirus (CoV) as the causative agent of COVID-19. CoV is a positive-sense, single-stranded RNA virus with a spherical envelope and crown-like spikes on its surface.[4, 5] This new CoV, which belongs to the betacoronavirus genus, shared 79.6% sequence identity with the severe acute respiratory syndrome coronavirus (SARS-CoV) that caused a deadly but localised outbreak in 2002-2003, and hence was named SARS-CoV-2.[4] Human infections by coronaviruses (HCoVs) are well precedented, with HCoV-229E, HCoV-OC43, HCoV-HKU1, and HCoV-NL63 causing the common cold. Additionally, both SARS-CoV and Middle East respiratory syndrome CoV (MERS-CoV; ~50% sequence identity to SARS-CoV-2)[6] have the potential to cause severe pneumonia.[7] However, in 2019 broadly useful prophylactic or therapeutic agents against these HCoVs had not been developed.[6]

The SARS-CoV-2 genome consists of ~29.9k nucleotides, over two thirds of which (at the 5′ end) comprises the overlapping open reading frames (ORFs) ORF1a and ORF1ab, which, respectively, encode the polyproteins (pps) pp1a and pp1ab.[8] The remaining genome encodes the structural

proteins, including spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins, as well as other accessory proteins.[9, 10] In particular, the receptor-binding domain (RBD) of the S protein is responsible for binding the host cell receptor angiotensin-converting enzyme 2 (ACE2) or transmembrane protein TMEM106B, with transmembrane protease serine 2 (TMPRSS2)-catalysed cleavage enabling viral entry into host cells.[11-14] The SARS-CoV-2 structural proteins have provided a basis for COVID-19 vaccine development.[10]

Table 1.1: Proposed roles of the SARS-CoV-2 non-structural proteins (nsps).[10, 15-18]

Polyprotein		nsp	Protein length	Examples of proposed protein role
pp1a	pp1ab	nsp1	180	Binds to host ribosome and alters host translation.
		nsp2	638	Interacts with host proteins PHB1/2 to regulate host intracellular signalling.
		nsp3	1945	Catalyses peptide hydrolysis to release nsps and to evade host immune response (PL^{pro} domain); counteracts host-mediated antiviral signalling (Mac1 domain).
		nsp4	500	Rearranges endoplasmic reticulum membrane and assembles double-membrane vesicles.
		nsp5	306	Catalyses peptide hydrolysis to release nsps (M^{pro}).
		nsp6	290	Regulates autophagy.
		nsp7	83	Forms primase complex with nsp8, and acts as cofactor for nsp12 in the replication and transcription complex.
		nsp8	198	Forms primase complex with nsp7, and acts as cofactor for nsp12 in the replication and transcription complex.
		nsp9	113	Binds single-stranded RNA/DNA; involved in formation of the replication and transcription complex.
		nsp10	139	Cofactor for nsp14 and nsp16.
		nsp11	13	Disordered; role unknown.
	pp1ab	nsp12	932	RNA-dependent RNA polymerase.
		nsp13	601	Helicase and RNA triphosphatase.
		nsp14	527	N7-guanine methyltransferase and exoribonuclease.
		nsp15	346	Uridine-specific endoribonuclease.
		nsp16	298	2′-O-Methyltransferase.

Following the release in the host cell cytoplasm, SARS-CoV-2 viral RNA binds to the ribosome and initiates the synthesis of pp1a and pp1ab; pp1ab is encoded by ~21.5k nucleobases.[19, 20] These polyproteins are long peptides that include the non-structural proteins (nsps), with pp1a comprising nsp1-11, and pp1ab comprising nsp1-16.[21, 22] Notably, the formation of pp1ab involves a ribosomal frame shift in nsp11,[20] a disordered peptide of 13 amino acids (AAs) which has the same first 9 N-terminal amino acids (AAs) as does nsp12.[15] Each of these nsps 1-16,

except for nsp11, plays indispensable roles in SARS-CoV-2 replication (**Table 1.1**),[15-17] necessitating its hydrolytic cleavage from the polyproteins by viral proteases at specific sites. This enables the folding and functioning of each nsp.[23] Replicated virions are subsequently assembled and released via exocytosis to infect other host cells, after which the infection lifecycle is repeated.[12, 19]

Potential antiviral treatments aim to stall the SARS-CoV-2 replication-infection lifecycle, in which various proteins or protein complexes can be targeted, including those involved in cell entry such as the S protein, ACE2, and TMPRSS2, the nsp12 RNA-dependent RNA polymerase, and the proteases responsible for polyprotein processing.[24-26] The potential of repurposing molecules with known safety profiles or previous clinical approvals for use as COVID-19 drugs has been explored, including large-scale screens.[27, 28]

SARS-CoV-2 employs two nucleophilic cysteine proteases, namely the nsp5 main protease (M^{pro}), and the papain-like protease (PL^{pro}), which is a domain within nsp3, to release nsps4-16 and nsps1-3, respectively (**Figure 1.1**).[29, 30] M^{pro} and PL^{pro} have essential roles of in the SARS-CoV-2 lifecycle; inhibition of either protease would hinder nsp release and block the subsequent SARS-CoV-2 maturation and replication processes.[10, 31] The potential of developing viral protease inhibitors as drugs is precedented, as exemplified by drugs inhibiting the human immunodeficiency virus 1 (HIV-1) serine protease.[32-34] Cysteine proteases are particularly attractive drug targets,[35, 36] because of the nucleophilicity of the cysteine residue and the availability of electrophilic warheads with tuneable reactivities.[37-42] These characteristics facilitate the design of potent covalently binding drugs that are less susceptible to mutation-led resistance and have more prolonged activities.[43-46]

Indeed, since 2019 both M^{pro} and PL^{pro} have attracted considerable interest amongst scientists worldwide as COVID-19 drug targets, including those in the computational chemistry community.[47] A wealth of computational modelling and simulation approaches have been utilised,

for instance, to predict the structures of the coronavirus, its variants and constituent proteins,[48] to understand biomolecular mechanisms and interactions,[49-51] and with the aim of facilitating inhibitor screening and optimisation.[52-54] The following sections discuss how structural, functional, and mechanistic understanding on M^{pro} and PL^{pro} has aided and/or guided COVID-19 drug discovery efforts.

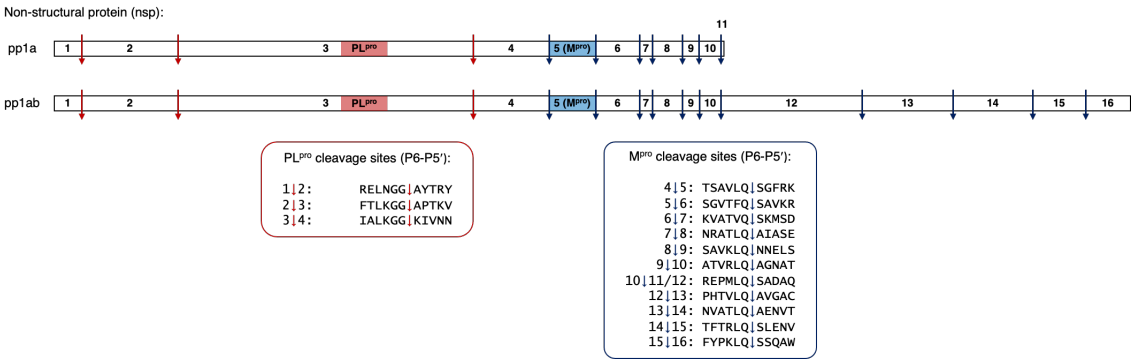


Figure 1.1: Polyprotein processing by M^{pro} and PL^{pro}.

1.2 Main protease (M^{pro})

The SARS-CoV-2 nsp5 main protease (M^{pro}), also known as the 3 chymotrypsin-like protease ($3CL^{pro}$),[55] is a nucleophilic cysteine protease responsible for catalysing hydrolysis at 11 distinct cleavage sites in the viral polyproteins to release nsps 4-16 (**Figure 1.1**).[56-58] M^{pro} shares 96% AA sequence identity with SARS-CoV M^{pro}, and lower sequence identity (50%) with MERS-CoV M^{pro}.[5, 57, 59] Since there are no known human proteases with a similar substrate selectivity, M^{pro} targeting drugs are proposed to be unlikely to lead to side effects.[56, 60] Given the essential role of M^{pro} in the viral lifecycle, its conserved nature across coronaviruses, and the lack of similar human proteases, SARS-CoV-2 M^{pro} presents an attractive drug target.

Functionally, M^{pro} recognises the consensus sequence in the SARS-CoV-2 polyproteins: [P4:Ala/Val/Pro/Thr]-[P3:X]-[P2:Leu/Phe/Val]-[P1:Gln]↓[P1′:Ser/Ala/Asn], where "X" denotes any proteinogenic AA and "↓" the cleavage site (**Figure 1.2**).[1, 61] Residues N-terminal and C-

terminal to the scissile amide are denoted as "P" and "P′", respectively, with an ascending number from 1 in both directions away from the scissile amide.[62, 63] Corresponding subsites on the protease that bind these substrate residues are denoted "S" and "S′", respectively. For instance, the conserved substrate P1-Gln residue binds in the S1 subsite on M^{pro}.

Comparison of the 11 native polyprotein cleavage site sequences of SARS-CoV and SARS-CoV-2 M^{pro}s reveals high similarity, except for the presence of P2-Met and P1′-Gly in certain substrates in SARS-CoV.[64, 65] Interestingly, M^{pro} is responsible for its own release from the polyproteins, via the initial autocleavage of its N-terminal at nsp4/5 and subsequently its C-terminal at nsp5/6; however, the exact autoprocessing mechanism is not fully understood.[66-69] In addition to the viral polyproteins, M^{pro} catalyses cleavage in certain human proteins such as TAB1, NLRP12,[23] MAP4K5,[70] HDAC2,[71] and NEMO,[72] which were suggested to relate to immune response regulation and to brain pathology.[73, 74]

Structurally, M^{pro} functions as a homodimer, with this dimer state being crucial for catalysis, at least when in isolated form (**Figure 1.2**).[75, 76] Indeed, preventing dimerization is a strategy for M^{pro} inhibition.[77] Interestingly, the N-terminal region of each M^{pro} chain is necessary for dimerization and forms part of the substrate binding pockets,[57, 67, 78, 79] which raises further questions on how autoprocessing occurs. Through mutational studies and molecular dynamics (MD) simulations, it has been observed that the two protomers in the M^{pro} homodimer exhibit asymmetric behaviour,[80-82] with only one protomer being catalytically active at a time.[83]

The substrate binding pockets on M^{pro} have been characterised using X-ray crystallography, initially with SARS-CoV M^{pro} [84] and later, during the pandemic, with SARS-CoV-2 M^{pro}.[85-88] At its active site, M^{pro} employs a catalytic dyad comprising the nucleophilic cysteine Cys145, and the general base His41. The exact ionic state of the Cys-His dyad has attracted considerable interest, with investigations conducted through measurements of the pH dependence of activity,[89-

[91] neutron diffraction,[92, 93] and computational studies in both *apo* and substrate- or inhibitor-bound M^{pro}.[94-98] These studies have yielded varying conclusions.

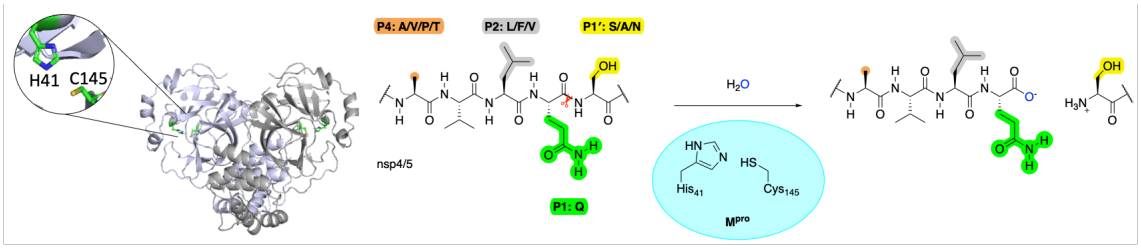


Figure 1.2: A view of the M^{pro} homodimeric structure (PDB 6YB7),[99] and a generalised reaction for M^{pro} catalysis. Adapted from **Manuscript I** (Reference **100**) with permission from the Royal Society of Chemistry.[100]

Efficient screening of M^{pro} inhibitors has been enabled by the development of *in vitro* assays that monitor hydrolytic activity towards peptidic or peptidomimetic substrates, often derived from the native polyprotein cleavage sites. For example, activity can be monitored by the release of a fluorescent tag C-terminal to the scissile amide bond;[101] however, such assays might not accurately reflect the binding of P′ residues of the native polyprotein substrates. Alternatively, the separation of the N- and C-terminal product fragments following hydrolysis can be monitored by fluorescence resonance energy transfer (FRET)-based techniques.[60, 102, 103] Malla *et al.* (2021) employed a solid phase extraction (SPE) coupled with mass spectrometry (MS) (SPE-MS) method, via a RapidFire (RF) sampling robot coupled to a quadrupole time-of-flight (Q-TOF) mass spectrometer, to directly monitor in the mass spectra substrate depletion and the formation of the N- and C-terminal product fragments over time to measure M^{pro} activity (**Manuscript VI**).[104] This method enables measurement of enzyme activity towards different peptide sequences, as well as high-throughput screening of inhibitors using the nsp4/5-derived TSAVLQ↓SGFRK-NH_2 substrate. The reaction mechanisms of M^{pro} with these assayed substrates have been studied using quantum mechanics/molecular mechanics (QM/MM) calculations,[105] although questions remain regarding the presence of a zwitterionic Cys-His dyad intermediate state and the role of His41 and the P′ leaving group in deprotonating the hydrolytic water molecule.[106, 107]

The discovery, design, and optimisation of M^{pro} inhibitors has focused around three primary targets: (i) the active site;[31, 54, 108, 109] (ii) dimer interface;[77] and (iii) other allosteric inhibition sites.[19] Efforts to search for potent M^{pro} inhibitors have often been accompanied, and sometimes driven, by computational modelling,[49, 110, 111] ranging from large-scale docking-based screening of compounds,[112-114] to free energy perturbation (FEP)-based optimisation[115-117] and machine learning (ML)-driven inhibitor discovery and prioritisation.[118, 119]

Most reported M^{pro} inhibitors have been designed to directly bind to the substrate binding active site of M^{pro}. Activity can be blocked by covalently modifying the essential active-site Cys145 residue, as exemplified by the potent inhibition shown by the thiol-reacting ebselen and its derivatives, which can react with up to 12 cysteines per M^{pro} protomer.[104, 120, 121] More targeted inhibition of Cys145 can be achieved by peptidomimetic scaffolds that mimic the native polyprotein cleavage sites (**Figure 1.3**).[95, 122, 123] Early in 2020, N3, which has a Michael acceptor moiety,[56] and a series of α-ketoamides were reported to be highly potent M^{pro} inhibitors that covalently modify Cys145.[57] Their inhibition mechanisms have been extensively studied by calculations.[95, 124-128] N3 and the α-ketoamides share a common feature: a γ-lactam-containing sidechain that binds the S1 pocket and closely mimics the Gln sidechain universally found across native polyprotein substrates.[129] Macrocyclic peptide scaffolds have also been explored in attempts to improve proteolytic stability and cell permeability relative to linear peptides.[12, 130] Numerous studies have considered repurposing or optimisation of known protease inhibitors, such as the hepatitis C virus NS3/4A serine protease inhibitors boceprevir, telaprevir, and narlaprevir.[55, 103, 131-134] In December 2021, Pfizer's Paxlovid, a combination of the M^{pro} inhibitor nirmatrelvir (PF-07321332) and the CYP3A inhibitor ritonavir, became the first oral COVID-19 treatment granted emergency use authorisation by the US Food and Drug Administration.[135, 136] Based on crystallographic observations, the binding mode of nirmatrelvir mimics those of substrates, with a Gln-mimicking γ-lactam-containing sidechain targeting S1, a cyclic Leu-mimicking 6,6-dimethyl-3-azabicyclo[3.1.0]hexane group targeting the S2 pocket, and a nitrile

electrophilic warhead.[135, 137] M^{pro}-nirmatrelvir interactions and nitrile reactivity have been extensively characterised by MD and QM/MM studies.[138-141]

Non-peptidic, non-covalent inhibitors of M^{pro} have been explored, such as the tyrosine kinase inhibitor masitinib,[142] and Shionogi's ensitrelvir (Xocova, S-217622) (**Figure 1.3**), an oral antiviral treatment that received emergency authorisation in Japan in late 2022.[143, 144] Both compounds bind in the S1 pocket via a nitrogen-containing heterocycle, and in the S2 pocket via aromatic moieties.[142, 144, 145]

Peptide inhibitors,[77, 146] and small molecule fragments, such as x1187 (**Figure 1.3**), have been investigated for disrupting the dimerization of SARS-CoV or SARS-CoV-2 M^{pro}, which has been confirmed via crystallography-based fragment screening.[147] Using protein native MS, x1187 was observed to disrupt SARS-CoV-2 M^{pro} dimerization and thus activity.[79]

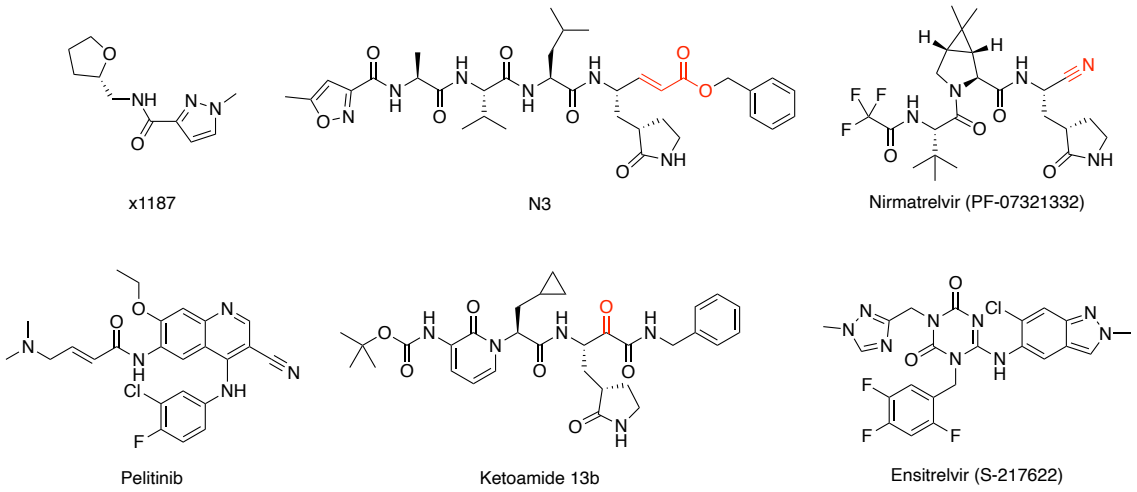


Figure 1.3: Examples of M^{pro} inhibitors, with electrophilic warheads in red where present.[56, 57, 135, 143, 147, 148]

Various allosteric M^{pro} ligands have been identified from multiple large-scale crystallographic screening experiments.[147, 148] However, whether such allosteric binding diminishes M^{pro} activity requires subsequent validation. For example, pelitinib (**Figure 1.3**) was observed to bind to a hydrophobic pocket formed by Ile213, Leu253, Gln256, Val297, and Cys300 on M^{pro} and was

subsequently validated to exhibit antiviral activity.[148] The identification of allosteric inhibition sites has been aided by computational methods, such as MD[5, 149-151] and protein residue network models.[152, 153]

Given the selection pressure on SARS-CoV-2 by antiviral treatments, unsurprisingly M^{pro} variants that lead to resistance against the existing drugs have been reported.[154-157] Experimental[158-162] and computational[163, 164] studies aimed at predicting and understanding how AA substitutions in M^{pro} create drug resistance are ongoing and will benefit next generation drug development.

1.3 Papain-like protease (PL^{pro})

The SARS-CoV-2 papain-like protease (PL^{pro}) is a domain within nsp3, which also includes an N-terminal ubiquitin (Ub)-like domain, a ADP-ribose phosphatase domain, a SARS-unique domain consisting of multiple macrodomains (Macs), a nucleic acid binding domain, transmembrane segments, and a C-terminal zinc finger domain.[165, 166] PL^{pro} is responsible for polyprotein cleavage at three sites, which are nsp1/2, nsp2/3, and nsp3/4.[23, 167] Among coronaviruses, the AA sequences of PL^{pro}s are less conserved than for M^{pro}s.[168] For instance, the SARS-CoV-2 PL^{pro} shares only 83% and 31% identities with its analogue in SARS-CoV and MERS-CoV, respectively.[169, 170] Other coronaviruses, such as the murine coronavirus mouse hepatitis virus and HCoV-229E, may encode an extra PL^{pro} domain termed PLP1, with the SARS-CoV-2 equivalent counterpart referred to as PLP2.[171] The sequence identities between PL^{pro}s of SARS-CoV and other coronaviruses are approximately 40%.[171]

Functionally, SARS-CoV-2 PL^{pro} recognises and selectively cleaves the following consensus sequence: [P4:Leu]-[P3:X]-[P2:Gly]-[P1:Gly]↓[P1′:X] (**Figure 1.4**). MERS-CoV PL^{pro} also recognises a P4:Ile at its nsp3/4 cleavage site.[8, 172, 173] The P3 position of PL^{pro} substrates apparently tolerates any proteinogenic AAs, although hydrophobic (Leu/Tyr) and positively charged (Lys/Arg) residues are preferred by SARS-CoV and SARS-CoV-2 PL^{pro}s.[174, 175]

Nevertheless, the requirements for glycine at P2 and P1 appear to be universal across SARS-CoV, SARS-CoV-2 and MERS-CoV PL^{pro}s.[8, 173, 175]

In addition to its role in polyprotein processing, PL^{pro} is known to cleave host protein substrates. Among known human substrates, PL^{pro} was observed to cleave the serine/threonine unc-51-like kinase (ULK1) C-terminal to a LGGG sequence, resulting in autophagy disruption,[176] and to cleave the interferon (IFN) regulatory factor 3 (IRF3) C-terminal to a LGGG sequence, thus suppressing the IFN immune response during infections.[23] Hydrolysis of sequences derived from human proteins such as the cardiac myosins MYH6/MYH7 (P4-P1 LKGG) and the plasma protein S PROS1 (P4-P1 LRGG) has been proposed to relate to cardiovascular and blood-related symptoms in COVID-19 infections.[177] Furthermore, the apparent lack of specificity at the P′ positions C-terminal to the scissile amide allows PL^{pro} to hydrolyse many substrates, including isopeptide bonds where the amide nitrogen is the N^{ε} of a lysine sidechain, a common motif in Ub and Ub-like modifier (UBL) scaffolds. PL^{pro} catalyses the cleavage of the UBL interferon stimulated gene 15 (ISG15), which has a C-terminal LRGG sequence,[177] from IRF3[178] and melanoma differentiation-associated protein 5 (MDA5).[179] Since ISG15 conjugation of IRF3 and MDA5 are critical markers for stimulating human antiviral type I IFN responses, deISG15ylation by PL^{pro} attenuates immune response against the virus.[178, 179] Similarly, PL^{pro} recognises the C-terminal LRGG sequence of Ub , and catalyses the removal of Lys48-linked di-Ub (K48-Ub_2) modifications, which are markers directing viral proteins towards human proteasomal degradation.[180-183] To summarise, PL^{pro}-catalysed hydrolysis of host proteins, especially its deubiquitinating and deISG15ylating capabilities, plays significant pro-viral roles in addition to its primary essential function of viral polyprotein processing.

Structurally, PL^{pro} is a single-chain enzyme and adopts a right-handed shape (**Figure 1.4**).[184] The N-terminal region of PL^{pro} consists of a Ub-like domain attached to the thumb domain of the hand. In SARS-CoV and MERS-CoV PL^{pro}s, this Ub-like domain is non-essential for PL^{pro} activity.[168, 185] In the palm region, the active site consists of a catalytic triad composed of the nucleophilic

cysteine residue Cys111, His272, and Asp286.[169] Studies on the pH-dependent activity of SARS-CoV PL^{pro} have suggested the presence of a Cys-His zwitterionic ion pair.[186] At the tip of the fingers domain, a zinc ion is tetrahedrally coordinated by four cysteine residues, Cys189, 192, 224, and 226.[169] The presence of this zinc finger is essential for the structural integrity and proteolytic activity of PL^{pro}, and zinc ejection by disulfiram and gold organometallics leads to loss of PL^{pro} activity.[187, 188] Residues located between the palm, fingers, and thumb regions interact with ISG15 and Ub,[189] while the ridge region of the thumb domain interacts with the distal Ub unit in PL^{pro}-bound K48-Ub$_2$.[180, 190] Differences in the identity and flexibility of PL^{pro} residues in these Ub/UBL-interacting regions likely contribute to substrate selectivity differences between SARS-CoV and SARS-CoV-2 PL^{pro}s, with SARS-CoV PL^{pro} preferentially cleaving K48-Ub$_2$ and SARS-CoV-2 PLpro preferentially cleaving ISG15.[8, 170, 182, 183, 191]

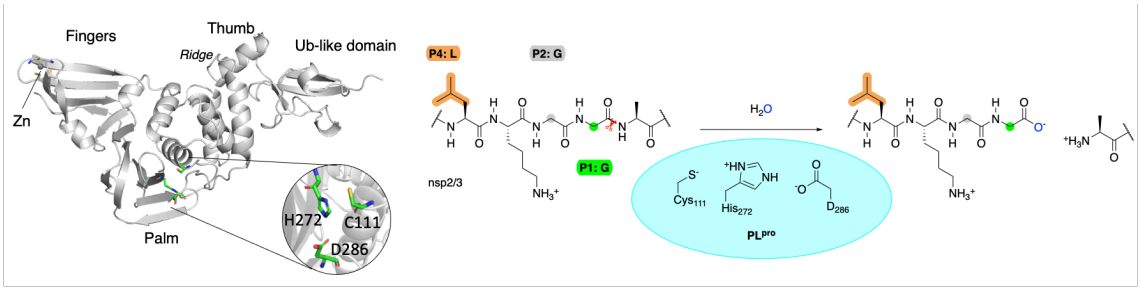


Figure 1.4: A view of PL^{pro} (PDB 6WX4; VIR251 inhibitor omitted)[175] and its general reaction scheme.

In similar manner to M^{pro}, multiple *in vitro* high throughput screening methods have been developed for identifying PL^{pro} inhibitors, with the majority based on the release of a C-terminal fluorescent tag[175, 178, 184, 192, 193] or FRET.[20, 194, 195] Considering the native polyprotein cleavage site sequences, Brewitz *et al.* (2022) reported a SPE-MS based assay, with the nsp2/3 derived oligopeptide observed to be the most efficient substrate, and thus utilised in subsequent inhibitor screening.[196]

Similar to M^{pro}, inhibition of PL^{pro} can be achieved by covalently modifying the active-site Cys111, such as with ebselen and its derivatives (**Figure 1.5**).[196, 197] Specificity can be improved

by considering the consensus P4-P1 LXGG substrate sequence during inhibitor design, as exemplified by the peptidomimetic, Michael acceptor-containing covalent inhibitors VIR250 and VIR251, the latter of which targets the S4 pocket with homotyrosine (hTyr), S3 with a 2,3-diaminopropionic acid (Dap) residue, and S2 and S1 with two glycine residues.[175] However, given the relative tightness of the S2-S1 channel proximate to Cys111, many ligands have instead been designed to bind the hydrophobic, Leu-recognising S4 PL^{pro} pocket.[167] Examples include the deubiquitinase inhibitor GRL0617 that binds both S4 and S3,[178, 194, 198] the antitumour agent sepantronium bromide (YM155),[184] natural products including tanshinone I, cryptotanshinone,[184] and dihydrotanshinone I,[20] and the antimalarial amodiaquine.[199] The binding modes of many of these inhibitors have been probed by experimental[192, 200] and computational studies.[190, 199, 201, 202] Alternative PL^{pro} inhibition strategies have been reported, including zinc ejection,[187, 188] allosteric inhibition via covalent modification of Cys270 with dimesna or pyritinol,[203] and the use of Ub variants that bind to and inhibit PL^{pro}.[204] Nevertheless, to date no PL^{pro} inhibitors have been authorised for clinical use against COVID-19.[8]

As seen from the observed differences in deubiquitinating and deISG15ylating activities across coronavirus PL^{pro}s, it is of significant interest to understand and predict how variations in PL^{pro} primary sequence would impact its activity towards protein substrates, and the efficacy of existing inhibitors. Indeed, while GRL0617 is a potent inhibitor of SARS-CoV and SARS-CoV-2 PL^{pro}s, inhibition is significantly deteriorated in MERS-CoV PL^{pro}, attributed to the substitution of Tyr268 (SARS-CoV-2 numbering) with Thr at the equivalent position.[8, 167, 168, 172]

Considering the potential development of resistance towards existing M^{pro} and PL^{pro} inhibitors caused by viral mutations, the development of dual inhibitors targeting both SARS-CoV-2 proteases is of interest, despite notable differences in the structure, catalytic mechanism, and substrate selectivity of these two cysteine proteases.[22, 205] Several screening efforts for dual inhibitors of M^{pro} and PL^{pro} have been reported, often guided by computational approaches such

as docking,[206, 207] MD,[208] ML, and their combinations,[209] although further validation and optimisation are likely required to develop these dual inhibitors into drugs.

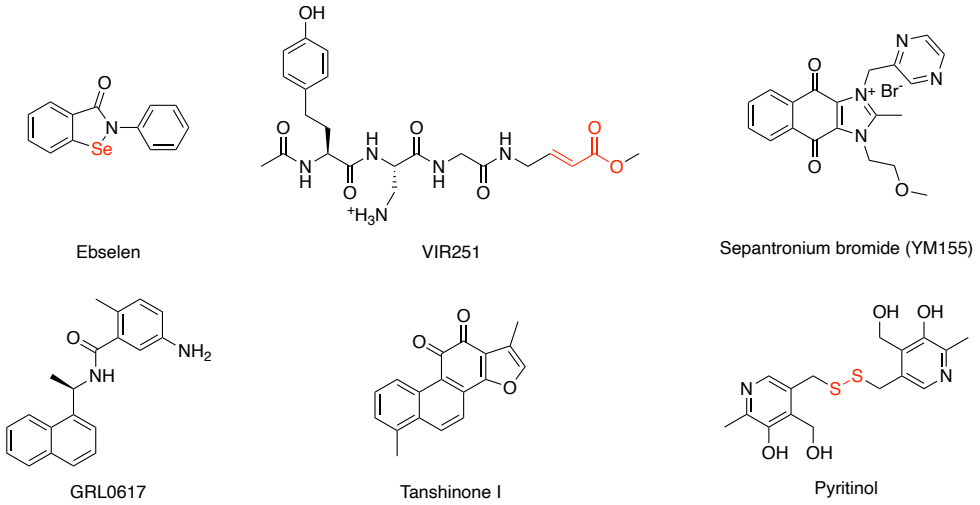


Figure 1.5: Examples of PL^{pro} inhibitors, with electrophilic warheads in red where present.[175, 178, 184, 197, 203]

1.4 Aims and thesis overview

Computational modelling has immense potential for investigating enzyme structure and dynamics, understanding substrate recognition and catalytic mechanisms, as well as facilitating inhibitor discovery and optimisation.[47, 49] Furthermore, SPE-MS-based high-throughput screens allow for the efficient acquisition of data on enzymatic activity and inhibition profiles, [104, 196] while also enabling validation of hypotheses from computational calculations, thus generating a feedback loop for investigations into M^{pro} and PL^{pro} activity and inhibition. The investigations described in this Thesis were initiated at the beginning of the pandemic in early 2020. The primary objective was to employ computational methods, the theoretical backgrounds of which are described in **Chapter 2**, to:

- investigate how M^{pro} binds its native substrates, when such crystal structures were not yet available, and elucidate how these interactions relate to the modes of inhibition exhibited by peptidic and small molecule M^{pro} inhibitors (**Chapter 3**);
- from these M^{pro}-substrate complexes, characterise the protein dynamics involved when M^{pro} adapts and responds to the association or dissociation of a whole substrate peptide chain (**Chapter 4**);
- dissect the M^{pro} dynamic responses on the binding or unbinding of each residue of the substrate, and explore how these responses could lead to substrate recognition (**Chapter 5**);
- enhance our understanding of the binding interactions between PL^{pro} and its three native polyprotein derived substrates, especially beyond the consensus P4-P1 LXGG sequence, and identify potential reasons for the experimentally observed differences in catalytic activity (**Chapter 6**);
- investigate how PL^{pro} recognises substrates that are representative of the deubiquitinating and deISG15ylating functions of PL^{pro} (**Chapter 7**).

It was intended that the obtained insights into enzyme dynamics, catalytic activity, and inhibition mechanisms would deepen our understanding of biological processes involving SARS-CoV-2 and guide the development of COVID-19 treatments.

1.5 References

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W*., et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265.
2. Telenti, A.; Hodcroft, E. B.; Robertson, D. L. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harb. Perspect. Med.* **2022,** *12*, a041390.
3. Wise, J. Covid-19: WHO declares end of global health emergency. *BMJ* **2023,** *381*, p1041.
4. Samuel, C. E. Interferon at the crossroads of SARS-CoV-2 infection and COVID-19 disease. *J. Biol. Chem.* **2023,** *299*, 104960.
5. Cho, E.; Rosa, M.; Anjum, R.; Mehmood, S.; Soban, M*., et al.* Dynamic Profiling of β-Coronavirus 3CL Mpro Protease Ligand-Binding Sites. *J. Chem. Inf. Model.* **2021,** *61*, 3058.
6. Shannon, A.; Le, N. T.-T.; Selisko, B.; Eydoux, C.; Alvarez, K*., et al.* Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antivir. Res.* **2020,** *178*, 104793.
7. Pillaiyar, T.; Meenakshisundaram, S.; Manickam, M. Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discov. Today* **2020,** *25*, 668.
8. Ullrich, S.; Nitsche, C. SARS-CoV-2 Papain-Like Protease: Structure, Function and Inhibition. *ChemBioChem* **2022,** *23*, e202200327.
9. Wang, R.; Hozumi, Y.; Yin, C.; Wei, G.-W. Decoding SARS-CoV-2 Transmission and Evolution and Ramifications for COVID-19 Diagnosis, Vaccine, and Medicine. *J. Chem. Inf. Model.* **2020,** *60*, 5853.
10. Yan, W.; Zheng, Y.; Zeng, X.; He, B.; Cheng, W. Structural biology of SARS-CoV-2: open the door for novel therapies. *Signal Transduct. Target. Ther.* **2022,** *7*, 26.
11. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T*., et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020,** *181*, 271.
12. Johansen-Leete, J.; Ullrich, S.; Fry, S. E.; Frkic, R.; Bedding, M. J*., et al.* Antiviral cyclic peptides targeting the main protease of SARS-CoV-2. *Chem. Sci.* **2022,** *13*, 3826.
13. Baggen, J.; Jacquemyn, M.; Persoons, L.; Vanstreels, E.; Pye, V. E*., et al.* TMEM106B is a receptor mediating ACE2-independent SARS-CoV-2 cell entry. *Cell* **2023,** *186*, 3427.
14. Wang, Y.; Liu, M.; Gao, J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 13967.
15. Jin, Y.; Ouyang, M.; Yu, T.; Zhuang, J.; Wang, W*., et al.* Genome-Wide Analysis of the Indispensable Role of Non-structural Proteins in the Replication of SARS-CoV-2. *Front. Microbiol.* **2022,** *13*, 907422.
16. Raj, R. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochem. Biophys. Rep.* **2021,** *25*, 100847.
17. Lubin, J. H.; Zardecki, C.; Dolan, E. M.; Lu, C.; Shen, Z*., et al.* Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. *Proteins* **2022,** *90*, 1054.
18. Stukalov, A.; Girault, V.; Grass, V.; Karayel, O.; Bergant, V*., et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **2021,** *594*, 246.
19. Alzyoud, L.; Ghattas, M. A.; Atatreh, N. Allosteric Binding Sites of the SARS-CoV-2 Main Protease: Potential Targets for Broad-Spectrum Anti-Coronavirus Agents. *Drug. Des. Devel. Ther.* **2022,** *16*, 2463.
20. Lim, C. T.; Tan, K. W.; Wu, M.; Ulferts, R.; Armstrong, L. A*., et al.* Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp3 papain-like protease. *Biochem. J.* **2021,** *478*, 2517.
21. Lv, Z.; Cano, K. E.; Jia, L.; Drag, M.; Huang, T. T*., et al.* Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development. *Front. Chem.* **2022,** *9*, 819165.
22. Yan, F.; Gao, F. An overview of potential inhibitors targeting non-structural proteins 3 (PLpro and Mac1) and 5 (3CLpro/Mpro) of SARS-CoV-2. *Comput. Struct. Biotechnol. J.* **2021,** *19*, 4868.
23. Moustaqil, M.; Ollivier, E.; Chiu, H. P.; Van Tol, S.; Rudolffi-Soto, P*., et al.* SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg. Microbes Infect.* **2021,** *10*, 178.
24. Nepali, K.; Sharma, R.; Sharma, S.; Thakur, A.; Liou, J.-P. Beyond the vaccines: a glance at the small molecule and peptide-based anti-COVID19 arsenal. *J. Biomed. Sci.* **2022,** *29*, 65.

25. Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.*, et al.* Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent. Sci.* **2020,** *6*, 315.
26. Cannalire, R.; Cerchia, C.; Beccari, A. R.; Di Leva, F. S.; Summa, V. Targeting SARS-CoV-2 Proteases and Polymerase for COVID-19 Treatment: State of the Art and Future Opportunities. *J. Med. Chem.* **2022,** *65*, 2716.
27. Riva, L.; Yuan, S.; Yin, X.; Martin-Sancho, L.; Matsunaga, N.*, et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **2020,** *586*, 113.
28. Redhead, M. A.; Owen, C. D.; Brewitz, L.; Collette, A. H.; Lukacik, P.*, et al.* Bispecific repurposed medicines targeting the viral and immunological arms of COVID-19. *Sci. Rep.* **2021,** *11*, 13208.
29. Kuo, C. J.; Chao, T. L.; Kao, H. C.; Tsai, Y. M.; Liu, Y. K.*, et al.* Kinetic Characterization and Inhibitor Screening for the Proteases Leading to Identification of Drugs against SARS-CoV-2. *Antimicrob. Agents Chemother.* **2021,** *65*, e02577.
30. Meyer, B.; Chiaravalli, J.; Gellenoncourt, S.; Brownridge, P.; Bryne, D. P.*, et al.* Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat. Commun.* **2021,** *12*, 5553.
31. Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021,** *9*, 622898.
32. Matthew, A. N.; Leidner, F.; Lockbaum, G. J.; Henes, M.; Zephyr, J.*, et al.* Drug Design Strategies to Avoid Resistance in Direct-Acting Antivirals and Beyond. *Chem. Rev.* **2021,** *121*, 3238.
33. Chang, K.-O.; Kim, Y.; Lovell, S.; Rathnayake, A. D.; Groutas, W. C. Antiviral Drug Discovery: Norovirus Proteases and Development of Inhibitors. *Viruses* **2019,** *11*, 197.
34. Anderson, J.; Schiffer, C.; Lee, S.-K.; Swanstrom, R., Viral Protease Inhibitors. In *Antiviral Strategies*, Kräusslich, H.-G.; Bartenschlager, R., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 85.
35. Gorbalenya, A. E.; Snijder, E. J. Viral cysteine proteinases. *Perspect. Drug Discov. Des.* **1996,** *6*, 64.
36. Verma, S.; Dixit, R.; Pandey, K. C. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Front. Pharmacol.* **2016,** *7*, 107.
37. Gehringer, M.; Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2019,** *62*, 5673.
38. Martin, J. S.; MacKenzie, C. J.; Fletcher, D.; Gilbert, I. H. Characterising covalent warhead reactivity. *Bioorg. Med. Chem.* **2019,** *27*, 2066.
39. Lonsdale, R.; Burgess, J.; Colclough, N.; Davies, N. L.; Lenz, E. M.*, et al.* Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *J. Chem. Inf. Model.* **2017,** *57*, 3124.
40. Petri, L.; Egyed, A.; Bajusz, D.; Imre, T.; Hetényi, A.*, et al.* An electrophilic warhead library for mapping the reactivity and accessibility of tractable cysteines in protein kinases. *Eur. J. Med. Chem.* **2020,** *207*, 112836.
41. Costa, A. M.; Bosch, L.; Petit, E.; Vilarrasa, J. Computational Study of the Addition of Methanethiol to 40+ Michael Acceptors as a Model for the Bioconjugation of Cysteines. *J. Org. Chem.* **2021,** *86*, 7107.
42. Cianni, L.; Feldmann, C. W.; Gilberg, E.; Gutschow, M.; Juliano, L.*, et al.* Can Cysteine Protease Cross-Class Inhibitors Achieve Selectivity? *J. Med. Chem.* **2019,** *62*, 10497.
43. Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **2011,** *10*, 307.
44. Sutanto, F.; Konstantinidou, M.; Dömling, A. Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **2020,** *11*, 876.
45. Boike, L.; Henning, N. J.; Nomura, D. K. Advances in covalent drug discovery. *Nat. Rev. Drug Discov.* **2022,** *21*, 881.
46. Bauer, R. A. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov. Today* **2015,** *20*, 1061.
47. Amaro, R. E.; Mulholland, A. J. Biomolecular Simulations in the Time of COVID19, and After. *Comput. Sci. Eng.* **2020,** *22*, 30.
48. Dommer, A.; Casalino, L.; Kearns, F.; Rosenfeld, M.; Wauer, N.*, et al.* #COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol. *Int. J. High Perform. Comput. Appl.* **2022,** *37*, 28.
49. Gao, K.; Wang, R.; Chen, J.; Cheng, L.; Frishcosy, J.*, et al.* Methodology-Centered Review of Molecular Modeling, Simulation, and Prediction of SARS-CoV-2. *Chem. Rev.* **2022,** *122*, 11287.

50. Ray, D.; Quijano, R. N.; Andricioaei, I. Point mutations in SARS-CoV-2 variants induce long-range dynamical perturbations in neutralizing antibodies. *Chem. Sci.* **2022,** *13*, 7224.
51. Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.*, et al.* Citizen Scientists Create an Exascale Computer to Combat COVID-19. *bioRxiv* **2020**, DOI: 10.1101/2020.06.27.175430.
52. Muratov, E. N.; Amaro, R.; Andrade, C. H.; Brown, N.; Ekins, S.*, et al.* A critical overview of computational approaches employed for COVID-19 drug discovery. *Chem. Soc. Rev.* **2021,** *50*, 9121.
53. Schuller, M.; Correy, G. J.; Gahbauer, S.; Fearon, D.; Wu, T.*, et al.* Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **2021,** *7*, eabf8711.
54. Achdout, H.; Aimon, A.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.*, et al.* COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning. *bioRxiv* **2020**, DOI: 10.1101/2020.10.29.339317.
55. Kneller, D. W.; Galanie, S.; Phillips, G.; O'Neill, H. M.; Coates, L.*, et al.* Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals. *Structure* **2020,** *28*, 1313.
56. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.*, et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
57. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.*, et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368*, 409.
58. Yadav, R.; Courouble, V. V.; Dey, S. K.; Harrison, J. J. E. K.; Timm, J.*, et al.* Biochemical and structural insights into SARS-CoV-2 polyprotein processing by Mpro. *Sci. Adv.* **2022,** *8*, eadd2191.
59. Biolatti, M.; Blangetti, M.; Baggieri, M.; Marchi, A.; Gioacchini, S.*, et al.* Strigolactones as Broad-Spectrum Antivirals against β-Coronaviruses through Targeting the Main Protease Mpro. *ACS Infect. Dis.* **2023,** *9*, 1310.
60. Milligan, J. C.; Zeisner, T. U.; Papageorgiou, G.; Joshi, D.; Soudy, C.*, et al.* Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp5 main protease. *Biochem. J.* **2021,** *478*, 2499.
61. MacDonald, E. A.; Frey, G.; Namchuk, M. N.; Harrison, S. C.; Hinshaw, S. M.*, et al.* Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **2021,** *7*, 2591.
62. Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967,** *27*, 157.
63. Berger, A.; Schechter, I. Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1970,** *257*, 249.
64. Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.*, et al.* Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J. Biol. Chem.* **2004,** *279*, 1637.
65. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003,** *300*, 1763.
66. Muramatsu, T.; Takemoto, C.; Kim, Y.-T.; Wang, H.; Nishii, W.*, et al.* SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113*, 12997.
67. Xia, B.; Kang, X. Activation and maturation of SARS-CoV main protease. *Protein Cell* **2011,** *2*, 282.
68. Noske, G. D.; Nakamura, A. M.; Gawriljuk, V. O.; Fernandes, R. S.; Lima, G. M. A.*, et al.* A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process. *J. Mol. Biol.* **2021,** *433*, 167118.
69. Chen, S.; Jonas, F.; Shen, C.; Hilgenfeld, R. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. *Protein Cell* **2010,** *1*, 59.
70. Pablos, I.; Machado, Y.; de Jesus, H. C. R.; Mohamud, Y.; Kappelhoff, R.*, et al.* Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CLpro substrate degradome. *Cell. Rep.* **2021,** *37*, 109892.
71. Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.*, et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020,** *583*, 459.
72. Hameedi, M. A.; T. Prates, E.; Garvin, M. R.; Mathews, I. I.; Amos, B. K.*, et al.* Structural and functional characterization of NEMO cleavage by SARS-CoV-2 3CLpro. *Nat. Commun.* **2022,** *13*, 5285.

73. Wenzel, J.; Lampe, J.; Müller-Fielitz, H.; Schuster, R.; Zille, M.*, et al.* The SARS-CoV-2 main protease Mpro causes microvascular brain pathology by cleaving NEMO in brain endothelial cells. *Nat. Neurosci.* **2021,** *24*, 1522.
74. Miczi, M.; Golda, M.; Kunkli, B.; Nagy, T.; Tőzsér, J.*, et al.* Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. *Int. J. Mol. Sci.* **2020,** *21*, 9523.
75. Zhong, N.; Zhang, S.; Zou, P.; Chen, J.; Kang, X.*, et al.* Without Its N-Finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can Form a Novel Dimer through Its C-Terminal Domain. *J. Virol.* **2008,** *82*, 4227.
76. Shi, J.; Sivaraman, J.; Song, J. Mechanism for Controlling the Dimer-Monomer Switch and Coupling Dimerization to Catalysis of the Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease. *J. Virol.* **2008,** *82*, 4620.
77. Goyal, B.; Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **2020,** *22*, 297.
78. Chen, S.; Chen, L.; Tan, J.; Chen, J.; Du, L.*, et al.* Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase N Terminus Is Indispensable for Proteolytic Activity but Not for Enzyme Dimerization. *J. Biol. Chem.* **2005,** *280*, 164.
79. El-Baba, T. J.; Lutomski, C. A.; Kantsadi, A. L.; Malla, T. R.; John, T.*, et al.* Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angew. Chem. Int. Ed.* **2020,** *59*, 23544.
80. Ansari, N.; Rizzi, V.; Carloni, P.; Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. *J. Am. Chem. Soc.* **2021,** *143*, 12930.
81. Grottesi, A.; Bešker, N.; Emerson, A.; Manelfi, C.; Beccari, A. R.*, et al.* Computational Studies of SARS-CoV-2 3CLpro: Insights from MD Simulations. *Int. J. Mol. Sci.* **2020,** *21*, 5346.
82. Jaffrelot-Inizan, T.; Celerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.*, et al.* High-Resolution Mining of SARS-CoV-2 Main Protease Conformational Space: Supercomputer-Driven Unsupervised Adaptive Sampling. *Chem. Sci.* **2021,** *12*, 4889.
83. Chen, H.; Wei, P.; Huang, C.; Tan, L.; Liu, Y.*, et al.* Only One Protomer Is Active in the Dimer of SARS 3C-like Proteinase. *J. Biol. Chem.* **2006,** *281*, 13894.
84. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.*, et al.* Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **2008,** *82*, 2515.
85. Kneller, D. W.; Zhang, Q.; Coates, L.; Louis, J. M.; Kovalevsky, A. Michaelis-like complex of SARS-CoV-2 main protease visualized by room-temperature X-ray crystallography. *IUCrJ* **2021,** *8*, 973.
86. Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.*, et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020,** *11*, 5877.
87. Shaqra, A. M.; Zvornicanin, S. N.; Huang, Q. Y. J.; Lockbaum, G. J.; Knapp, M.*, et al.* Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nat. Commun.* **2022,** *13*, 3556.
88. Zhao, Y.; Zhu, Y.; Liu, X.; Jin, Z.; Duan, Y.*, et al.* Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* **2022,** *119*, e2117142119.
89. Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.*, et al.* The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 13190.
90. Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L. 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism. *Biochemistry* **2004,** *43*, 4568.
91. Solowiej, J.; Thomson, J. A.; Ryan, K.; Luo, C.; He, M.*, et al.* Steady-State and Pre-Steady-State Kinetic Evaluation of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CLpro Cysteine Protease: Development of an Ion-Pair Model for Catalysis. *Biochemistry* **2008,** *47*, 2617.
92. Kneller, D. W.; Phillips, G.; Weiss, K. L.; Pant, S.; Zhang, Q.*, et al.* Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* **2020,** *295*, 17365.
93. Kneller, D. W.; Phillips, G.; Weiss, K. L.; Zhang, Q.; Coates, L.*, et al.* Direct Observation of Protonation State Modulation in SARS-CoV-2 Main Protease upon Inhibitor Binding with Neutron Crystallography. *J. Med. Chem.* **2021,** *64*, 4991.
94. Paasche, A.; Zipper, A.; Schafer, S.; Ziebuhr, J.; Schirmeister, T.*, et al.* Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry* **2014,** *53*, 5930.

95. Awoonor-Williams, E.; Abu-Saleh, A. A.-A. A. Covalent and non-covalent binding free energy calculations for peptidomimetic inhibitors of SARS-CoV-2 main protease. *Phys. Chem. Chem. Phys.* **2021,** *23*, 6746.
96. Pavlova, A.; Lynch, D. L.; Daidone, I.; Zanetti-Polzi, L.; Smith, M. D.*, et al.* Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *Chem. Sci.* **2021,** *12*, 1513.
97. Verma, N.; Henderson, J. A.; Shen, J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *J. Am. Chem. Soc.* **2020,** *142*, 21883.
98. Díaz, N.; Suárez, D. Influence of charge configuration on substrate binding to SARS-CoV-2 main protease. *Chem. Commun.* **2021,** *57*, 5314.
99. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
100. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
101. Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.*, et al.* SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2021,** *17*, 222.
102. Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.*, et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020,** *368*, 1331.
103. Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.*, et al.* Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020,** *30*, 678.
104. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
105. Khrenova, M. G.; Tsirelson, V. G.; Nemukhin, A. V. Dynamical properties of enzyme–substrate complexes disclose substrate specificity of the SARS-CoV-2 main protease as characterized by the electron density descriptors. *Phys. Chem. Chem. Phys.* **2020,** *22*, 19069.
106. Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem. Sci.* **2020,** *11*, 10626.
107. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020,** *10*, 12544.
108. La Monica, G.; Bono, A.; Lauria, A.; Martorana, A. Targeting SARS-CoV-2 Main Protease for Treatment of COVID-19: Covalent Inhibitors Structure–Activity Relationship Insights and Evolution Perspectives. *J. Med. Chem.* **2022,** *65*, 12500.
109. Thakur, A.; Sharma, G.; Badavath, V. N.; Jayaprakash, V.; Merz, K. M., Jr.*, et al.* Primer for Designing Main Protease (Mpro) Inhibitors of SARS-CoV-2. *J. Phys. Chem. Lett.* **2022,** *13*, 5776.
110. Bray, S.; Dudgeon, T.; Skyner, R.; Backofen, R.; Grüning, B.*, et al.* Galaxy workflows for fragment-based virtual screening: a case study on the SARS-CoV-2 main protease. *J. Cheminform.* **2022,** *14*, 22.
111. El Khoury, L.; Jing, Z.; Cuzzolin, A.; Deplano, A.; Loco, D.*, et al.* Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation. *Chem. Sci.* **2022,** *13*, 3674.
112. Ghahremanpour, M. M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J. A.; Zhang, C.-H.*, et al.* Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **2020,** *11*, 2526.
113. Macip, G.; Garcia-Segura, P.; Mestres-Truyol, J.; Saldivar-Espinoza, B.; Ojeda-Montes, M. J.*, et al.* Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Med. Res. Rev.* **2022,** *42*, 744.
114. Zev, S.; Raz, K.; Schwartz, R.; Tarabeh, R.; Gupta, P. K.*, et al.* Benchmarking the Ability of Common Docking Programs to Correctly Reproduce and Score Binding Modes in SARS-CoV-2 Protease Mpro. *J. Chem. Inf. Model.* **2021,** *61*, 2957.
115. Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.*, et al.* Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021,** *7*, 467.
116. Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.*, et al.* Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 27381.

117. Martí, S.; Arafet, K.; Lodola, A.; Mulholland, A. J.; Świderek, K., *et al.* Impact of Warhead Modulations on the Covalent Inhibition of SARS-CoV-2 Mpro Explored by QM/MM Simulations. *ACS Catal.* **2022,** *12*, 698.
118. Chenthamarakshan, V.; Hoffman, S. C.; Owen, C. D.; Lukacik, P.; Strain-Damerell, C., *et al.* Accelerating drug target inhibitor discovery with a deep generative foundation model. *Sci. Adv.* **2023,** *9*, eadg7865.
119. Nguyen, D. D.; Gao, K.; Chen, J.; Wang, R.; Wei, G.-W. Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning. *Chem. Sci.* **2020,** *11*, 12036.
120. Amporndanai, K.; Meng, X.; Shang, W.; Jin, Z.; Rogers, M., *et al.* Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives. *Nat. Commun.* **2021,** *12*, 3061.
121. Thun-Hohenstein, S. T. D.; Suits, T. F.; Malla, T. R.; Tumber, A.; Brewitz, L., *et al.* Structure-Activity Studies Reveal Scope for Optimisation of Ebselen-Type Inhibition of SARS-CoV-2 Main Protease. *ChemMedChem* **2021,** *17*, e202100582.
122. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016,** *59*, 6595.
123. Bai, B.; Belovodskiy, A.; Hena, M.; Kandadai, A. S.; Joyce, M. A., *et al.* Peptidomimetic α-Acyloxymethylketone Warheads with Six-Membered Lactam P1 Glutamine Mimic: SARS-CoV-2 3CL Protease Inhibition, Coronavirus Antiviral Activity, and in Vitro Biological Stability. *J. Med. Chem.* **2022,** *65*, 2905.
124. Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V., *et al.* Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021,** *12*, 1433.
125. Hatada, R.; Okuwaki, K.; Mochizuki, Y.; Handa, Y.; Fukuzawa, K., *et al.* Fragment Molecular Orbital Based Interaction Analyses on COVID-19 Main Protease − Inhibitor N3 Complex (PDB ID: 6LU7). *J. Chem. Inf. Model.* **2020,** *60*, 3593.
126. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design. *Chem. Sci.* **2021,** *12*, 3489.
127. Mondal, D.; Warshel, A. Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of α-Ketoamide Inhibitors of the Main Protease of SARS-CoV-2. *Biochemistry* **2020,** *59*, 4601.
128. Genovese, L.; Dawson, W.; Nakajima, T.; Cristiglio, V.; Vallet, V., *et al.* Protein–ligand interactions from a quantum fragmentation perspective: The case of the SARS-CoV-2 main protease interacting with α-ketoamide inhibitors. *J. Chem. Phys.* **2023,** *158*, 214121.
129. Stubbing, L. A.; Hubert, J. G.; Bell-Tyrer, J.; Hermant, Y. O.; Yang, S. H., *et al.* P1 Glutamine isosteres in the design of inhibitors of 3C/3CL protease of human viruses of the Pisoniviricetes class. *RSC Chem. Biol.* **2023,** *4*, 533.
130. Miura, T.; Malla, T. R.; Owen, C. D.; Tumber, A.; Brewitz, L., *et al.* In vitro selection of macrocyclic peptide inhibitors containing cyclic γ2,4-amino acids targeting the SARS-CoV-2 main protease. *Nat. Chem.* **2023,** *15*, 998.
131. Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q., *et al.* Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020,** *11*, 4417.
132. Oerlemans, R.; Ruiz-Moreno, A. J.; Cong, Y.; Dinesh Kumar, N.; Velasco-Velazquez, M. A., *et al.* Repurposing the HCV NS3–4A protease drug boceprevir as COVID-19 therapeutics. *RSC Med. Chem.* **2021,** *12*, 370.
133. Qiao, J.; Li, Y.-S.; Zeng, R.; Liu, F.-L.; Luo, R.-H., *et al.* SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021,** *371*, 1374.
134. Kneller, D. W.; Li, H.; Phillips, G.; Weiss, K. L.; Zhang, Q., *et al.* Covalent narlaprevir- and boceprevir-derived hybrid inhibitors of SARS-CoV-2 main protease. *Nat. Commun.* **2022,** *13*, 2268.
135. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M., *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374*, 1586.
136. US Food & Drug Administration Coronavirus (COVID-19) Update: FDA Authorizes First Oral Antiviral for Treatment of COVID-19. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19 (accessed 2021-12-22).
137. Zhao, Y.; Fang, C.; Zhang, Q.; Zhang, R.; Zhao, X., *et al.* Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein Cell* **2021,** *13*, 689.

138. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Computational simulations on the binding and reactivity of a nitrile inhibitor of the SARS-CoV-2 main protease. *Chem. Commun.* **2021,** *57*, 9096.
139. Wang, Y.-T.; Liao, J.-M.; Lin, W.-W.; Li, C.-C.; Huang, B.-C.*, et al.* Structural insights into Nirmatrelvir (PF-07321332)-3C-like SARS-CoV-2 protease complexation: a ligand Gaussian accelerated molecular dynamics study. *Phys. Chem. Chem. Phys.* **2022,** *24*, 22898.
140. Ngo, S. T.; Nguyen, T. H.; Tung, N. T.; Mai, B. K. Insights into the binding and covalent inhibition mechanism of PF-07321332 to SARS-CoV-2 Mpro. *RSC Adv.* **2022,** *12*, 3729.
141. Grigorenko, B. L.; Polyakov, I. V.; Khrenova, M. G.; Giudetti, G.; Faraji, S.*, et al.* Multiscale Simulations of the Covalent Inhibition of the SARS-CoV-2 Main Protease: Four Compounds and Three Reaction Mechanisms. *J. Am. Chem. Soc.* **2023,** *145*, 13204.
142. Drayman, N.; DeMarco, J. K.; Jones, K. A.; Azizi, S.-A.; Froggatt, H. M.*, et al.* Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science* **2021,** *373*, 931.
143. Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.*, et al.* Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022,** *65*, 6499.
144. Lin, M.; Zeng, X.; Duan, Y.; Yang, Z.; Ma, Y.*, et al.* Molecular mechanism of ensitrelvir inhibiting SARS-CoV-2 main protease and its variants. *Commun. Biol.* **2023,** *6*, 694.
145. Noske, G. D.; de Souza Silva, E.; de Godoy, M. O.; Dolci, I.; Fernandes, R. S.*, et al.* Structural basis of nirmatrelvir and ensitrelvir activity against naturally occurring polymorphisms of the SARS-CoV-2 main protease. *J. Biol. Chem.* **2023,** *299*, 103004.
146. Wei, P.; Fan, K.; Chen, H.; Ma, L.; Huang, C.*, et al.* The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochem. Biophys. Res. Commun.* **2006,** *339*, 865.
147. Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.*, et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020,** *11*, 5047.
148. Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.*, et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021,** *372*, 642.
149. DasGupta, D.; Chan, W. K. B.; Carlson, H. A. Computational Identification of Possible Allosteric Sites and Modulators of the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2022,** *62*, 618.
150. Bissaro, M.; Bolcato, G.; Pavan, M.; Bassani, D.; Sturlese, M.*, et al.* Inspecting the Mechanism of Fragment Hits Binding on SARS-CoV-2 M(pro) by Using Supervised Molecular Dynamics (SuMD) Simulations. *ChemMedChem* **2021,** *16*, 2075.
151. Sztain, T.; Amaro, R.; McCammon, J. A. Elucidation of Cryptic and Allosteric Pockets within the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2021,** *61*, 3495.
152. Strömich, L.; Wu, N.; Barahona, M.; Yaliraki, S. N. Allosteric Hotspots in the Main Protease of SARS-CoV-2. *J. Mol. Biol.* **2022,** *434*, 167748.
153. Dubanevics, I.; McLeish, T. C. B. Computational analysis of dynamic allostery and control in the SARS-CoV-2 main protease. *J. R. Soc. Interface* **2021,** *18*, 20200591.
154. Lee, J. T.; Yang, Q.; Gribenko, A.; Perrin, B. S.; Zhu, Y.*, et al.* Genetic Surveillance of SARS-CoV-2 Mpro Reveals High Sequence and Structural Conservation Prior to the Introduction of Protease Inhibitor Paxlovid. *mBio* **2022,** *13*, e00869.
155. Moghadasi, S. A.; Heilmann, E.; Khalil, A. M.; Nnabuife, C.; Kearns, F. L.*, et al.* Transmissible SARS-CoV-2 variants with resistance to clinical protease inhibitors. *Sci. Adv.* **2023,** *9*, eade8778.
156. Hu, Y.; Lewandowski, E. M.; Tan, H.; Zhang, X.; Morgan, R. T.*, et al.* Naturally Occurring Mutations of SARS-CoV-2 Main Protease Confer Drug Resistance to Nirmatrelvir. *ACS Cent. Sci.* **2023**, DOI: 10.1021/acscentsci.3c00538.
157. Clayton, J.; de Oliveira, V. M.; Ibrahim, M. F.; Sun, X.; Mahinthichaichan, P.*, et al.* Integrative Approach to Dissect the Drug Resistance Mechanism of the H172Y Mutation of SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2023,** *63*, 3521.
158. Flynn, J. M.; Samant, N.; Schneider-Nachum, G.; Barkan, D. T.; Yilmaz, N. K.*, et al.* Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms. *eLife* **2022,** *11*, e77433.
159. Flynn, J. M.; Huang, Q. Y. J.; Zvornicanin, S. N.; Schneider-Nachum, G.; Shaqra, A. M.*, et al.* Systematic Analyses of the Resistance Potential of Drugs Targeting SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **2023,** *9*, 1372.
160. Iketani, S.; Hong, S. J.; Sheng, J.; Bahari, F.; Culbertson, B.*, et al.* Functional map of SARS-CoV-2 3CL protease reveals tolerant and immutable sites. *Cell Host Microbe* **2022,** *30*, 1354.

161. Jochmans, D.; Liu, C.; Donckers, K.; Stoycheva, A.; Boland, S.*, et al.* The Substitutions L50F, E166A, and L167F in SARS-CoV-2 3CLpro Are Selected by a Protease Inhibitor In Vitro and Confer Resistance To Nirmatrelvir. *mBio* **2023,** *14*, e02815.
162. Zhou, Y.; Gammeltoft, K. A.; Ryberg, L. A.; Pham, L. V.; Tjørnelund, H. D.*, et al.* Nirmatrelvir-resistant SARS-CoV-2 variants with high fitness in an infectious cell culture system. *Sci. Adv.* **2022,** *8*, eadd7197.
163. Sasi, V. M.; Ullrich, S.; Ton, J.; Fry, S. E.; Johansen-Leete, J.*, et al.* Predicting Antiviral Resistance Mutations in SARS-CoV-2 Main Protease with Computational and Experimental Screening. *Biochemistry* **2022,** *61*, 2495.
164. Diessner, E. M.; Takahashi, G. R.; Cross, T. J.; Martin, R. W.; Butts, C. T. Mutation Effects on Structure and Dynamics: Adaptive Evolution of the SARS-CoV-2 Main Protease. *Biochemistry* **2023,** *62*, 747.
165. Armstrong, L. A.; Lange, S. M.; Dee Cesare, V.; Matthews, S. P.; Nirujogi, R. S.*, et al.* Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLOS ONE* **2021,** *16*, e0253364.
166. Báez-Santos, Y. M.; St. John, S. E.; Mesecar, A. D. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* **2015,** *115*, 21.
167. Tan, H.; Hu, Y.; Jadhav, P.; Tan, B.; Wang, J. Progress and Challenges in Targeting the SARS-CoV-2 Papain-like Protease. *J. Med. Chem.* **2022,** *65*, 7561.
168. Báez-Santos, Y. M.; Mielech, A. M.; Deng, X.; Baker, S.; Mesecar, A. D. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J. Virol.* **2014,** *88*, 12511.
169. Osipiuk, J.; Azizi, S.-A.; Dvorkin, S.; Endres, M.; Jedrzejczak, R.*, et al.* Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **2021,** *12*, 743.
170. Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.*, et al.* Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect. Dis.* **2020,** *6*, 2099.
171. Harcourt, B. H.; Jukneliene, D.; Kanjanahaluethai, A.; Bechill, J.; Severson, K. M.*, et al.* Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J. Virol.* **2004,** *78*, 13600.
172. Lee, H.; Lei, H.; Santarsiero, B. D.; Gatuz, J. L.; Cao, S.*, et al.* Inhibitor Recognition Specificity of MERS-CoV Papain-like Protease May Differ from That of SARS-CoV. *ACS Chem. Biol.* **2015,** *10*, 1456.
173. Rut, W.; Zmudzinski, M.; Snipas, S. J.; Bekes, M.; Huang, T. T.*, et al.* Engineered unnatural ubiquitin for optimal detection of deubiquitinating enzymes. *Chem. Sci.* **2020,** *11*, 6058.
174. Drag, M.; Mikolajczyk, J.; Bekes, M.; Reyes-Turcu, F. E.; Ellman, J. A.*, et al.* Positional-scanning fluorigenic substrate libraries reveal unexpected specificity determinants of DUBs (deubiquitinating enzymes). *Biochem. J.* **2008,** *415*, 367.
175. Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D.*, et al.* Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **2020,** *6*, eabd4596.
176. Mohamud, Y.; Xue, Y. C.; Liu, H.; Ng, C. S.; Bahreyni, A.*, et al.* The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem. Biophys. Res. Commun.* **2021,** *540*, 75.
177. Reynolds, N. D.; Aceves, N. M.; Liu, J. L.; Compton, J. R.; Leary, D. H.*, et al.* The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect. Dis.* **2021,** *7*, 1483.
178. Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.*, et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020,** *587*, 657.
179. Liu, G.; Lee, J.-H.; Parker, Z. M.; Acharya, D.; Chiang, J. J.*, et al.* ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat. Microbiol.* **2021,** *6*, 467.
180. Ratia, K.; Kilianski, A.; Baez-Santos, Y. M.; Baker, S. C.; Mesecar, A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. *PLOS Pathog.* **2014,** *10*, e1004113.
181. Békés, M.; van der Heden van Noort, Gerbrand J.; Ekkebus, R.; Ovaa, H.; Huang, Tony T.*, et al.* Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. *Mol. Cell* **2016,** *62*, 572.
182. Patchett, S.; Lv, Z.; Rut, W.; Békés, M.; Drag, M.*, et al.* A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. *Cell. Rep.* **2021,** *36*, 109754.

183. Wydorski, P. M.; Osipiuk, J.; Lanham, B. T.; Tesar, C.; Endres, M.*, et al.* Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. *Nat. Commun.* **2023,** *14*, 2366.
184. Zhao, Y.; Du, X.; Duan, Y.; Pan, X.; Sun, Y.*, et al.* High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. *Protein Cell* **2021,** *12*, 877.
185. Clasman, J. R.; Báez-Santos, Y. M.; Mettelman, R. C.; O'Brien, A.; Baker, S. C.*, et al.* X-ray Structure and Enzymatic Activity Profile of a Core Papain-like Protease of MERS Coronavirus with utility for structure-based drug design. *Sci. Rep.* **2017,** *7*, 40292.
186. Han, Y.-S.; Chang, G.-G.; Juo, C.-G.; Lee, H.-J.; Yeh, S.-H.*, et al.* Papain-Like Protease 2 (PLP2) from Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV): Expression, Purification, Characterization, and Inhibition. *Biochemistry* **2005,** *44*, 10349.
187. Gil-Moles, M.; Basu, U.; Büssing, R.; Hoffmeister, H.; Türck, S.*, et al.* Gold Metallodrugs to Target Coronavirus Proteins: Inhibitory Effects on the Spike-ACE2 Interaction and on PLpro Protease Activity by Auranofin and Gold Organometallics. *Chem. Eur. J.* **2020,** *26*, 15140.
188. Lin, M.-H.; Moses, D. C.; Hsieh, C.-H.; Cheng, S.-C.; Chen, Y.-H.*, et al.* Disulfiram can inhibit MERS and SARS coronavirus papain-like proteases via different modes. *Antivir. Res.* **2018,** *150*, 155.
189. Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.*, et al.* Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* **2020,** *39*, e106275.
190. Bosken, Y. K.; Cholko, T.; Lou, Y.-C.; Wu, K.-P.; Chang, C.-e. A. Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. *Front. Mol. Biosci.* **2020,** *7*, 174.
191. Leite, W. C.; Weiss, K. L.; Phillips, G.; Zhang, Q.; Qian, S.*, et al.* Conformational Dynamics in the Interaction of SARS-CoV-2 Papain-like Protease with Human Interferon-Stimulated Gene 15 Protein. *J. Phys. Chem. Lett.* **2021,** *12*, 5608.
192. Fu, Z.; Huang, B.; Tang, J.; Liu, S.; Liu, M.*, et al.* The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **2021,** *12*, 488.
193. Xu, Y.; Chen, K.; Pan, J.; Lei, Y.; Zhang, D.*, et al.* Repurposing clinically approved drugs for COVID-19 treatment targeting SARS-CoV-2 papain-like protease. *Int. J. Biol. Macromol.* **2021,** *188*, 137.
194. Ma, C.; Sacco, M. D.; Xia, Z.; Lambrinidis, G.; Townsend, J. A.*, et al.* Discovery of SARS-CoV-2 Papain-like Protease Inhibitors through a Combination of High-Throughput Screening and a FlipGFP-Based Reporter Assay. *ACS Cent. Sci.* **2021,** *7*, 1245.
195. Ma, C.; Wang, J. Validation and Invalidation of SARS-CoV-2 Papain-like Protease Inhibitors. *ACS Pharmacol. Transl. Sci.* **2022,** *5*, 102.
196. Brewitz, L.; Kamps, J.; Lukacik, P.; Strain-Damerell, C.; Zhao, Y.*, et al.* Mass spectrometric assays reveal discrepancies in inhibition profiles for the SARS-CoV-2 papain-like protease. *ChemMedChem* **2022,** *17*, e202200016.
197. Weglarz-Tomczak, E.; Tomczak, J. M.; Talma, M.; Burda-Grabowska, M.; Giurg, M.*, et al.* Identification of ebselen and its analogues as potent covalent inhibitors of papain-like protease from SARS-CoV-2. *Sci. Rep.* **2021,** *11*, 3640.
198. Ratia, K.; Pegan, S.; Takayama, J.; Sleeman, K.; Coughlin, M.*, et al.* A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 16119.
199. Ribaudo, G.; Yun, X.; Ongaro, A.; Oselladore, E.; Ng, J. P. L.*, et al.* Combining computational and experimental evidence on the activity of antimalarial drugs on papain-like protease of SARS-CoV-2: A repurposing study. *Chem. Biol. Drug Des.* **2023,** *101*, 809.
200. Shiraishi, Y.; Shimada, I. NMR Characterization of the Papain-like Protease from SARS-CoV-2 Identifies the Conformational Heterogeneity in Its Inhibitor-Binding Site. *J. Am. Chem. Soc.* **2023,** *145*, 16669.
201. Pitsillou, E.; Liang, J.; Ververis, K.; Hung, A.; Karagiannis, T. C. Interaction of small molecules with the SARS-CoV-2 papain-like protease: In silico studies and in vitro validation of protease activity inhibition using an enzymatic inhibition assay. *J. Mol. Graph. Model.* **2021,** *104*, 107851.
202. Sencanski, M.; Perovic, V.; Milicevic, J.; Todorovic, T.; Prodanovic, R.*, et al.* Identification of SARS-CoV-2 Papain-like Protease (PLpro) Inhibitors Using Combined Computational Approach. *ChemistryOpen* **2022,** *11*, e202100248.
203. Hu, H.; Wang, Q.; Su, H.; Shao, Q.; Zhao, W.*, et al.* Identification of Cysteine 270 as a Novel Site for Allosteric Modulators of SARS-CoV-2 Papain-Like Protease. *Angew. Chem. Int. Ed.* **2022,** *61*, e202212378.

204. van Vliet, V. J. E.; Huynh, N.; Palà, J.; Patel, A.; Singer, A.*, et al.* Ubiquitin variants potently inhibit SARS-CoV-2 PLpro and viral replication via a novel site distal to the protease active site. *PLOS Pathog.* **2022,** *18*, e1011065.
205. Santos, L. H.; Kronenberger, T.; Almeida, R. G.; Silva, E. B.; Rocha, R. E. O.*, et al.* Structure-Based Identification of Naphthoquinones and Derivatives as Novel Inhibitors of Main Protease Mpro and Papain-like Protease PLpro of SARS-CoV-2. *J. Chem. Inf. Model.* **2022,** *62*, 6553.
206. Di Sarno, V.; Lauro, G.; Musella, S.; Ciaglia, T.; Vestuto, V.*, et al.* Identification of a dual acting SARS-CoV-2 proteases inhibitor through in silico design and step-by-step biological characterization. *Eur. J. Med. Chem.* **2021,** *226*, 113863.
207. Elseginy, S. A.; Fayed, B.; Hamdy, R.; Mahrous, N.; Mostafa, A.*, et al.* Promising anti-SARS-CoV-2 drugs by effective dual targeting against the viral and host proteases. *Bioorg. Med. Chem. Lett.* **2021,** *43*, 128099.
208. Qayed, W. S.; Ferreira, R. S.; Silva, J. R. A. In Silico Study towards Repositioning of FDA-Approved Drug Candidates for Anticoronaviral Therapy: Molecular Docking, Molecular Dynamics and Binding Free Energy Calculations. *Molecules* **2022,** *27*, 5988.
209. Zhang, L.-c.; Zhao, H.-l.; Liu, J.; He, L.; Yu, R.-l.*, et al.* Design of SARS-CoV-2 Mpro, PLpro dual-target inhibitors based on deep reinforcement learning and virtual screening. *Future Med. Chem.* **2022,** *14*, 393.

2. Theoretical Background

2.1 Molecular mechanics

To describe the molecular interactions and chemical reactions in complex biomolecular systems such as M^{pro} and PL^{pro} (**Chapter 1**), an ideal approach would be to employ the Schrödinger Equation to describe the nuclei and electrons. However, the computational expense of solving the electronic Schrödinger Equation in such a system, including in an aqueous environment as in experimental conditions, is immense for a static structure. Moreover, the additional cost of sampling to obtain an equilibrium ensemble from which properties can be correlated with experimental observations is enormous.[1] Consequently, it becomes necessary to describe the potential energy (PE) of the system U in a simplified manner, such as according to Equation 2.1 (**Figure 2.1**). The functional form of Equation 2.1 and its constituent parameters are together referred to as a molecular mechanics (MM) force field (FF).[2] Commonly used MM FFs for biomolecules include the AMBER,[3] CHARMM,[4] OPLS,[5] and GROMOS families.[6] In atomistic FFs, every atom is treated as an elementary particle, whereas united atom FFs merge aliphatic hydrogens into their bonded carbons. In coarse-grained FFs, groups of atoms such as those within the same AA are treated together as one single bead. For the purpose of illustrating the components of a FFs, we use the atomistic AMBERFF99SB-ILDN FF ("SB" denoting Stony Brook University and "ILDN" denoting the improved dihedral parameters in isoleucine, leucine, aspartate, and asparagine).[7, 8]

$$\begin{aligned} U = & \frac{1}{2}\sum_{\text{bonds}} k_b\left(b - b_0\right)^2 + \frac{1}{2}\sum_{\text{angles}} k_\theta\left(\theta - \theta_0\right)^2 \\ & + \sum_{\text{dihedrals}} k_\phi\left(\cos n\phi - \phi_0\right)^2 \\ & + \sum_{i,j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\ & + \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \end{aligned} \tag{2.1}$$

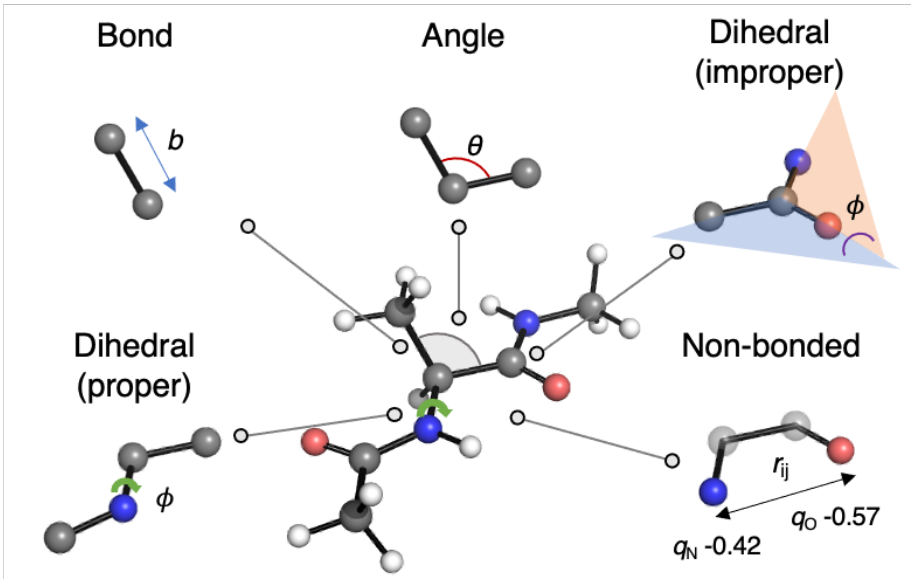


Figure 2.1: MM FF potential energy terms exemplified by an N-acetyl-Ala-NHMe dipeptide.

The first three terms correspond to bonded interactions, where covalently bonded atoms are described as balls connected by springs. The first and second terms correspond to harmonic stretching and bending of the distance b and angle θ in covalent bonds, respectively, away from the natural bond length b_0 and angle θ_0, with force constants k_b and k_θ. The third term involves a summation of periodic functions of the dihedral angle ϕ with force constant k_ϕ, multiplicity n, and phase ϕ_0, for describing "proper" dihedral angles involving four consecutively bonded atoms, as well as "improper" dihedral angles for maintaining planarity in functional groups such as amides.

The last two terms are the non-bonded interactions, that is van der Waals (vdW) interactions and Coulomb interactions, both of which depend on the distance r_{ij} between any two atoms i and j in the system. The exceptions are 1-2 and 1-3 atom pairs connected via covalent bonds, where all the vdW and Coulomb interactions are omitted, and 1-4 atom pairs, where the vdW and Coulomb interactions are scaled down (by factors of 0.5 and 0.8333, respectively, in AMBERFF99SB-ILDN).[8] The vdW interactions are typically described by the Lennard-Jones (LJ) potential, with the dispersion attraction varying as the theoretically valid r_{ij}^{-6}, and the short-term Pauli repulsion varying as r_{ij}^{-12}, selected to minimise computational expense by squaring the r_{ij}^{-6} term.[9] The

parameters of an atom pair, σ_{ij} and ε_{ij}, with respect to length (zero of the LJ curve) and energy (at the minimum of the LJ curve) dimensions, respectively, are derived from the σ and ε parameters of the two individual atom types by taking the arithmetic and geometric means respectively, according to the Lorentz-Berthelot combination rules.[10] Others FFs such as the OPLS All-Atom FF (OPLS-AA) use the geometric averaging of both σ and ε.[11] Coulomb interactions vary as r_{ij}^{-1} according to Coulomb's Law, with ε_0 and ε_r being the vacuum and relative permittivities, respectively. The sign and magnitude of Coulomb interactions depend on the two atomic charges q_i and q_j. In the aforementioned FFs, charges are pre-assigned and subsequently fixed; hence, electronic polarisation and multi-body effects are not explicitly modelled. In these pairwise-additive, non-polarisable FFs, removing a certain molecule in the system will not affect the interaction energies between the remaining molecules.[2]

As most biomolecular entities are normally considered to exist in aqueous environments, water molecules often account for a significant portion of the modelled systems. Due to its high polarity and extensive hydrogen bond (HB) networks, efficient and accurate modelling of water is a non-trivial task.[12] One popular model is the TIP3P water model developed in 1983, a three-site model in which oxygen and hydrogen are assigned charges of -0.834 and +0.417, respectively.[13] Additional interaction sites can be added on a water molecule to mimic the oxygen lone pairs as seen in the four-site TIP4P or five-site TIP5P model. While these models may lead to improved properties such as density, radial distribution function, and melting temperature, they also increase the number of charges and thus the number of Coulombic interaction calculations required.[9] Note that FF parametrisation is often performed with specific water models; for instance, the AMBER99FFSB-ILDN was parameterised with the TIP3P water model.[8]

While parameters for standard AAs, nucleic acids, solvent, and ions are often readily available, the modelling of novel entities such as non-standard AAs and ligands requires definition of bonded and non-bonded parameters. The general Amber FF (GAFF), developed to be compatible with AMBER FFs, enables efficient parametrisation of organic molecules.[14, 15] The assignment of

atomic charges in a novel molecule is performed based on the calculated electrostatic potential (ESP) around the molecule to reflect intermolecular Coulomb interactions. These charges are obtained for the molecule in vacuum at the HF/6-31G(d) level of theory, which is deliberately chosen to overestimate dipole moments by 10-15% to account for aqueous polarisation.[2] A restraint algorithm (RESP) is used to account for conformational dependence, uncertainty in assignments of buried atomic charges, and symmetry.[16, 17] Alternatively, AM1 (a semi-empirical quantum method) derived population charges with additive bond charge corrections (AM1-BCC) is a charge assignment method designated to emulate RESP charges.[18]

The explicit definition of bonded parameters in conventional MM FFs prohibits the breaking and formation of covalent bonds. This limitation can in part be overcome by methods such as the reactive FF (ReaxFF)[19] and the Morse potential-based empirical valence bond (EVB)[20, 21] methods, or by the use of hybrid quantum mechanics (QM) MM approaches (**Section 2.7**). Since X-ray crystallography is based on scattering of electron density, in protein structures protons are normally effectively invisible. Hence, the protonation states of titratable residues, especially of histidine residues, need to be pre-assigned, often based on HB network optimisation.[22] Similarly, differences between N and CH in histidines, and between O and NH_2 in glutamine and asparagine sidechains, are often not clearly resolved, and assignment of the conformational states of these residues requires careful inspection of potential interactions prior to subsequent modelling.[22]

2.2 Molecular dynamics

Following on from Equation 2.1, the force vector $\boldsymbol{F}$ acting on an atom can be calculated using the negative gradient of the PE, and be related to the acceleration $\boldsymbol{a}$ by dividing $\boldsymbol{F}$ by the atomic mass m according to Newton's second law (Equation 2.2):

$$\boldsymbol{F} = -\nabla U = -\left(\frac{\partial U}{\partial x}\vec{i} + \frac{\partial U}{\partial y}\vec{j} + \frac{\partial U}{\partial z}\vec{k}\right) = m\boldsymbol{a} \tag{2.2}$$

In a molecular dynamics (MD) simulation, integrations are performed over timesteps of Δt to update the velocity $\boldsymbol{v}$ and position $\boldsymbol{r}$ of every atom. In the default leap-frog integrator in the GROMACS MD program,[10] positions and velocities are updated at time points every t and $t + \Delta t/2$ (Equations 2.3 and 2.4), respectively.[9] Forces $\boldsymbol{F}$, velocities $\boldsymbol{v}$, and positions $\boldsymbol{r}$ are then repeatedly calculated over time to generate a trajectory.

$$\text{Velocity: } \boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right) = \boldsymbol{v}\left(t-\frac{1}{2}\Delta t\right) + \boldsymbol{a}\Delta t \tag{2.3}$$

$$\text{Position: } \boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right)\Delta t \tag{2.4}$$

For accurate numerical integrations, the timestep Δt should reflect the shortest timescale of motion. This is often the stretching motion, or vibration, of a covalent bond involving a hydrogen atom. A guideline is that a minimum of five numerical integration steps should be performed per period of a harmonic oscillation.[10] For instance, consider a hydroxyl OH group (atom types OH and HO) described by AMBER99FFSB-ILDN,[8] where k_b = 462750.4 kJ mol^{-1} nm^{-2} = 768.4331 N m^{-1} and the reduced mass μ = 0.9483 amu = 1.575 × 10^{-27} kg:

$$\text{Oscillation Period} = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{\mu}{k_b}} = 9 \text{ fs} \tag{2.5}$$

As a result, the timestep is usually limited to 1 femtosecond (fs). However, considering that most (biologically relevant) timescales of interest are often in nanoseconds (ns) or beyond, many force evaluation and numerical integration steps would be required for adequate sampling. To enable the use of a larger timestep, bonds involving hydrogen atoms, except for proton transfer processes, are often constrained to a particular length, using algorithms such as the linear constraint solver (LINCS). In LINCS, the bond undergoes an unconstrained update, a force projection along the bond, and correction for rotational lengthening.[23] Use of such constraints enables the timestep to be doubled to 2 fs.

In MD simulations, periodic boundary conditions (PBC) are used to represent experimental conditions and remove boundary effects, where molecules near the boundaries would experience a reduced number of intermolecular interactions than those in the middle, and without suitable restraints, escape into vacuum space, when in reality they are in a bulk solution environment. PBC eliminate these effects by infinitely repeating the simulation box in three dimensions, such that one molecule leaving the box at the boundary would re-enter in the opposite face. The box size is chosen such that the protein does not interact with its nearest periodic image in the short range given by the cut-off used for non-covalent interactions;[24] a box edge defined to be ≥1 nm away from the outermost protein atom is usually sufficient, such that any protein atom is at least ≥2 nm from a protein atom in the neighbouring image.[25]

Given the size and periodicity of the system, the number of calculations required for non-bonded interactions is significant and scales as N^2, where N is the number of atoms.[26] Considering the dependence of the non-bonded interactions on r_{ij}, the LJ repulsive r_{ij}^{-12} term is only significant in the very short length scale ~σ_{ij}, which is typically up to a few Å, and the r_{ij}^{-6} attraction also decays quickly with r_{ij}, being insignificant beyond ~1 nm. Hence, LJ interactions are often only evaluated for atom pairs within ~1 nm cut-off radius.[25] By contrast, the Coulomb interactions decay more slowly (r_{ij}^{-1}) and their sign depends on the relative signs of q. To overcome this, Coulomb interactions are evaluated in two different ways: short-range interactions (< ~1 nm) are directly evaluated in real space, while longer-range interactions are more efficiently evaluated in the reciprocal, Fourier space using Ewald summation.[27] The smooth particle-mesh Ewald (PME) method, where charges are smeared on a grid via interpolation, leads to a more efficient $N \log(N)$ scaling and has enabled the efficient MD sampling of large biomolecules.[28] PBC can be applied for any triclinic box, but the number of solvent molecules required can be minimised by using shapes of higher symmetry, such as cube, octahedron, or rhombic dodecahedron, with the latter being the most efficient.[29]

To initiate an MD simulation starting from a static (crystal) structure, velocities are first generated, with a random number generator, according to the Maxwell-Boltzmann velocity distribution (MB) at a designated temperature (T). Replica simulations can be initiated via different initial velocities.[30] To mimic experimental conditions where the sample can exchange heat with the surroundings to maintain constant T, a thermostat is applied to simulations. The velocity-rescaling thermostat regularly rescales atomic velocities according to MB, with first order decay of deviations, and an extra stochastic term ensures correct equipartition kinetic energy (KE) distribution.[31] Another popular thermostatting method is the use of stochastic dynamics according to the Langevin equation (Equation 2.6), where an additional friction term (second term on the right-hand side) and a random force term $\boldsymbol{R}$ dependent on temperature ($\boldsymbol{R} \propto T^{1/2}$) generates the correct KE ensemble.[26] A simulation box of constant volume (V) and of a fixed number of atoms (N) coupled to a thermostat at constant temperature (T) is referred to as a canonical (NVT) ensemble.

$$m\boldsymbol{a} = -\nabla U - m\gamma\boldsymbol{v} + \boldsymbol{R} \tag{2.6}$$

Experiments are often conducted in constant pressure (P) instead of constant volume. To sample the isothermal-isobaric NPT ensemble, a barostat is applied, such as the Berendsen or Parrinello-Rahman pressure coupling scheme.[32-34] The box size is allowed to vary over the course of simulation.

Following initiation by velocity generation, the system is firstly equilibrated under NVT conditions and then under NPT conditions. These equilibration phases enable the system to reach the correct temperature and pressure conditions and be fully solvated at the correct density. The structure is also relaxed from any crystallographic biases or artefacts from resolution limitations, symmetry constraints, or other experimental conditions, or from post-processing of the crystal structure such as the modelling of unresolved loops or atoms.[35]

2.3 Umbrella sampling

MD is an ergodic method, meaning that all configurations on the free energy surface (FES) spanned by the phase space of the modelled system will eventually be sampled over time, and that phase space averages can be obtained by time averages over the trajectory.[24] However, the vast number of degrees of freedom in biomolecules and the presence of kinetic barriers separating local energy minima often make full phase space sampling unattainable. For instance, a typical kinetic barrier of ~10 kcal mol^{-1}, seen in structural transitions between different functional states of biomolecules,[36] would have a half-life of >2 μs, corresponding to >1 billion timesteps of 2 fs. Enhanced sampling techniques aim to accelerate FES sampling by increasing the probability of sampling rare transition events over these kinetic barriers.

One such technique is umbrella sampling (US), where a collective variable (CV) that relates to the process of interest, *e.g.*, the protein-ligand centre of mass (COM) distance during ligand dissociation, is defined. Intermediate d_i values are chosen between the end states representing the complexed and the fully dissociated species. For each d_i value, the system is sampled at a biased potential U', which is the unbiased potential U summed with a biasing potential W harmonically restraining d to d_i (Equation 2.7) with force constant k_d. A different k_d value can be chosen for each window.

$$\text{Biased potential: } U' = U + W = U + \frac{1}{2} k_d \left(d - d_i \right)^2 \tag{2.7}$$

To obtain the unbiased FES, the weighted histogram analysis method (WHAM) is popularly employed.[37] In brief, the observed d values in each sampling window are arranged into bins to generate histograms, and the statistical errors in the probability distributions are minimised. This process yields the potential of mean force (PMF), which describes the free energy (G) variation along the CV path. Accurate PMFs rely on the choice of CV(s) that are relevant to the transition of interest, and suitable force constant values for adequate sampling overlap between windows.[26]

2.4 Dynamical nonequilibrium molecular dynamics

The MD and enhanced sampling methods discussed so far relate to the sampling of configurations in the equilibrium ensemble, either at the unbiased or biased potential, with the observed properties being time independent due to ergodicity. MD can also be employed to measure dynamical properties outside of equilibrium, such as the dynamical response of a system to a change or perturbation.[26] Such a perturbation may not be spontaneously observed over the duration of conventional equilibrium MD simulations, but can be introduced manually.[38] One approach for measuring dynamical response is dynamical nonequilibrium MD (D-NEMD), which was initially introduced by Ciccotti and Jacucci (1975) to measure the mobility of a charged particle in a LJ liquid.[39]

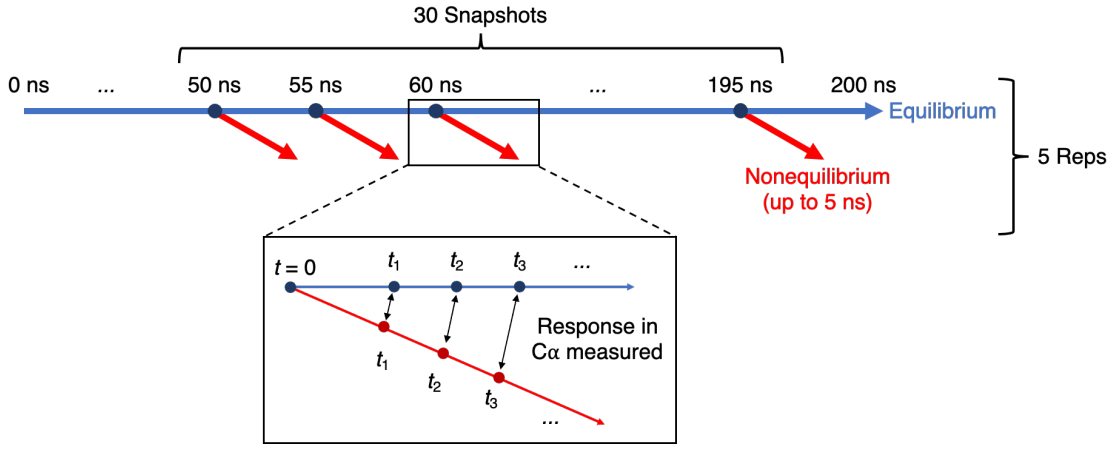


Figure 2.2: Scheme of the procedures used to set up and analyse D-NEMD simulations. Reprinted with permission from the Supporting Information of **Manuscript II**: *JACS Au* **2023,** *3*, 1767. Copyright 2023 American Chemical Society.[40]

In D-NEMD (**Figure 2.2**), initially long MD simulations are performed to sample equilibrium configurations. These configurations are then used to initiate nonequilibrium trajectories by introducing the perturbation. Different perturbations have been reported, such as the application of external electric fields[39, 41] and the removal of ligands from their protein binding site.[38] Subsequently, a subtraction between the nonequilibrium and equilibrium configurations is

performed between equivalent time points following the perturbation. Such a subtraction technique enables the cancellation of random thermal fluctuations in the trajectories, and the difference thus represents the systematic response to the perturbation.[41] According to the Kubo-Onsager relation, the nonequilibrium average response can be obtained from the expected value of the time-evolved response averaged over the initial equilibrium ensemble.[38, 42] Hence, by averaging the response over many (>50) trajectory pairs, the response induced by the perturbation can be obtained, with assessment of its statistical significance.

2.5 Binding energy calculations

Computational chemistry is often used in drug discovery campaigns to predict how strongly a (potential) ligand binds to a protein target. While visual inspection and analysis of protein-ligand properties or interactions can sometimes provide useful information, calculation of the binding free energy (ΔG_{bind}) is often pivotal. ΔG_{bind} corresponds to the free energy difference between the solvated protein-ligand complex, and the isolated systems of the solvated protein and the solvated ligand (**Figure 2.3**).

Reliable calculations of ΔG_{bind} from direct simulations of ligand association or dissociation are challenging, partly due to the difficulties in defining the fully isolated, non-interacting state. Since G is a state function, this difficulty can be circumvented by constructing intermediate states to compute ΔG_{bind}, as long as the processes complete the thermodynamic cycle (**Figure 2.3**). Various methods at different levels of trade-off between accuracy and speed exist, ranging from fast estimations by scoring functions during docking (**Section 2.6**), to molecular mechanics/Poisson Boltzmann or generalised Born surface area (MM/PBSA or MM/GBSA) methods at intermediate computational expense and accuracy, and to the gold standard determination of absolute binding free energy (ABFE) using free energy perturbation (FEP) methods.[43]

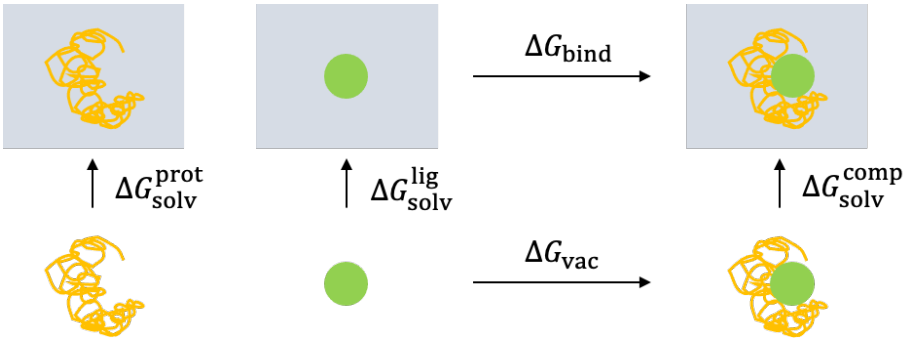


Figure 2.3: Thermodynamic cycle for ΔG_{bind} calculation using the MM/PBSA or MM/GBSA method.

The MM/PBSA and MM/GBSA methods consider the protein (prot), ligand (lig), and complex (comp) in vacuum as intermediate states (**Figure 2.3**).[44-47] ΔG_{bind} is the sum of the binding free energy in vacuum (ΔG_{vac}) and the solvation energy (ΔG_{solv}) of the complex, minus the solvation energies of the isolated protein and ligand (Equation 2.8).

$$\Delta G_{bind} = \Delta G_{vac} + \Delta G_{solv}^{comp} - \Delta G_{solv}^{prot} - \Delta G_{solv}^{lig} \tag{2.8}$$

ΔG_{vac} can be split into the MM energy difference (ΔU_{MM}), which consists of differences in the bonded, Coulomb, and vdW terms, and an entropy term ($-T\Delta S$) (Equation 2.9).

$$\Delta G_{vac} = \Delta U_{MM} - T\Delta S = \left(\Delta U_{bonded} + \Delta U_{Coulomb} + \Delta U_{vdW}\right) - T\Delta S \tag{2.9}$$

Each ΔG_{solv} term can be split into a non-polar part, which has been found to be a linear function of the surface accessible surface area (SASA) (A) with surface tension coefficient γ and constant c, and a polar part, which is calculated by either the Poisson Boltzmann (PB) or generalised Born (GB) equation (Equation 2.10).

$$\Delta G_{solv} = \Delta G_{polar} + \Delta G_{nonpolar} = \Delta G_{PB\ or\ GB} + (\gamma A + c) \tag{2.10}$$

The PB equation, either in full (Equation 2.11) or linearised by assuming low ionic strength and electric field (Equation 2.12), is solved numerically, for the electrostatic potential Φ given the

dielectric (ε) and inverse Debye length (κ).[48] The GB approach (Equation 2.13) considers atoms as spheres with a lower ε in its "internal" immediate surroundings (int) than in the bulk solvent (sol), depending on the effective Born radius α.[47] The GB equation offers greater simplicity compared to PB but sacrifices some accuracy.[46, 47]

$$\text{PB:}\ -\nabla\cdot\varepsilon(\boldsymbol{r})\,\nabla\Phi(\boldsymbol{r})+\kappa^2\sinh\phi(\boldsymbol{r})=4\pi\rho(\boldsymbol{r}) \tag{2.11}$$

$$\text{Linearised PB:}\ -\nabla\cdot\varepsilon(\boldsymbol{r})\,\nabla\Phi(\boldsymbol{r})+\kappa^2\phi(\boldsymbol{r})=4\pi\rho(\boldsymbol{r}) \tag{2.12}$$

$$\text{GB:}\ \Delta G_{\text{GB}}=-\left(\frac{1}{\varepsilon_{\text{int}}}-\frac{1}{\varepsilon_{\text{sol}}}\right)\sum_{\text{i,j}}\frac{q_{\text{i}}q_{\text{j}}}{f_{\text{GB}}},\ \text{where}\ f_{\text{GB}}=\sqrt{r_{\text{ij}}^{\,2}+\alpha_{\text{ij}}^{\,2}e^{\frac{r_{\text{ij}}^{\,2}}{4\alpha_{\text{ij}}^{\,2}}}} \tag{2.13}$$

In MM/PBSA or MM/GBSA calculations, configurations of the protein, ligand, and complex can be obtained from separate trajectories of the three species, or alternatively, from a single trajectory of the solvated complex, assuming that the protein and ligand conformations in their isolated solvated forms are similar to those when complexed. The methods enable one to decompose the binding energy in terms of per residue or per residue pair contributions, allowing binding hotspots to be identified.[49, 50] However, the results are dependent on the definition of parameters, such as the Born radii set and dielectric constant values. The combination of mbondi2 radii and parameters by Onufriev, Bashford, and Case (2004) is a popular choice for GB calculations.[51] The entropy contribution ($-T\Delta S$) can be calculated by normal mode analysis, but due to the significant expense of diagonalising a $3N \times 3N$ matrix, this term is often omitted,[46] particularly when comparing similar ligands, with likely comparable ΔS values.[47]

A more accurate approach to evaluating ΔG_{bind} involves unphysical, alchemical intermediate states, in which interactions between the solvated protein and its bound ligand are partially or completely deactivated (**Figure 2.4**).[52] This approach, known as FEP, is commonly used as the gold standard for ABFE calculations.[53] The thermodynamic cycle for ΔG_{bind} comprises five processes: (i) turning off the interactions between the isolated ligand and the solvent ($-\Delta G_{\text{lig-sol}}$);

(ii) restraining the non-interacting ligand in solvent ($\Delta G_{rest(sol)}$); (iii) transferring the non-interacting, restrained ligand to its protein-bound location, which involves zero ΔG since the ligand is not interacting with its surroundings; (iv) restoring the interactions between the restrained ligand and the protein ($-\Delta G_{decoup}$); and (v) releasing the ligand restraints in the complexed state ($-\Delta G_{rest(comp)}$) (Equation 2.14).[53] In the non-interacting intermediate states, the ligand may still retain its intramolecular interactions in a "decoupled" manner, or the intramolecular interactions can be turned off to result in an "annihilated" state.[54]

$$\Delta G_{bind} = -\Delta G_{lig-sol} + \Delta G_{rest(sol)} + 0 - \Delta G_{decoup} - \Delta G_{rest(comp)} \quad (2.14)$$

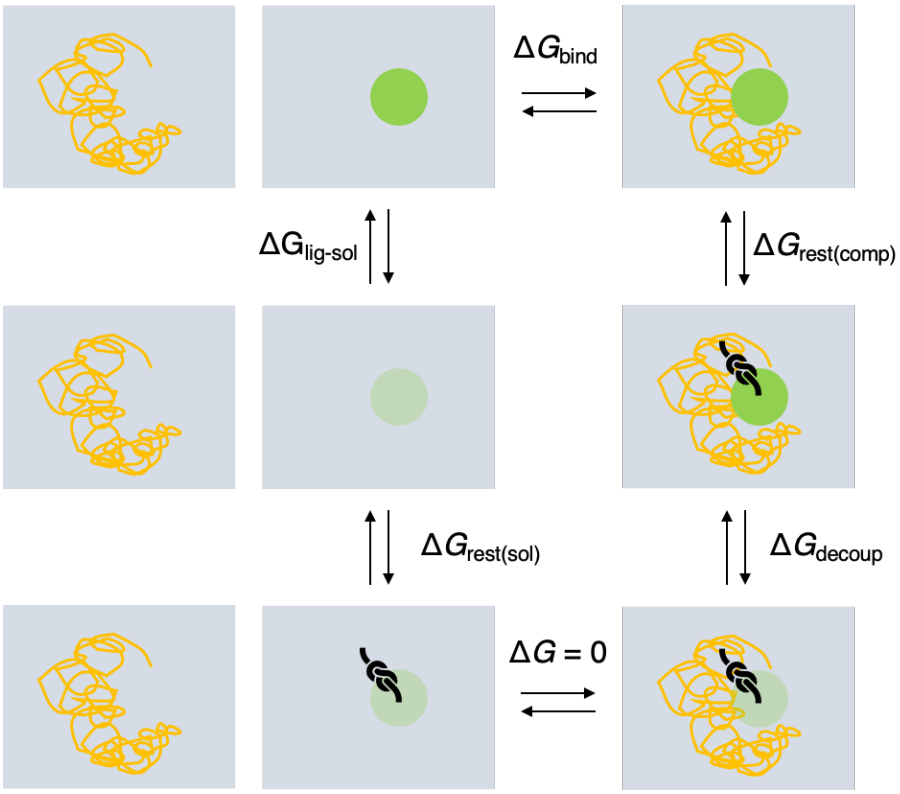


Figure 2.4: Thermodynamic cycle for ΔG_{bind} calculation (ABFE) via alchemical intermediates.

To minimise the sampling problem when the ligand is not interacting with the protein, harmonic restraints are applied on one distance, two angles, and three dihedrals between the protein and the ligand. This approach prevents the non-interacting ligand from drifting freely in the simulation box, in which case extensive sampling of every possible position and orientation of the ligand

relative to the protein would be required to achieve convergence.[54] The $\Delta G_{rest(sol)}$ term in process (ii) can be calculated analytically, as reported by Boresch *et al.* (2003).[55]

Other than processes (ii) and (iii), MD simulations are employed for processes (i), (iv), and (v), with each simulation having a λ mixing parameter between the $\lambda = 0$ and $\lambda = 1$ end states. The ΔG value can then be determined from the calculated energy differences between the λ-states (ΔU) via WHAM, exponential averaging (EXP, also referred to as FEP), or (multistate) Bennett acceptance ratio (BAR or MBAR), or from the derivative of U with respect to λ via thermodynamic integration (TI).[56]

The reliability of calculated ΔG values depends on the quality of sampling in each λ window and overlap between sampling windows. Hence, multiple windows between $\lambda = 0$ and $\lambda = 1$ are required, contributing to the computational cost of FEP calculations. Unlike MM/PBSA or MM/GBSA, where only the end points of bound and unbound species are considered, FEP requires the computation of the full set of intermediate λ-windows along the transformation pathways. Hence, FEP is often only used in the lead optimisation stage of drug discovery for few selected compounds, or for accurate determination of binding pose when this is unknown. In lead optimisation, FEP is often used to predict the $\Delta\Delta G_{bind}$ when a local modification is made on the ligand. In these cases, λ is directly varied from 0 in the original ligand structure to 1 in the new ligand in the complex, in a relative binding free energy (RBFE) calculation.[57]

2.6 Docking

Binding energy calculations serve as valuable tools for comparing binding affinities of different ligands or of the same ligand in distinct binding modes However, their utility might be limited when the binding mode is itself unknown. Docking, the process of aligning two molecules optimally relative to each other,[9] is commonly employed to predict binding modes. Docking methodologies vary based on the treatment of flexibility of molecules. In rigid docking, both

protein and ligand are considered as rigid bodies, while in semi-flexible docking, the ligand conformation is flexible. In flexible docking, protein movements are also allowed.[58] These approaches rely on efficient methods to explore possible bound conformations, and to score them according to estimated binding affinities.[59]

Sampling of binding modes can be performed systematically, such as by an exhaustive search over all rotatable bonds in the ligand, or by generating an ensemble of conformational isomers before attempting to fit each of them into the target binding site. Alternatively, stochastic methods such as genetic algorithm (GA) and Monte Carlo (MC) can be used. In GA, bound conformations referred to as "genes" are allowed to mix between them, such as by the swapping of dihedrals, and to mutate by random dihedral modifications, to generate new "children" structures. The new structures are scored and ranked, with the lower-scoring ones being eliminated. Repeated rounds result in the survival of the "fittest", best-scoring conformations.[9] In MC, a predicted bound conformation is subjected to a random movement. If this improves the binding score, the new pose is accepted. Otherwise, the pose is accepted or rejected with a probability based on the score difference, thus enabling conformational sampling.

The energy equation used to score these generated poses is known as the scoring function. For example, the AutoDock4 (AD4) scoring function mimics the AMBER FF, but integrates additional terms relating to HB, desolvation, and entropy loss based on number of rotatable torsions N_{tors} (Equations 2.15-2.17), each weighted by a coefficient w.[60-62]

$$\Delta G = \Delta U^{\text{lig(intramolecular)}}_{\text{bound-unbound}} + \Delta U^{\text{prot(intramolecular)}}_{\text{bound-unbound}} + U^{\text{comp}}_{\text{bound}} + \Delta S_{\text{conf}} \tag{2.15}$$

$$U = w_{\text{vdW}} \sum_{\text{i,j}} \left(\frac{A_{\text{ij}}}{r_{\text{ij}}^{12}} - \frac{B_{\text{ij}}}{r_{\text{ij}}^{6}} \right) + w_{\text{HB}} \sum_{\text{i,j}} E(\theta) \left(\frac{C_{\text{ij}}}{r_{\text{ij}}^{12}} - \frac{D_{\text{ij}}}{r_{\text{ij}}^{10}} \right) + w_{\text{elec}} \sum_{\text{i,j}} \frac{q_{\text{i}} q_{\text{j}}}{\varepsilon(r_{\text{ij}}) r_{\text{ij}}} + w_{\text{solv}} \sum_{\text{i,j}} \left(S_{\text{i}} V_{\text{j}} + S_{\text{j}} V_{\text{i}} \right) e^{-\frac{r_{\text{ij}}^2}{2\sigma^2}} \tag{2.16}$$

$$\Delta S_{\text{conf}} = w_{\text{conf}} N_{\text{tors}} \tag{2.17}$$

Compared to docking of small molecule ligands, docking of peptides to proteins is more challenging due to the vast degrees of freedom in the peptide, including both the overall peptide fold and the conformation of each AA sidechain. AutoDock CrankPep (ADCP) is a recently developed peptide docking program.[63] Based on the AutoDock affinity grid approach, it utilises a MC conformational search via crankshaft motion along two randomly selected peptide Cα atoms and the sidechain rotamer library compiled by Dunbrack Jr. and Cohen (1997),[64] with scoring dependent on peptide conformation and protein-peptide interactions.

After obtaining multiple highest-scoring poses from the sampling algorithm, clustering is performed to group similarly binding poses together. Clustering is often based on ligand root mean square deviation (RMSD). For larger entities such as peptides, overall RMSD may be skewed by conformational deviations in more flexible regions that form few or no interactions with the protein. Native contacts, defined in ADCP as a list of protein-peptide AA pairs within 5 Å of each other,[63] can be used for clustering instead, with greater emphasis on the binding in the peptide regions that form more interactions with the protein.

2.7 Quantum mechanics/molecular mechanics

In the classical MM potential (Equation 2.1), the definition of bonded parameters implies the pre-definition of covalent bond connectivities prior to MD simulations. In most enzyme catalysed reactions, covalent bonds are formed and broken. One method to overcome this limitation is the use of hybrid quantum mechanics/molecular mechanics (QM/MM) methods, pioneered by Warshel, Levitt, and Karplus,[65, 66] and recognised with the Nobel Prize in Chemistry in 2013 "for the development of multiscale models for complex chemical systems".[67] In QM/MM, the reactive part of the system is treated by QM methods, such as with semi-empirical methods, Hartree-Fock (HF) or other *ab initio* methods, or density functional theory (DFT), all of which enable bond formation and breaking; these QM-treated atoms are referred to as the QM region. The

surroundings, which often include regions of the enzyme not directly involved in reaction and the bulk solvent, are treated by a classical MM FF and referred to as the MM region.

In most cases, the partitioning of atoms into the QM or MM region is pre-defined and fixed, although in some cases an adaptive partitioning may be necessary, such as in the case when reactive and bulk waters exchange.[68] Since the enzyme environment may significantly impact transitional state (TS) stabilisation, accurate treatment of the coupling between the QM and MM regions is crucial, with both subtractive and additive schemes being developed.[65] In the subtractive scheme, the full system PE is calculated by summing the MM energy of the full system and the QM energy of the QM region, and taking away the MM energy of the QM region (Equation 2.18), as in the ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) method.[69] While this scheme removes the need for explicit QM-MM coupling terms, the coupling is entirely treated by MM, thus relying on accurate MM parametrisation of the reactive species.[68]

$$U(\text{System}) = U_{\text{MM}}(\text{System}) + U_{\text{QM}}(\text{QM region}) - U_{\text{MM}}(\text{QM region}) \tag{2.18}$$

By contrast, the additive approach calculates the full system PE by summing the QM energy of the QM region, the MM energy of the MM region, and the QM-MM coupling (Equation 2.19). The coupling term involves non-bonded vdW and Coulomb interactions, and bonded terms if the QM/MM boundary cuts across covalent bonds, each of which is discussed below.

$$U(\text{System}) = U_{\text{QM}}(\text{QM region}) + U_{\text{MM}}(\text{MM region}) + U(\text{QM-MM}) \tag{2.19}$$

The vdW coupling is usually calculated at the MM level using the LJ equation, and thus requires definitions of σ and ε for atoms in the QM region. While this might be difficult in the case of reactions in which the nature of the QM atom changes significantly, such as in the fluorinase-catalysed S_N2 attack of the fluoride ion on *S*-adenosyl methionine (SAM) to a carbon-bonded

fluorine, the short-ranged nature of LJ interactions limits the impact of the LJ parameters on the FES, if the QM/MM boundary is placed further away from such atoms.[70-72]

By contrast, the treatment of long-ranged electrostatics across the QM/MM boundary requires more careful attention, and is characterised by different embedding schemes.[71] In mechanical embedding, Coulomb interactions are treated in an MM manner, and the QM electron density is not polarised by the MM region. In electrostatic embedding, electrostatic interactions are treated with QM, by including the MM environment as point charges in the calculation of the QM energy. Polarised embedding further accounts for the polarisation of the MM region by charges of the QM atoms, and thus requires the use of polarisable MM FFs. Electrostatic embedding currently remains the most widely used method.[68, 71]

If the QM/MM boundary cuts across covalent bonds, bonded interactions can be treated by classical MM bond, angle, and dihedral terms. To satisfy the incomplete valence in the QM region, a frozen set of hybrid orbitals can be placed on the QM atom at the QM/MM boundary.[73] Alternatively, a link atom, typically modelled as a hydrogen, can be added, with its position constrained depending on the QM and MM atoms involved in the cut covalent bond.[71, 73] To minimise artefacts in the QM/MM electrostatics interactions, cuts are usually only performed across non-polar carbon-carbon bonds.[72] In enzyme-catalysed reactions where AA sidechains are involved in reaction, such as the cysteine and histidine residues in M^{pro} and PL^{pro}, the Cα—Cβ bond presents a reasonable point for defining the QM/MM boundary.

2.8 References

1. Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.*, et al.* Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *WIREs Comput. Mol. Sci.* **2019,** *9*, e1393.
2. Lemkul, J. A., Chapter One - Pairwise-additive and polarizable atomistic force fields for molecular dynamics simulations of proteins. In *Progress in Molecular Biology and Translational Science*, Strodel, B.; Barz, B., Eds. Academic Press: 2020; Vol. 170, pp 1.
3. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.*, et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984,** *106*, 765.
4. Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996,** *105*, 1902.
5. Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988,** *110*, 1657.
6. Daura, X.; Mark, A. E.; Van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* **1998,** *19*, 535.
7. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.*, et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65*, 712.
8. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
9. Jensen, F., *Introduction to Computational Chemistry*. 3rd ed.; Wiley: Chichester, UK, 2016.
10. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
11. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225.
12. Kadaoluwa Pathirannahalage, S. P.; Meftahi, N.; Elbourne, A.; Weiss, A. C. G.; McConville, C. F.*, et al.* Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2021,** *61*, 4521.
13. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
14. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157.
15. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247.
16. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269.
17. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993,** *115*, 9620.
18. Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23*, 1623.
19. Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.*, et al.* The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mater.* **2016,** *2*, 15011.
20. Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **1980,** *102*, 6218.
21. Åqvist, J.; Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* **1993,** *93*, 2523.
22. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.*, et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018,** *27*, 293.
23. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18*, 1463.

24. Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.*, et al.* Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comp. Mol. Sci.* **2018,** *1*, 5957.
25. Gapsys, V.; de Groot, B. L. On the importance of statistics in molecular simulations for thermodynamics, kinetics and simulation box size. *eLife* **2020,** *9*, e57589.
26. Tuckerman, M., *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press: Oxford, UK, 2010.
27. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089.
28. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.*, et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.
29. Lemkul, J. A. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *Living J. Comp. Mol. Sci.* **2018,** *1*, 5068.
30. Knapp, B.; Ospina, L.; Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018,** *14*, 6127.
31. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126*, 014101.
32. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81*, 3684.
33. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52*, 7182.
34. Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50*, 1055.
35. Wells, S. A.; van der Kamp, M. W.; McGeagh, J. D.; Mulholland, A. J. Structure and Function in Homodimeric Enzymes: Simulations of Cooperative and Independent Functional Motions. *PLOS ONE* **2015,** *10*, e0133372.
36. Miao, Y.; McCammon, J. A. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Simul.* **2016,** *42*, 1046.
37. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992,** *13*, 1011.
38. Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur. Phys. J. B* **2021,** *94*, 144.
39. Ciccotti, G.; Jacucci, G. Direct Computation of Dynamical Response by Molecular Dynamics: The Mobility of a Charged Lennard-Jones Particle. *Phys. Rev. Lett.* **1975,** *35*, 789.
40. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3*, 1767.
41. Ciccotti, G.; Jacucci, G.; McDonald, I. R. "Thought-experiments" by molecular dynamics. *J. Stat. Phys.* **1979,** *21*, 1.
42. Ciccotti, G.; Ferrario, M. Non-equilibrium by molecular dynamics: a dynamical approach. *Mol. Simul.* **2016,** *42*, 1385.
43. Gutierrez-de-Teran, H.; Aqvist, J. Linear interaction energy: method and applications in drug design. *Methods Mol. Biol.* **2012,** *819*, 305.
44. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401.
45. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.*, et al.* Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889.
46. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015,** *10*, 449.
47. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.*, et al.* End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019,** *119*, 9478.
48. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98*, 10037.

49. Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003,** *330*, 891.
50. Metz, A.; Pfleger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K.-H.*, et al.* Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein–Protein Interface. *J. Chem. Inf. Model.* **2012,** *52*, 120.
51. Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004,** *55*, 383.
52. Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017,** *46*, 531.
53. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016,** *7*, 207.
54. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006,** *125*, 084902.
55. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003,** *107*, 9535.
56. Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.*, et al.* Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living J. Comp. Mol. Sci.* **2020,** *2*, 18378.
57. Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017,** *57*, 2911.
58. Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.* **2018,** *9*, 923.
59. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided Drug Des.* **2011,** *7*, 146.
60. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007,** *28*, 1145.
61. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.*, et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30*, 2785.
62. Morris, G. M.; Lim-Wilby, M., Molecular Docking. In *Molecular Modeling of Proteins*, Kukol, A., Ed. Humana Press: Totowa, NJ, 2008; pp 365.
63. Zhang, Y.; Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019,** *35*, 5121.
64. Dunbrack Jr., R. L.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997,** *6*, 1661.
65. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976,** *103*, 227.
66. Warshel, A.; Karplus, M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *J. Am. Chem. Soc.* **1972,** *94*, 5612.
67. Smith, Jeremy C.; Roux, B. Eppur Si Muove! The 2013 Nobel Prize in Chemistry. *Structure* **2013,** *21*, 2102.
68. van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013,** *52*, 2708.
69. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.*, et al.* ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996,** *100*, 19357.
70. Senn, H. M.; O'Hagan, D.; Thiel, W. Insight into Enzymatic C−F Bond Formation from QM and QM/MM Calculations. *J. Am. Chem. Soc.* **2005,** *127*, 13643.
71. Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009,** *48*, 1198.
72. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev.* **2012,** *41*, 3025.
73. Rodríguez, A.; Oliva, C.; González, M.; van der Kamp, M.; Mulholland, A. J. Comparison of Different Quantum Mechanical/Molecular Mechanics Boundary Treatments in the Reaction of the Hepatitis C Virus NS3 Protease with the NS5A/5B Substrate. *J. Phys. Chem. B* **2007,** *111*, 12909.

3. M^{pro}-Substrate Interactions to Guide Inhibitor Design

3.1 Overview

Given the attractiveness of SARS-CoV-2 M^{pro} as a potential drug target (**Chapter 1**), an understanding of its activity towards native polyprotein substrates and its inhibition by peptide and small molecule inhibitors is of importance to drug discovery. This Chapter describes an early collaborative effort to gain mechanistic insights into the activity and inhibition of M^{pro} using various computational methods, when crystal structures of its substrate-bound complex were unavailable. The M^{pro}-substrate models enabled characterisation of key interactions and provided a basis for designing novel peptide inhibitors that were then experimentally validated. The interactions formed with substrates were compared with those associated with the binding of inhibitors, some of which target Cys145 via a β-lactam electrophilic scaffold.

3.2 Manuscript I

The following manuscript (**Manuscript I**) has been published as the following and is adapted with modifications for this Thesis, with permission from the Royal Society of Chemistry.

I. Chan, H. T. H.‡; Moesser, M. A.‡; Walters, R. K.‡; Malla, T. R.‡; Twidale, R. M.; John, T.; Deeks, H. M.; Johnston-Wood, T.; Mikhailov, V.; Sessions, R. B.; Dawson, W.; Salah, E.; Lukacik, P.; Strain-Damerell, C.; Owen, C. D.; Nakajima, T.; Świderek, K.; Lodola, A.; Moliner, V.; Glowacki, D. R.; Spencer, J.; Walsh, M. A.; Schofield, C. J.*; Genovese, L.*; Shoemark, D. K.*; Mulholland, A. J.*; Duarte, F.*; Morris, G. M.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021**, *12* (41), 13686-13703.

I performed the classical MD simulations (with hydrogen bond and MM/GBSA analysis) and ADCP peptide docking, and wrote the manuscript with other co-authors. A Statement of Authorship is appended following the manuscript. The Supplementary Information is in the **Appendix for Chapter 3, Section A**.

Discovery of SARS-CoV-2 M^{pro} Peptide Inhibitors from Modelling Substrate and Ligand Binding

Abstract

The main protease (M^{pro}) of SARS-CoV-2 is central to viral maturation and is a promising drug target, but little is known about structural aspects of how it binds to its 11 natural cleavage sites. We used biophysical and crystallographic data and an array of biomolecular simulation techniques, including automated docking, molecular dynamics (MD) and interactive MD in virtual reality, QM/MM, and linear-scaling DFT, to investigate the molecular features underlying recognition of the natural M^{pro} substrates. We extensively analysed the subsite interactions of modelled 11-residue cleavage site peptides, crystallographic ligands, and docked COVID Moonshot-designed covalent inhibitors. Our modelling studies reveal remarkable consistency in the hydrogen bonding patterns of the natural M^{pro} substrates, particularly on the N-terminal side of the scissile bond. They highlight the critical role of interactions beyond the immediate active site in recognition and catalysis, in particular plasticity at the S2 site. Building on our initial M^{pro}-substrate models, we used predictive saturation variation scanning (PreSaVS) to design peptides with improved affinity. Non-denaturing mass spectrometry and other biophysical analyses confirm these new and effective 'peptibitors' inhibit M^{pro} competitively. Our combined results provide new insights and highlight opportunities for the development of M^{pro} inhibitors as anti-COVID-19 drugs.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of coronavirus disease 2019 (COVID-19) that caused the World Health Organization to declare a global pandemic in March 2020. At the time of writing, >210 million COVID-19 cases and >4.4 million deaths have been reported worldwide.[1] A key step in maturation of SARS-CoV-2, a single-stranded positive-sense RNA virus, is hydrolysis of its polyproteins pp1a and pp1ab. Most of the cleavage events—at 11 sites—are performed by the SARS-CoV-2 main protease (M^{pro}; 3 chymotrypsin-like or 3CL proteinase, 3C-like protease, $3CL^{pro}$; or non-structural protein 5, Nsp5).

M^{pro} is a nucleophilic cysteine protease, which in solution is predominantly homodimeric. Each protomer consists of three domains, and the active site contains a cysteine-histidine catalytic dyad, Cys145 and His41, located near the dimer interface.[2] SARS-CoV-2 M^{pro} is 96% identical to the M^{pro} of SARS-CoV, which causes SARS.[3] Dimerization of M^{pro} is proposed as a prerequisite for catalysis: the N-terminus of one protomer contributes part of the active site of the other.[4] Indeed, the monomeric form of SARS-CoV M^{pro} is reported to be inactive.[5] Evidence from non-denaturing mass spectrometry (MS)-based assays indicates that M^{pro} monomers are not only inactive (at least with tested substrates), but do not bind 11-mer substrates with high affinity.[6]

SARS-CoV-2 M^{pro} and SARS-CoV M^{pro} have similar substrate specificities, both recognising the motif: [P4:Small] [P3:X] [P2:Leu/Phe/Val/Met] [P1:Gln] ↓ [P1′:Gly/Ala/Ser/Asn], "Small" denoting Ala, Val, Pro or Thr; "X" any residue; and "↓" the scissile amide (**Figure 3.1**).[7, 8] In part, because such sequences are not known to be recognised by any human protease, M^{pro} represents an attractive drug target.[4] Although no clinically approved M^{pro} drugs are available, small molecule inhibitors and peptidomimetics have been designed to inhibit SARS-CoV M^{pro} and, more recently, SARS-CoV-2 M^{pro}.[9, 10] Indeed, a covalent M^{pro} inhibitor from Pfizer has recently entered clinical trials.[11, 12]

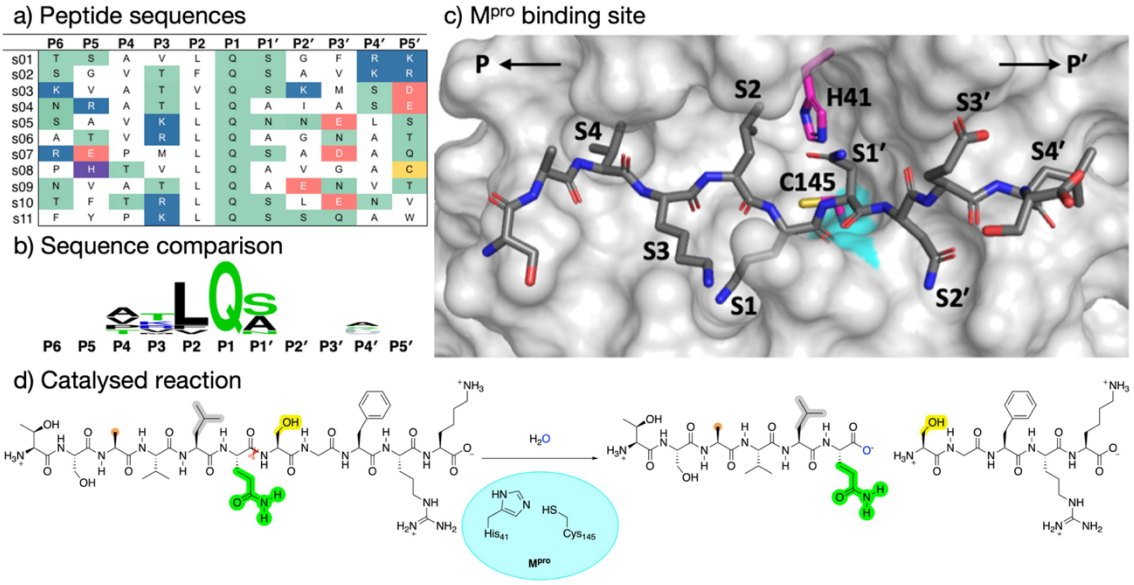


Figure 3.1: Substrates processed by SARS-CoV-2 M^{pro}. a) The 11 SARS-CoV-2 M^{pro} cleavage sites, and the corresponding 11-residue peptides, s01-s11; positively/negatively charged residues are blue/red, respectively; histidine is purple; residues with polar sidechains are green; and cysteine is yellow. b) Comparison between the 11 substrate sequences (generated by WebLogo)[13] highlighting the completely conserved Gln at P1 and the highly conserved Leu at P2. c) View of an energy minimised model, built using *apo* M^{pro} (PDB 6YB7, light grey surface)[14], of M^{pro} complexed with s05 (dark grey sticks); subsites S4-S4′ are labelled. The oxyanion hole formed by the M^{pro} backbone NHs of Gly143, Ser144 and Cys145 is cyan. d) The reaction catalysed by M^{pro} exemplified by s01. Substrate residues important in recognition (see main text) are highlighted.

Multiple crystallographic and computational modelling studies concerning the M^{pro} mechanism[15-18] and inhibition are available;[19-24] the CORD-19 database[25] documents many such studies. It is proposed that during M^{pro} catalysis, His41 deprotonates the Cys145 thiol, which reacts with the carbonyl of the scissile amide to give an acyl-enzyme intermediate. This intermediate is stabilised by a hydrogen bond network that holds the scissile amide carbonyl in an 'oxyanion hole'. The C-terminal part of the product likely leaves the active site at this stage. The acyl-enzyme intermediate is subsequently hydrolysed with loss of the N-terminal product regenerating active M^{pro}. Computational and mechanistic studies on SARS-CoV M^{pro} [26-29] and SARS-CoV-2 M^{pro} [30-32] suggest that in the resting state His41 and Cys145 are likely neutral and that the protonation states of nearby histidines (*e.g.* His163, 164, and 172) affect the structure of the catalytic machinery — although it has been suggested in SARS-CoV M^{pro} that the protonation state of the catalytic dyad may change in the presence of an inhibitor or substrate.[33] A different picture has

been obtained from neutron crystallographic studies, which indicate that an ion pair form of the dyad is favoured at pH 6.6.[34] While neutron crystallography, in principle, enables the direct determination of hydrogen atom positions, questions remain about how pH and the presence of active site-bound ligands influence the precise—and likely dynamic—protonation state(s) of the dyad.

Important questions remain regarding M^{pro} catalysis, including to what extent the active site protonation state, solvent accessibility, induced fit, and substrate sequence influence activity. The lack of this knowledge makes it difficult to carry out effective computational studies on M^{pro} catalysis and inhibition.

With the aim of helping to combat COVID-19, in April 2020 we embarked on a collaborative effort involving weekly virtual meetings, initially to investigate the relationship between M^{pro} substrate selectivity and activity. We employed an array of classical molecular mechanics (MM) and quantum mechanical (QM) techniques, including non-covalent and covalent automated docking, molecular dynamics (MD) simulations, density functional theory (DFT), combined quantum mechanics/molecular mechanics (QM/MM) modelling, and interactive MD in virtual reality (iMD-VR). Our results provide consensus atomic-level insights into the interactions of M^{pro} with 11-residue peptides derived from the 11 natural cleavage sites (named "s01" to "s11", in order of occurrence in the viral polyprotein, **Figure 3.1a**). The identification of key interactions between M^{pro} and its substrates, together with analysis of fragment/inhibitor structures,[35, 36] led to the design of peptides proposed to bind more tightly than the natural substrates, several of which inhibit M^{pro}. The results are freely available via GitHub (https://github.com/gmm/SARS-CoV-2-Modelling).

2. Results and Discussion: Understanding substrate binding and recognition

Protonation state of the catalytic dyad

The protonation state of the dyad following substrate binding was studied with s01-bound (**Figure 3.1a**) M^{pro} [37] using QM/MM umbrella sampling simulations at the DFTB3/MM and ωB97X-D/6-31G(d)/MM levels of theory (**Supplementary Information** in **Appendix for Chapter 3: Section S3.1a**). The protonation states of nearby histidines were also evaluated (**Table S3.2.1** and **Figures S3.2.5-6)**.

His41 was treated as Nδ-protonated in the neutral state of the dyad, as reported by Pavlova *et al.* to be preferred for both uncomplexed and N3 inhibitor-bound M^{pro} based on MD studies.[30] The protonation state of the dyad-neighbouring His163, which interacts with Tyr161, Phe140 and the substrate P1 Gln sidechain, was also studied. Three His163 protonation states were considered: (i) Nδ-protonated, neutral ("HID"); (ii) Nε-protonated, neutral ("HIE"); and (iii) Nδ and Nε-protonated, positively charged ("HIP").[38] For all three His163 protonation states, the forward trajectories (neutral, N to ion pair, IP) showed the anionic Cys145 form to be stabilised solely by interaction with positively charged His41 (**Figures 3.2a** and **S3.2.2A-C**), and the thiolate in an unreactive conformation for nucleophilic attack. This suggests that such a zwitterionic state is transient, with concerted proton transfer and simultaneous nucleophilic attack of the thiolate onto the scissile amide carbon being more likely than a stepwise mechanism.[15] By contrast, the backwards proton transfer (PT) trajectories (from IP to N) showed stabilisation of the Cys145 thiolate by His41 and the P1 Gln backbone N-H, in the case of HID163 and HIP163 (**Figure S3.2.2E, F**). For HIE163, additional stabilisation comes from interactions with the backbone N-H and the hydroxyl of the P1′ Ser, and an additional water that diffuses into the active site (**Figures 3.2b** and **S3.2.2D**).

The zwitterionic state with HID163 was less stable than the neutral state and the zwitterionic states with HIE163 and HIP163 (**Figure S3.2.1C**). This was due to perturbation of the interaction

network with Tyr161 and Phe140, suggesting that a Nδ-protonated His163 is unlikely. Double protonation of His163 results in a loss of both interactions in the forwards and backwards PT trajectories. Despite both HIP163 and HIE163 giving similar PT free energy profiles, the loss of these interactions suggests HIP163 is unfavourable for productive catalysis. These QM/MM results therefore suggest that an Nε-protonated neutral His163 is most likely. Along with conserving interactions with Tyr161 and Phe140, an Nε-protonated His163 also formed a hydrogen bond with the P1 Gln side chain (**Figure 3.2a, b**), an interaction not observed in PT trajectories with HID163 and HIP163.

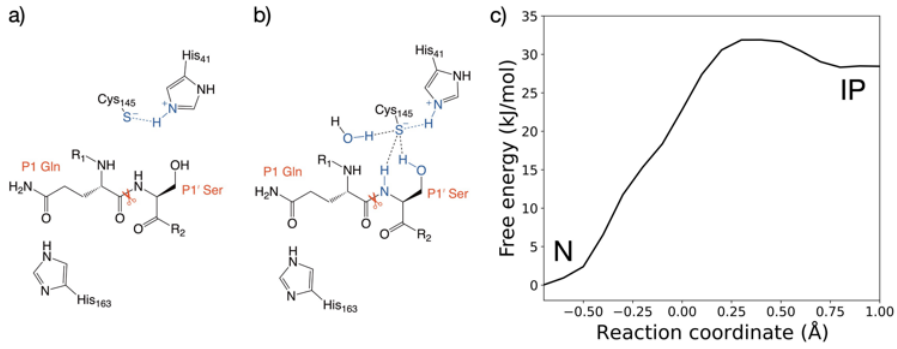


Figure 3.2: QM/MM umbrella sampling of the proton transfer in the catalytic dyad. Representation of the interactions of the Cys145 thiolate in the ion pair (IP) state from the a) forwards and b) backwards simulations. c) Free energy profile for interconversion of the neutral (N) and ion pair states of the dyad in the HIE163 system, from the combined forwards and backwards QM/MM umbrella sampling MD simulations, corrected to the ωB97X-D/6-31G(d)/MM level of theory.

Considering that DFTB3 overestimates the proton affinity of methylimidazole, it is expected that this method will over-stabilise the zwitterionic state relative to the neutral state.[39] To account for this, the backwards PT reaction with a Nδ-protonated His163 was modelled at the ωB97X-D/6-31G(d)/MM level of theory. This showed the zwitterionic state was 24.3 kJ mol^{-1} above the neutral state, an increase of 26.4 kJ mol^{-1} compared to DFTB3/MM (**Figure S3.2.3**). Applying the free energy difference between ωB97X-D/6-31G(d) and DFTB3 at each reaction coordinate value as a correction to the combined QM/MM free energy profile in the case of HIE163, the neutral

catalytic dyad is preferred, with the ion pair being 28.5 kJ mol^{-1} higher in energy than the neutral state (**Figure 3.2c**). Similar results were obtained with a different QM approach (**Figure S3.2.4**).

Models of SARS-CoV-2 M^{pro}-substrate peptide complexes

To understand substrate specificity and to assess their relative binding affinities, we constructed 11 models of SARS-CoV-2 M^{pro} complexed with its cleavage site-derived substrates[40] as 11-amino acid peptides, from P6 to P5′ (see main text **Method**; **Figures 3.1a** and **S3.2.7**). We refer to these peptides as 'substrates' as their hydrolytic sites are all cleaved by M^{pro} (*vide infra*). The substrates were modelled in crystallographic chain A of the M^{pro} dimer (PDB 6YB7)[14] with a neutral dyad; unless otherwise stated, all M^{pro} residue numbers and names in the following discussions refer to chain A. Initial models were subjected to three independent explicit-solvent MD simulations each of 200 ns.

Three of the 11 cleavage site-derived peptides (s01, s02, and s05) were also modelled with interactive MD using virtual reality (iMD-VR), as an alternative to comparative modelling and traditional MD. iMD-VR provides an immersive 3D environment for users to interact with physically rigorous MD simulations.[41-43] The three substrates were chosen because s01 and s02 have the highest relative efficiencies (of SARS-CoV M^{pro}) of all substrates; while s05 has the second-lowest catalytic efficiency but the same P2 and P1 residues as s01.[44]

Throughout both the explicit-solvent MD and implicit-solvent iMD-VR simulations, all the substrates remained tightly bound in the active site (**Figures S3.2.8-11**). Substrate backbone stability was maintained especially in the central region, with only the N- and C-terminal residues showing substantial flexibility. Local sidechain fluctuations were present, notably at the solvent-facing P3 residue (**Figures 3.1c** and **S3.2.9**). C-terminal P′-residues consistently fluctuated more than the N-terminal P-side residues, likely in part because of fewer protein-substrate hydrogen bonds on the P′ side (*vide infra*).

Conserved hydrogen bond interactions. Crystallographic studies on SARS-CoV M^{pro} revealed the importance of hydrogen bonds in binding of substrate s01,[37] which has the same cleavage site sequence in SARS-CoV-2 (**Figure 3.1a**). To investigate whether this was true for SARS-CoV-2 M^{pro} complexed with its 11 substrates, we analysed the hydrogen bond (HB) prevalence at each subsite in both explicit-solvent MD and implicit-solvent iMD-VR simulations. Twelve HBs were consistently identified (**Figure 3.3**) and their distance and angular distributions analysed (**Figure S3.2.12)**.

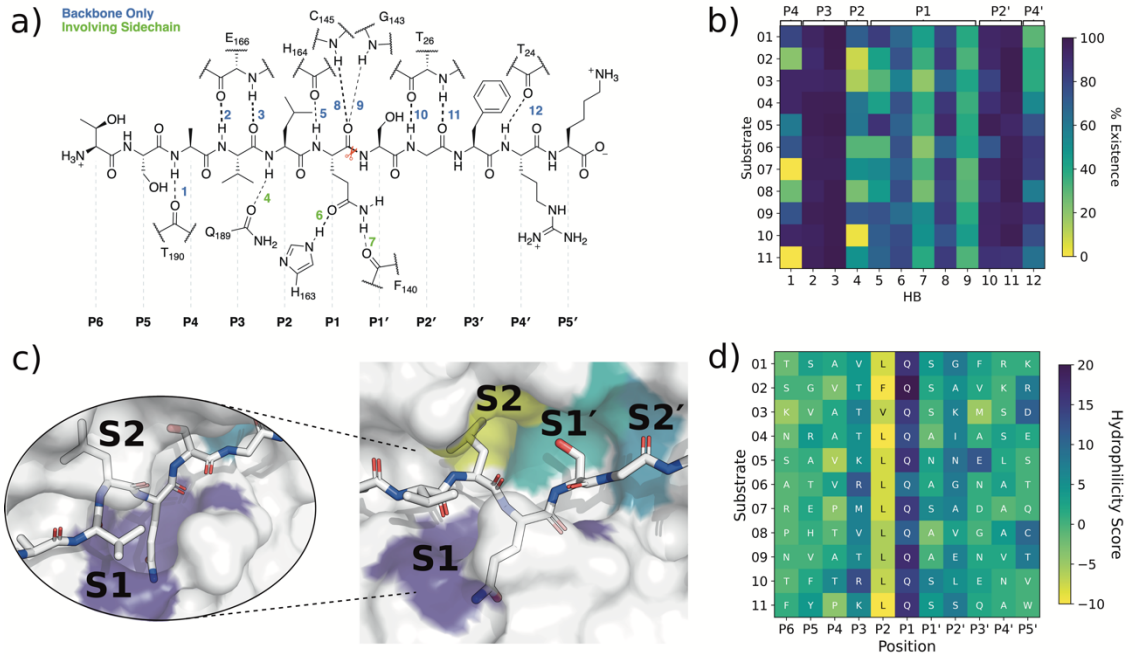


Figure 3.3: Interactions between SARS-CoV-2 M^{pro} and its substrates. a) M^{pro}-substrate hydrogen bonds (HBs) exemplified by substrate s01. b) An annotated heat-map showing the frequency of each HB, with blue indicating highest frequency. Frames were extracted every ns from 600 ns of cumulative explicit-solvent MD conducted per system. c) Close-up of the MD-generated binding mode of SARS-CoV-2 M^{pro} substrate s01 with subsites S1, S2, S1′ and S2′ labelled. Different views of the S1 subsite are shown, emphasising the deep S1 pocket that accommodates the P1 Gln sidechain. Subsite surface colour corresponds to the hydrophilicity score, with hydrophobic subsites shown in yellow, hydrophilic subsites in dark blue, and amphiphilic subsites in turquoise. d) Hydrophilicity map for the 11 substrates calculated as the sum of hydrophilic interactions (dark blue) subtracted from the sum of hydrophobic interactions (yellow). Interactions were identified from substrate-M^{pro} frames using Arpeggio[45] (**SI Methods Section S3.1c**).

In both explicit-solvent MD and iMD-VR simulations, all 11 substrates are primarily held in place by four consistently formed backbone-backbone HBs: Glu166 at S3 (2 & 3) and Thr26 at S2′ (10

& 11; **Figure 3.3**). Backbone HBs 1 and 12 further from the cleavage site show greater variation between the MD and iMD-VR studies. Although P1 Gln is conserved in all 11 M^{pro} cleavage sites (**Figure 3.1a, b**), HBs 5-9 formed in the S1 site are observed less often than 2, 3, 10 and 11, but outnumber other sites. In both types of simulations, HB 8 from the Cys145 backbone amide is consistently formed, suggesting that this could play a fundamental role in catalysis. HB 8 forms part of the oxyanion hole, which stabilises the tetrahedral intermediate formed upon nucleophilic attack by Cys145 Sγ on the scissile amide. The exquisite specificity of M^{pro} for Gln at P1 is likely due to formation of HB 6 with His163, and to a lesser extent HB 7, along with the narrowness of the S1 pocket accommodating the Gln sidechain in an extended conformation.

Hydrophobicity analysis. Hydrophilicity analysis shows that the S1 subsite is substantially hydrophilic in all 11 substrate-M^{pro} complexes, while S2 is consistently hydrophobic (**Figure 3.3d**). In accord with the MD-HB analysis (**Figure 3.3b**), the conservation of subsite interactions decreases further from the cleavage site. Although S3 and S2′ show a slight bias towards hydrophilic interactions, none of the other subsites show a consistent pattern.

Other non-covalent interactions. Beyond HBs, several other interaction types are conserved across the substrates; these were identified by running Arpeggio[45] on snapshots extracted from the explicit-solvent MD simulations (**Figure 3.4**). Six of the eight most common P1 interactions are present in most (*i.e.,* ≥9/11) substrates. This includes the previously described HBs between the P1 backbone oxygen and the backbone NHs of Cys145 (HB 8) and Gly143 (HB 9) that constitute the oxyanion hole, as well as HB 6 (His163) and HB 7 (Phe140) that stabilise the P1 Gln sidechain. In addition, interactions with Ser144, His163, His164, Glu166 and, to a lesser extent, Phe140 were prevalent at P1. None of the other subsites show this level of consistency in residue-level contacts, although some interactions such as hydrophobic contacts with Met49 and Met165 were always present at P2. Furthermore, important stabilising backbone HBs (HBs 2-3 between Glu166 and P3; and HBs 10-11 between Thr26 and P2′) were conserved in all substrates. Finally, P′ interactions are less common than those on the P side (**Figure 3.4**). The same trend

was found when docking s01, s02, and s05 using iMD-VR, where P′ residues tended to be more flexible than P-side residues.

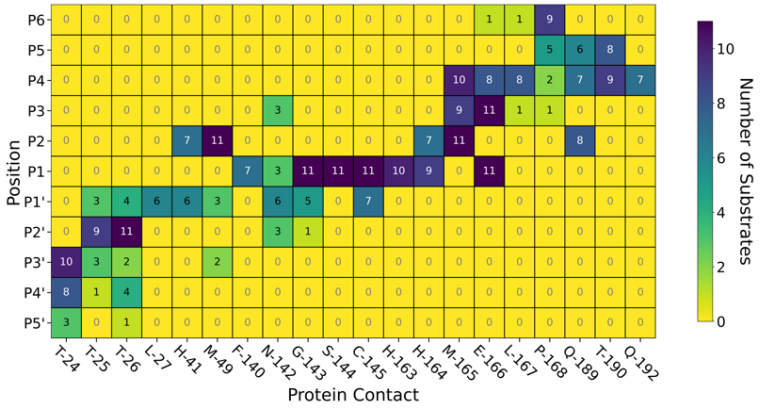


Figure 3.4: M^{pro}-substrate contacts. Map of HBs and other non-covalent interactions between the 11 substrates and M^{pro} from Arpeggio[45] analysis of the most representative pose for each substrate generated by MD. Dark blue indicates the interaction is formed by all 11 substrates at that substrate position, while yellow indicates no substrates form this interaction.

MM/GBSA analysis. We analysed van der Waals and electrostatic contributions to protein-substrate interactions by employing the molecular mechanics/generalised Born surface area (MM/GBSA) method.[46-49] The ten M^{pro} residues contributing most to the binding energy were identified for each of the 11 substrate complexes (**Figure S3.2.13**). These residues are all within proximity to the complexed substrate (**Figure 3.5**) and were also identified by Arpeggio as conserved contacts (**Figure 3.4**).

As anticipated, residues that form the most stable HBs (HBs 2, 3, 10, 11), namely Glu166 and Thr26, display large favourable contributions (**Figure 3.5**); Glu166 sidechain has a larger total contribution due to its interaction with P1 Gln. Other HB-forming residues (backbones of Cys145, Thr190, and Thr24; and sidechains of His163 and Gln189) are also identified as hotspots. For the remaining consistently interacting residues, contributions to binding are dominated by van der Waals interactions (**Figure S3.2.14**), exemplified by Met165 and His41, both of which engage in non-polar contacts with the hydrophobic P2 residue. Optimising interactions with these hotspot residues could help guide the design of optimal M^{pro} inhibitors.

By applying a similar per-residue decomposition of MM/GBSA binding energy at each substrate position (**Figure S3.2.15**), contributions from the P2 sidechain are significant (**Figure S3.2.16**). The S2 site appears to tolerate Phe (s02) well, in addition to the more common Leu that is found in nine of the eleven substrates, while Val (s03) is less favourable. The possibility of filling the S2 pocket with a larger, aromatic moiety is of interest in designing inhibitors.

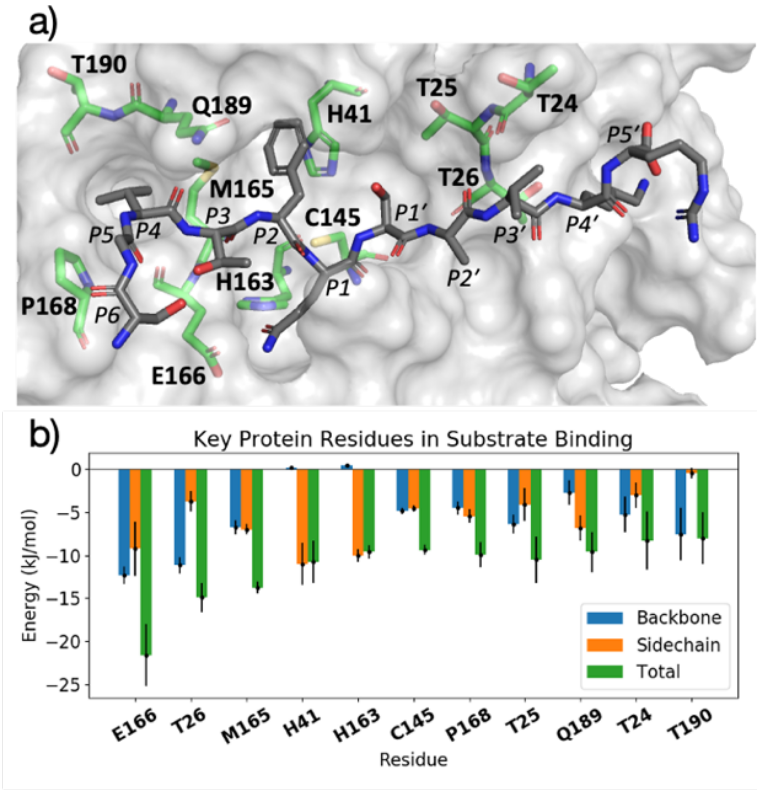


Figure 3.5: M^{pro} residues involved in substrate binding identified by MM/GBSA. a) View of the energy minimised structure of M^{pro} (PDB 6YB7, light grey surface)[14] in complex with s02 (dark grey sticks), with hotspot residues as identified by MM/GBSA per-residue decomposition shown (green sticks). b) Contributions to the MM/GBSA binding energy by each hotspot residue (average ± standard deviation across 11 systems), in descending order of consistency across the 11 substrates.

Density functional theory analysis of the interaction network. We performed linear-scaling DFT (BigDFT[50]) calculations using representative snapshots extracted from explicitly solvated MD trajectories of M^{pro} complexed with s01, s02, and s05. By automatically decomposing large molecular systems into coarse-grained subsets of atoms (or 'fragments') in an unbiased manner,[51] quantities like inter-fragment bond order and interaction strengths, E_{cont}, can be easily calculated

(**Figure 3.6a**, **SI Section S3.2d** and **Figures S3.2.17-20)**. This analysis supports the essential roles of Glu166 and Thr26, with interactions observed in all three peptides s01, s02, and s05, consistent with the HB analysis described earlier (**Figure 3.3**). Gln189 consistently hydrogen-bonds with P2 (HB 4) in s01 and s05, but rarely in s02.This weakening of HB 4 in P2 may be due the greater bulk of Phe in s02 (**Figure 3.6b**).

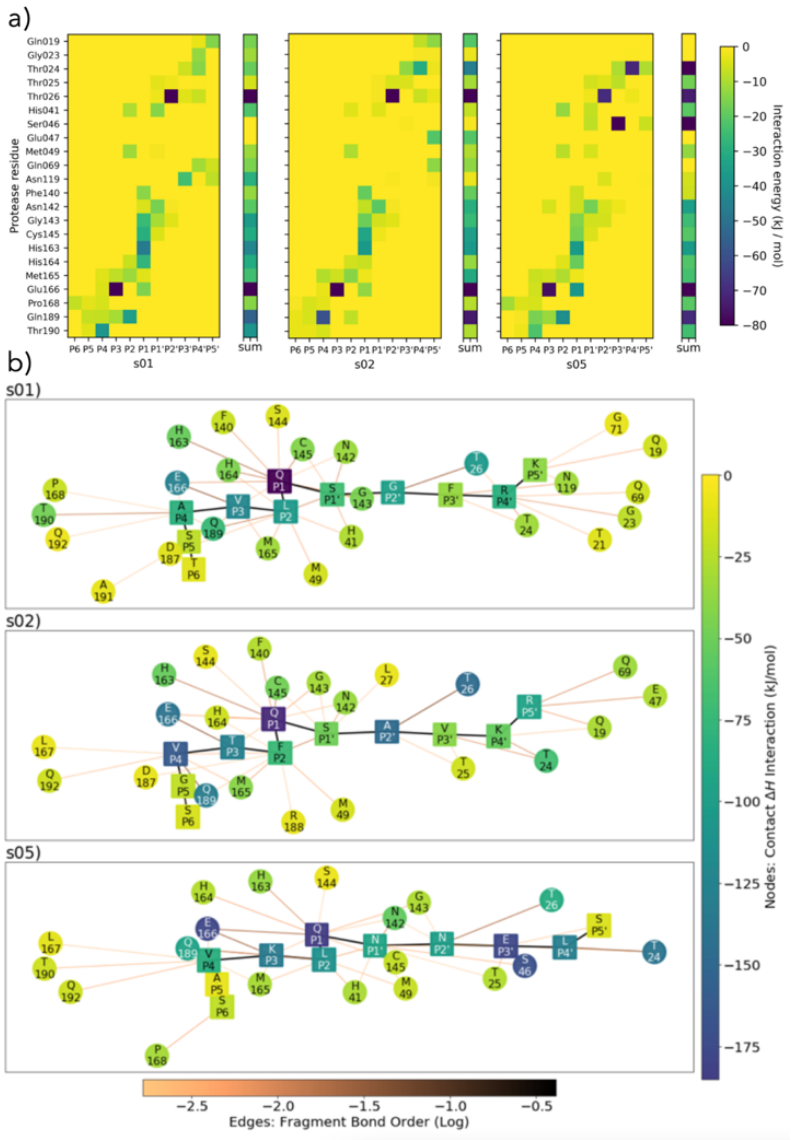


Figure 3.6: BigDFT analysis of M^{pro}-substrate interactions. a) Heatmaps showing QM interaction energies between 22 selected residues of M^{pro} and s01, s02 and s05. b) QM interaction networks where node colour indicates interaction strength, from dark blue (strongest) through green to yellow (weakest). Square nodes denote substrate, while circular ones denote M^{pro}. The thickness and colour of the edges show the fragment bond order between residues, a unitless measure associated with bond strength and analogous to bond order; black is strongest, orange is weakest.[51] Interaction energies and bond orders were computed using BigDFT and ensemble-averaged results of MD snapshots.

Conserved contacts are present in the three substrates between Cys145 and both P1 and P1′ residues. Interactions between His41 and P2/P1′ are observed for s01 and s05, and between Glu166 and P1/P3 (and, to some extent, to P4) for all three substrates. This analysis singles out the character of s02, which is dominated by the bulky character of its P2 Phe. Substitutions at P2 may have a substantial effect on the interaction network close to the catalytic site. While the P side exhibits an interconnected character especially from P1 to P4 (**Figure 3.6b**), the network on the P′ side has a more linear character, once again indicating that hotspot residues responsible for binding are present on the P side. Distributions of E_{cont} are shown in **Figure S3.2.19**.

The following trends emerge from our studies on M^{pro} in complex with models of its 11 substrates: (i) Binding stability is partly conferred by a series of HBs from P4 to P4′, in particular between the backbones of M^{pro} Glu166 and Thr26 and substrate positions P3 and P2′ respectively, as well as HBs involving the conserved P1 Gln sidechain; (ii) substrate residues N-terminal of the cleavage site (P side) form more, and more consistent, contact interactions with M^{pro} compared to the P′ side, with interactions at Met49, Gly143, Ser144, Cys145, His163, His164, Met165 and Glu166 being most conserved. We conclude that the S1 and S2 pockets are prime targets for active site substrate-competing inhibitor design due to their well-defined hydrophilic character, large energy contributions to substrate binding, and vital conserved hydrogen bonds in S1 for substrate recognition.

Conformational plasticity in M^{pro} crystal structures. Previous studies have compared the dynamics of ligand binding sites across SARS-CoV-2, SARS-CoV and MERS-CoV M^{pro}.[52] Here, we investigated the conformational plasticity of the SARS-CoV-2 M^{pro} active site upon binding by comparing 333 M^{pro}-ligand co-crystal structures obtained from Fragalysis[53] to a reference *apo* structure of M^{pro}, PDB 6YB7 [14] (**Figure 3.7**). A high degree of plasticity was observed at residues Thr24, Thr25, His41, Thr45, Ser46, Met49, Asn142, Met165, Glu166, Arg188, Gln189 and Ala191. Unsurprisingly, the S1 subsite is particularly rigid, with almost no change in residue conformations across all 333 crystal structures.

In all cases, P1 Gln recognition is mainly driven by interactions with Gly143, Ser144, Cys145 (oxyanion hole) and His163 and Phe140. The S2 subsite, however, is highly flexible, especially at Thr45, Ser46 and Met49. Although P2 is conserved in terms of hydrophobicity (Leu, Phe, Val), the S2 pocket is highly flexible and can adapt to accommodate functional groups of varying sizes, including aliphatic and aromatic groups. The outer regions of the active site (S3-S6 and S2′-S5′) vary in flexibility, echoing our MD simulations.

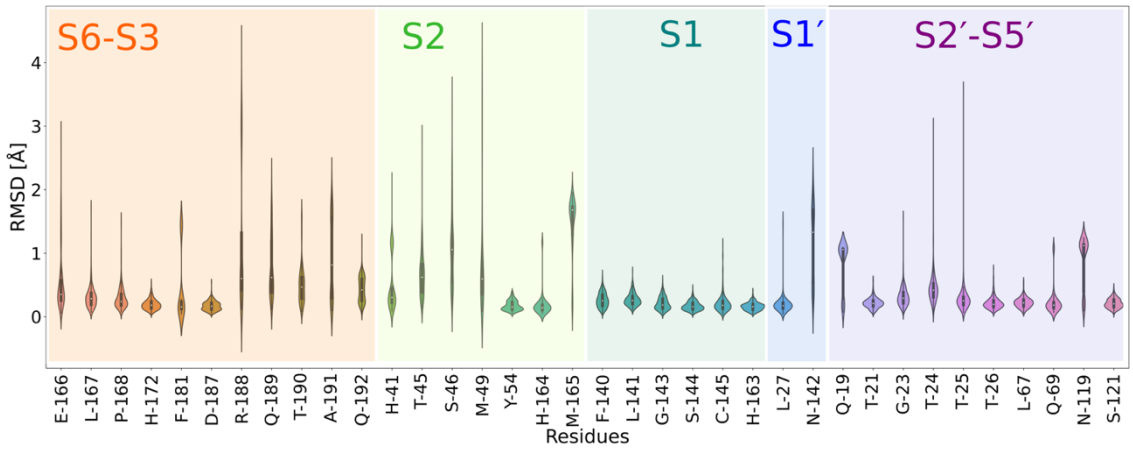


Figure 3.7: Analysis of the active site plasticity of 333 M^{pro} co-crystal structures. Active site residues (residues 19, 21, 23- 27, 41, 45, 46, 49, 54, 67, 69, 119, 121, 140-145, 163-168, 172, 181, 187-192) were chosen based on the MD analysis of the 11 substrate-M^{pro} models and correspond to all M^{pro} residues that contact any substrate. The violin plots show the distributions of per-residue heavy atom root mean square deviation (RMSD) values between the 333 M^{pro}-ligand co-crystal structures[53] and a reference uncomplexed structure (PDB 6YB7).[14] Each M^{pro} subsite is colour-coded.

Monitoring of substrate sequence hydrolysis by mass spectrometry

To rank the SARS-CoV-2 M^{pro} preferences for hydrolysis of the 11 cleavage sites, we monitored turnover of 11-mer peptides by solid-phase extraction (SPE) coupled to mass spectrometry (MS). Interestingly, after the N-terminal autocleavage site s01, s11 was the next preferred substrate for catalysis (**Figure S3.2.21**). Peptides s06, s02, and s10 were hydrolysed less efficiently than s11. Slow turnover was observed for s07 and s09. Evidence for low turnover of s05 was obtained after prolonged incubation with M^{pro} (9.56%) (**Figure S3.2.22**). Under our standard conditions, no evidence for cleavage was observed for s03, s04, and s08.

We then examined turnover under non-denaturing MS conditions using ammonium acetate buffer (**Figure 3.8a**). Peptides s01, s06, s08, s10 and s11 evidenced fast turnover. The level of substrate ion depletion was >70% after 1-min and >90% after 6-min incubation. Peptides s02, s04 and s09 showed substrate ion depletion from 35 to 45% after 1-min incubation, >70% depletion after 6 mins, and >90% depletion after 12 mins. Peptides s03, s05 and s07 demonstrated slow turnover that was below 50% after 12-min incubation.

In the protein region of the mass spectra, complexes of the M^{pro} dimer and the cleavage products were observed after 1 min of incubation for the fast-turnover substrates s01, s06, s08, s10 and s11, and also s02, s04 and s09 (**Figure 3.8b1-4**). For the slow-turnover substrates s03, s05 and s07, only M^{pro} complexes with intact substrates were observed after 1-min incubation. For longer incubation times, complexes between M^{pro} and the products from these substrates emerged and increased in abundance (**Figure 3.8b5-6**).

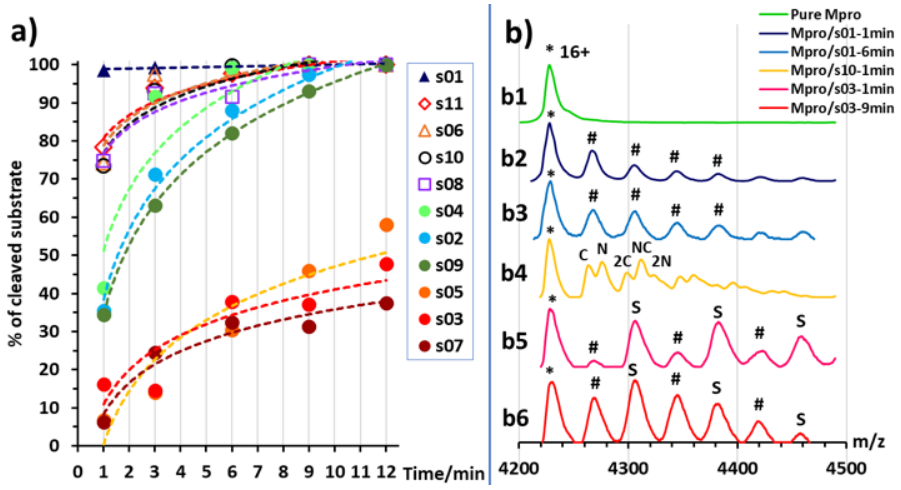


Figure 3.8: Non-denaturing mass spectrometry of M^{pro} substrate turnover. a) Substrate turnover *versus* incubation time as measured by non-denaturing MS. Trend lines are given for visual guidance only. b) Examples of mass spectra showing normalised intensity in the *m/z* region around the 16+ charge state of M^{pro} dimer (asterisk, *): b1) pure M^{pro} solution (5 µM); b2) M^{pro} and s01 solution after 1-min incubation, hashes (#) indicate the mass peaks corresponding to the s01 cleaved fragments sequentially attached to the M^{pro} dimer; note: the resolution is not sufficient to distinguish between the N- or C-terminal fragments (mass shifts of 617 and 593 Da, respectively); b3) same solution as b2) after 6-min incubation; b4) M^{pro} and s10 solution after 1-min incubation, 'N' labels N-terminal fragment(s) attached (765 Da) , 'C' labels C-terminal fragment(s) attached (560 Da); b5) M^{pro} and s03 solution after 3-min incubation, 'S' labels intact substrate(s) attached, hashes (#) label attached substrate fragments, but the N- and C-terminal fragments cannot be distinguished (mass shift 644 and 566 Da, respectively); b6) same solution as b5) after 9-min incubation.

The rank order of the substrates in part depends on the MS method used, likely due to the differences in the buffers and concentrations used: *i.e.,* non-denaturing MS used ammonium acetate buffer and an M^{pro} concentration of 5 μM, which is higher than the 0.15 μM used in denaturing MS assays. Higher concentrations of both enzyme and substrate in the non-denaturing MS experiments explain the faster substrate turnover than seen with denaturing MS, especially as the concentration of catalytically active M^{pro} dimer would be higher at higher enzyme concentrations.[6, 54]

Regardless of the MS method used, a clear trend is observed in the catalytic turnover of the cleavage site-derived peptides. The rank order of substrate preference under denaturing MS conditions was s01 > s11 > s06 > s02 > s10 > s07 > s09 > s05 (**Figures S3.2.21-22**). Under non-denaturing conditions (**Figure 3.8**) turnover was: fast (s01, s11, s06, s10, and s08), medium (s04, s02, and s09), and slow (s05, s03, and s07). Substrates s01, s11 and s06 turned over fastest; while s07, s05 and s03 were slow as measured by both methods. This is in broad agreement with the reversed phase high performance liquid chromatography analysis of substrate turnover by SARS-CoV M^{pro}, where s01 and s02 display fast turnover; s10, s11 and s06 manifest medium turnover; and the rest (s09, s08, s04, s03, s05, s07) show slow turnover.[44] Both of our MS studies on SARS-CoV-2 M^{pro} indicate that s02 consistently displayed slower turnover than s11. Previous reports on SARS-CoV M^{pro} have shown evidence for cooperativity between subsites during substrate binding, in particular during autocleavage of the M^{pro} C-terminal site (s02), where the Phe at P2 induces formation of the S3′ subsite to accommodate the P3′ Phe residue.[55] SARS-CoV-2 M^{pro} substrate s02 has a Phe at P2, but not at P3′ (**Figure 3.1a**). The absence of a Phe at the P3′ position may in part explain the reduced activity of SARS-CoV-2 M^{pro} for s02 relative to s01, compared to the same pair in SARS-CoV M^{pro}.[44]

The observed turnover of all 11 SARS-CoV-2 cleavage-site-derived peptides by M^{pro} is consistent with our atomistic models, where the peptides remain bound in the active site during MD simulations and where the scissile amide carbonyl remains well-positioned in the oxyanion hole

(*e.g.*, HB 8 in **Figure 3.3**) for reaction initiation. The stability of the M^{pro}-peptide interactions involving the S2 and S1 subsites, as well as backbone-backbone HBs 2, 3, 10 and 11, could explain the observation using non-denaturing MS of complexes of M^{pro} with products — because of slow product dissociation. Nevertheless, we envisage that the order of substrate turnover rates is likely determined by various factors, including peptide conformations, the influence of the P2 and P1′ residues on the catalytic dyad (as highlighted by the BigDFT analysis), entropic effects, and rates of product dissociation, all of which prompt ongoing experimental and computational investigations.

3. *In silico* mutational analysis of substrate peptides enables peptide inhibitor design

Building on insights gained from our binding studies of SARS-CoV-2 M^{pro} and the 11 SARS-CoV-2 polypeptide substrate sequences, we designed peptides that could bind more tightly than the native substrates. We quantified the per-residue energetic contributions of these sequences to the overall binding in the M^{pro} active site and proposed substitutions that would increase affinity. We hypothesised these peptides would: a) behave as competitive inhibitors, and b) provide counterpoints for comparison with natural substrates, shedding light on requirements for M^{pro} binding and, perhaps, turnover.

In silico alanine scanning and predictive saturation variation scanning

We used the interactive web application BAlaS to perform Computational Alanine-Scanning mutagenesis (CAS) using BudeAlaScan[56] and the BUDE_SM algorithm[57] for Predictive Saturation Variation Scanning (PreSaVS).[58] Both are built on the docking algorithm BUDE,[59] which uses a semi-empirical free energy force field to calculate binding energies.[60] To identify key binding interactions of the natural substrate peptides to M^{pro}, the 11 substrate-M^{pro} complexes were first subjected to CAS using BAlaS. By sequentially substituting for alanine, the energetic

contribution of each substrate residue to the overall interaction energy between the singly mutated peptide and M^{pro} is calculated using:

$$\Delta\Delta G = \Delta G_{\text{Ala}} - \Delta G_{\text{wt}} \tag{3.1}$$

where ΔG_{wt} is the interaction energy between the peptide and M^{pro}, and ΔG_{Ala} is the interaction energy for the peptide with a single alanine mutation at a given position. The more positive the value for each residue, the greater the contribution from that substrate residue to binding. This method was used later to evaluate potential inhibitor peptides.

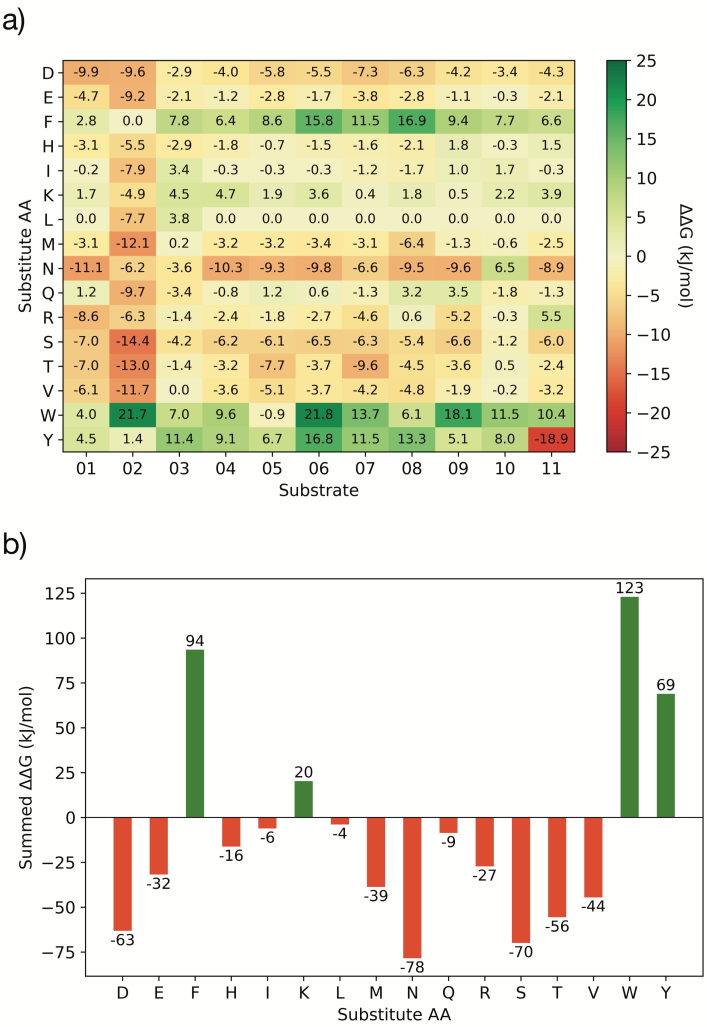


Figure 3.9: BUDE_SM PreSaVS for the P2 position. a) Heat-map for BUDE_SM PreSaVS saturation mutagenesis at P2, showing the $\Delta\Delta G = \Delta G_{wt} - \Delta G_{mut}$ value calculated for each substitution and each M^{pro} substrate. Mutations predicted to improve peptide binding have a positive $\Delta\Delta G$ and are greener; those disfavouring binding are in red. b) The summed $\Delta\Delta G$ values for each residue type substituted at P2.

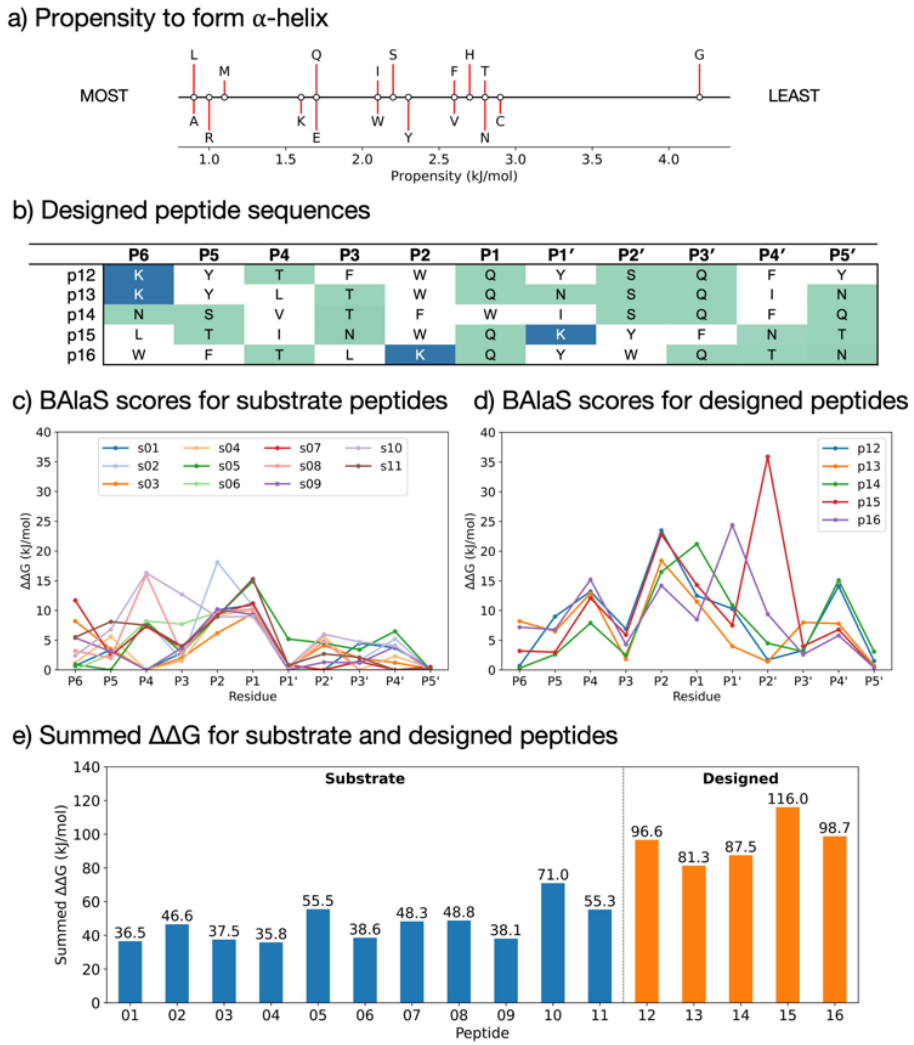
Having identified residues contributing most to the binding energy of the natural M^{pro} substrates, each of the sequences was subjected to PreSaVS using the BUDE_SM algorithm. This sequentially substitutes each substrate residue with a range of residues (D, E, F, H, I, K, L, M, N, Q, R, S, T, V, W and Y). BUDE_SM calculates the $\Delta\Delta G = \Delta G_{wt} - \Delta G_{mut}$ for the binding interaction of each, entire, singly mutated peptide with M^{pro}. Substitutions predicted to improve binding over wildtype sequences have a positive $\Delta\Delta G$. **Figure 3.9** shows an example of the BUDE_SM PreSaVS results for all the P2 substitutions for the 11 substrate peptides (normally Leu, Phe, or Val in the 11 substrates). The most positive results suggest that Phe, Trp and Tyr favour increased predicted affinity at the P2 position (**Figure 3.9b**). However, although Tyr generally increased the predicted binding affinity ($\Delta\Delta G_{sum}$ = 68.8 kJ mol^{-1}), it was not considered for substitution at P2 due to its negative effect at this position in s11 (scoring -18.9, **Figure 3.9a**). Candidate residues for each position, from P6 to P5′, were shortlisted similarly based on those with the best total, and the fewest unfavourable, scores.

In addition to the computed $\Delta\Delta G$ values, we considered the propensity of each residue to promote an extended conformation. All bound substrates are largely extended, so entropic penalties may be avoided if inherently extended conformations could be favoured in the designed peptide. Thus, the best β-forming (and therefore least α-forming) residues from the first triage were selected (**Figure 3.10a**).[61] We also considered solubility. This was achieved by limiting the number of hydrophobic residues in each designed peptide and ensuring a net positive charge (except p14, which was neutral).

Designed peptide sequences

Employing the criteria described above, five new peptides, p12-p16, were designed (**Figure 3.10b**). Comparison of the computed ΔΔG values for s01-s11 (**Figure 3.10c**) and p12-p16 (**Figure 3.10d**) reveals that substitutions at the P sites provide only occasional, moderate improvements to binding energy over the corresponding substrate P sites, with the notable exception of P2, which

can accommodate Trp, Phe or Lys. These results agree with the HB analysis, which predicts that the sidechains of residues that are on the N-terminal side of the cleavage site (P sites) contribute more to binding than C-terminal, P′ sites. The most striking difference between substrates and designed peptides is in this P′ region, where the predicted binding energy contributions for the designed peptides exceed those of the substrates, an advantage that is distributed over most of the designed P′ positions.



	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′
p12	K	Y	T	F	W	Q	Y	S	Q	F	Y
p13	K	Y	L	T	W	Q	N	S	Q	I	N
p14	N	S	V	T	F	W	I	S	Q	F	Q
p15	L	T	I	N	W	Q	K	Y	F	N	T
p16	W	F	T	L	K	Q	Y	W	Q	T	N

Figure 3.10: BAlaS-guided design of tight-binding peptides. a) Propensity scale of each amino acid to form an α-helical peptide conformation. b) Sequences of designed peptides p12-p16. Scatter plots with predicted BAlaS $\Delta\Delta G = \Delta G_{Ala} - \Delta G_{wt}$ values on substitution to alanine for each residue of c) the 11 M^{pro} natural substrates and d) designed peptides based on these. The more positive the value, the greater the contribution made by the sidechain to the overall binding energy. e) The BAlaS $\Delta\Delta G_{sum}$ comparing values between complexes of M^{pro} with substrate and designed peptides as a proxy for predicting relative binding affinity (larger score = tighter binder).

The final step in design was to assess the relative binding affinities of the substrates and designed peptides. Hence the summed ΔΔGs (**Figure 3.10e**) provide a proxy for the binding energies (BAlaS)[62] for the substrates and designed peptides with M^{pro}. The substrate-M^{pro} complexes are stabilised by an average of 46.5 kJ mol^{-1}, whereas the designed peptide-M^{pro} complexes are predicted to have, in some cases, double the interaction stability of the substrates, with an average of 96.0 kJ mol^{-1}.

Synthesis and analysis of designed peptides

To test the designed sequences, p12, p13, p15 and p16 were synthesised with a carboxyl-amide C-terminus by solid phase synthesis. Their M^{pro} inhibitory activity was determined by dose-response analysis (**Table 3.1**) using SPE MS, monitoring both substrate s01 (1191.68 Da) depletion and N-terminally cleaved product (617 Da) formation. Ebselen which reacts multiple times with M^{pro}[63] was used as a standard (IC_{50} = 0.14 ± 0.04 μM; **Figure 3.11**).

Table 3.1: Designed peptides inhibit M^{pro} in a dose-dependent manner. The assay conditions were 0.15 μM M^{pro}, 2 μM s01 in 20 mM HEPES, pH 7.5, and 50 mM NaCl.

Peptide	*IC_{50} / μM*	*Hill Slope*
p12	5.36 ± 2.17	1.25 ± 0.06
p13	3.11 ± 1.80	0.94 ± 0.09
p15	5.31 ± 1.08	1.17 ± 0.16
p16	3.76 ± 0.51	1.19 ± 0.16

All four designed peptides manifested similar potency with IC_{50} values ranging from 3.11 μM to 5.36 μM (**Table 3.1** and **Figure S3.3.1**). Strikingly, despite the presence of Gln at P1 in all the designed peptides assayed, no evidence for hydrolysis was observed by SPE MS. This observation was supported by LCMS of the peptides incubated overnight with M^{pro} (**Figure S3.3.2**). We probed the inhibition mode of the designed peptides by monitoring changes in IC_{50} while varying the substrate concentration (2 μM, 10 μM, 20 μM and 40 μM TSAVLQ↓SGFRK-NH_2 s01; K_m ~14.4 μM).[63] The results indicated a linear dependency between substrate concentration and IC_{50} values (**Figure 3.11a-d**). This was not observed with a control 15-mer peptide or ebselen (**Figure 3.11e, f**). Analysis of the data by the procedure of Wei *et al*.[64] implies competitive

inhibition (**Figure S3.3.3** and **Tables S3.3.1-2**). By contrast, the same analysis for ebselen did not support competitive inhibition, consistent with MS studies showing it has a complex mode of inhibition.[63]

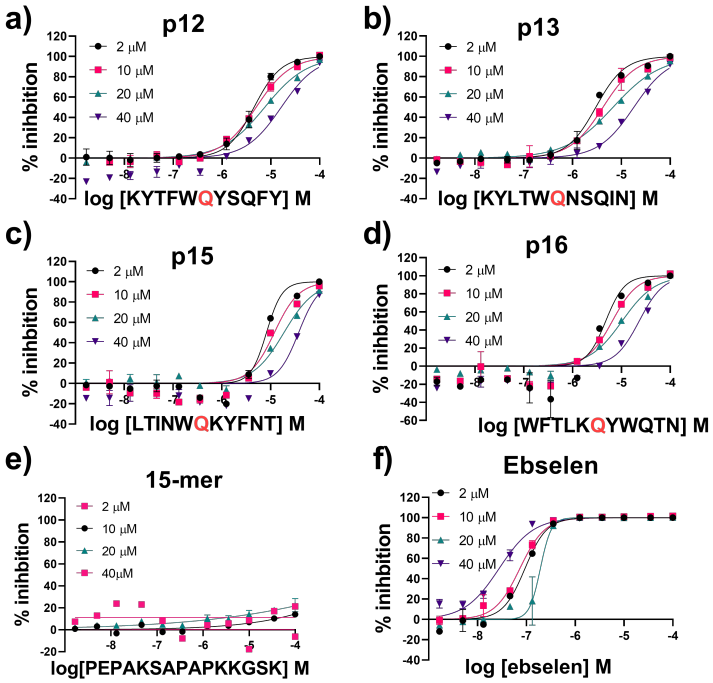


Figure 3.11: IC_{50} of designed peptides against M^{pro} with varied substrate concentrations. IC_{50}s for a) p12, b) p13, c) p15, d) p16, e) 15-mer control peptide and f) ebselen with 2 μM, 10 μM, 20 μM and 40 μM of substrate peptide s01. IC_{50} values were calculated from technical duplicates (**Table S3.3.2)**. See **SI Experimental Section S3.1e** for assay details.

Three of the synthesised peptides — p12, p13, and p15 — have a Trp at P2 (**Figure 3.10b)** while the other, p16, has a Lys at P2. The 11 M^{pro} substrates all have hydrophobic residues (Leu, Val or Phe) at P2 (**Figure 3.1a**). To investigate if the nature of the hydrophobic P2 residue, or the hydrophilic nature of the Gln at P1, alters the interaction of the peptide and hence its reactivity at the active site, we synthesised p13-WP2L, s01-LP2W, and s01- QP1W. There was no evidence for cleavage of p13-WP2L or s01-QP1W. However, s01-LP2W underwent partial cleavage (12.6 ± 4.5) % after overnight incubation. These results suggest that the presence of a Trp at P2 hinders catalytically productive binding, at least with these peptides, and that other residues (including the P1′ and P2′ residues) play roles in orienting the substrates for cleavage (*vide infra*).

We then used non-denaturing protein MS to study enzyme-substrate/product/inhibitor complexes simultaneously with turnover. Complexes between M^{pro} dimer and p12 and p13 were observed, together with the uncomplexed M^{pro} dimer in the protein region of the mass spectra. No binding was observed for p15 and p16, due to relatively high noise in that *m/z* region. None of the designed peptides were cleaved by M^{pro}, as recorded in the peptide region. As a control, s01 was added to the protein/inhibitor mixtures; for all the inhibitors, turnover of s01 was observed after 3-min incubation. Depletion of s01 was 95%, 91%, 70% and 78% in the presence of p12, p13, p15 and p16, respectively, with an 8-fold excess of inhibitor over M^{pro}, *versus* >98% depletion for the M^{pro}/s01 mixture without the inhibitor. In the protein region of the mass spectra, complexes between M^{pro} dimers and the s01-cleavage products were observed in the presence of p13, but the abundance of these complexes was lower than the abundance of M^{pro}/p13 complexes (**Figure 3.12**). These results validate the above-described evidence that the peptide inhibitors both bind and competitively inhibit M^{pro}.

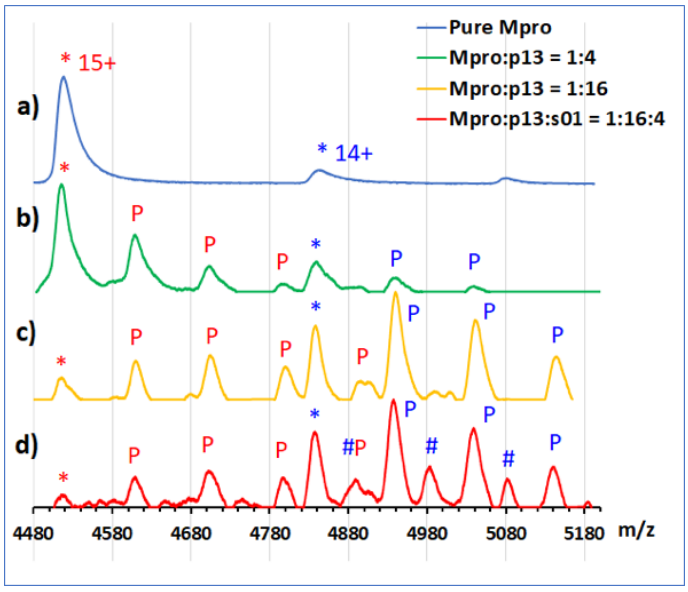


Figure 3.12: Non-denaturing MS analysis of designed peptides binding to the M^{pro} dimer. Inhibitor binding from non-denaturing MS showing normalised intensity in the *m/z* region around the 14+ and 15+ charge states of the M^{pro} dimer. (*) indicates unbound M^{pro} dimer. a) 5 µM M^{pro} solution; b) 4-fold excess of p13 relative to the M^{pro} dimer; 'P' indicates sequential binding of p13 peptides to M^{pro} in the 15+ charge state (red) and 14+ state (blue); c) 16-fold excess of p13; d) 16-fold excess of p13 and 4-fold excess of s01; hash (#) indicates sequential binding of s01-cleavage products (note: the resolution is not sufficient to distinguish between the N- and C-terminal fragments; some non-specific binding of p13 is also observed in c) and d) due to the high concentration of the peptide).

Understanding the basis of SARS-CoV-2 M^{pro} inhibition by the designed peptides

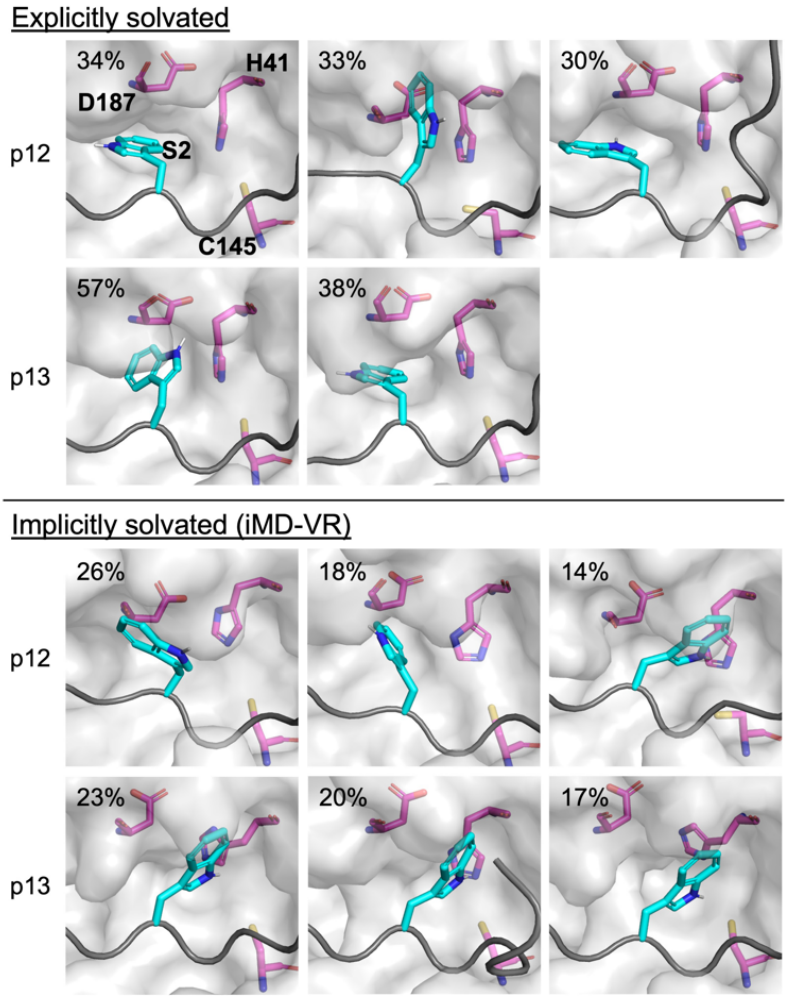


Figure 3.13: Binding of P2 Trp in the designed peptides. Conformations adopted by the P2 Trp sidechain (cyan sticks; non-polar hydrogens omitted for clarity) in p12 and p13 (grey ribbon) observed during explicit and implicit solvent MD simulations, showing representative structures obtained by RMSD clustering. His41, Cys145 and Asp187 are shown in magenta. See **Figure S3.3.13** for cluster populations formed during MD. For each peptide, conformations are displayed in decreasing order of occurrence (above 10%).

Modelling of the designed peptides. Modelling of p12 and p13 shows that both bind stably at the active site during MD simulation (**Figures S3.3.4-8**). Like the natural substrates, key HBs form with Glu166, Thr26, Thr24, and the oxyanion hole- contributing Cys145 (**Figures S3.3.9-11**). However, HBs involving the P1-Gln sidechain of p12 and p13 showed greater variability. The favourability of the P2 Trp mutation, as predicted by the BAlaS scores, prompted us to investigate its binding. In line with the plasticity observed at S2, a variety of conformations are observed

during MD simulations at this position, showing varying degrees of immersion in S2 (**Figures 3.13** and **S3.3.12-13**). Similar results were obtained using iMD-VR (**Figure 3.13**).

Analysis of the conformations of the most populated cluster from MD using Arpeggio-generated hydrophilicity maps (**Figure S3.3.14**) reveals that the P2 Trp is more deeply buried within S2 than the native P2 residues in the natural substrates, forming more than double the number of hydrophobic contacts in the cases of p12 and p13. Some conformations involve the indole ring π-π-stacking, or hydrogen-bonding via its indole N-H, with the catalytic His41 sidechain, forming an extended HB network (**Figure 3.13**). It is possible that these interactions may hamper the ability of His41 to deprotonate Cys145 at the start of peptide hydrolysis, which could be tested using QM/MM calculations. Interestingly, DFT-based interaction analysis reveals that one of the slowest turnover substrates, s05 (**Figure 3.6**), and inhibitor p13 (**Figure 3.14**), share similar short-range interaction networks.

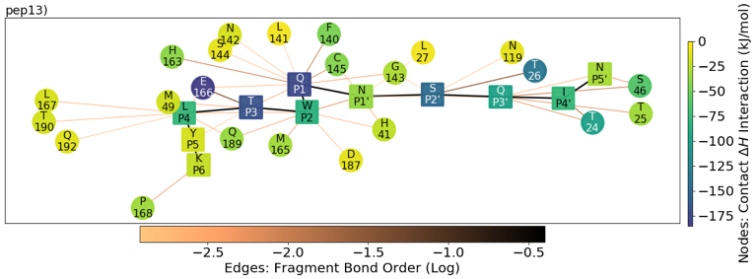


Figure 3.14: **QM contact interaction graph for p13 and M^{pro}.** Interactions are computed using ensemble-averaged results of MD snapshots with the BigDFT code.[50]

Comparative peptide docking. To investigate the ability of the M^{pro} subsites to recognise residues in the designed sequences, AutoDock CrankPep (ADCP) was used (**Table S3.3.3**).[65] A trial was performed by redocking s01 into the H41A SARS-CoV M^{pro} structure originally complexed with s01 (PDB 2Q6G).[37] ADCP successfully positioned the peptide mostly correctly in its top solution, with the Cα positions from P5 to P1′ deviating by less than 1 Å (**Table S3.3.4** and **Figure S3.3.15**). Deviations increased up to 16 Å at P5′ as the peptide coiled up in the P′ positions, but this is deemed acceptable since the S′ subsites are less well defined, as discussed earlier.

Following the promising redocking results with ADCP, s01-s11, p12, p13, p15, and p16 were docked with an M^{pro} structure originally complexed with the N3 inhibitor (PDB 7BQY; 1.7 Å resolution).[2] Substrate-docked structures were found to have the P4 and P2 residues correctly positioned in their corresponding S4 and S2 pockets (**Table S3.3.5**). From P1 to P5′ the poses were more variable, with some peptide backbones turning through S1 rather than continuing an extended conformation, likely due to the less well-defined S′ subsites (**Figure S3.3.16**). For the designed peptides, by contrast, docking appeared less successful (except p16), with none of the top 10 solutions positioning the peptide in the manner observed in our MD simulations (**Figure S3.3.17**). The S2 pocket in PDB 7BQY binds the Leu sidechain of N3 and is probably too shallow to accommodate the larger Trp sidechain, given the assumption of a rigid receptor in ADCP docking. Hence, the four designed peptides were also docked to the C145A M^{pro} structure in complex with the s02 cleaved product (PDB 7JOY; 2 Å resolution),[66] which has a deeper S2 pocket that binds the P2 Phe sidechain in s02. Interestingly, for both p12 and p16, the top docked solution matched our design more closely (**Table S3.3.6** and **Figure S3.3.18**). Docking of p13 and p15 was challenging, possibly due to the difficulty of recognising a larger Leu (p13) or Ile (p15) residue in the S4 pocket, which originally accommodated a Val sidechain. This highlights the ability of the M^{pro} active site to adapt when binding to different substrates or inhibitors.

Summary – designed peptides

We used *in silico* Predictive Saturation Variation Scanning to design peptides that were shown *in vitro* to inhibit M^{pro} competitively. Structures of p12 and p13 generated by both iMD-VR docking and comparative modelling behaved similarly, in terms of HB formation and peptide backbone root mean square deviation (RMSD) and fluctuation (RMSF), when performing MD. These studies highlight how the S2 subsite can adapt its size and interaction network via induced fit to accommodate different substrate or inhibitor P2 residues.

Notably, while these models suggested similarly stable binding modes as seen with the natural substrates, turnover of the inhibitor peptides by M^{pro} was not detected. This may relate to the more favourable predicted binding affinity of the designed peptide-M^{pro} complexes, both in terms of higher overall interaction energies, and greater contribution of the P′ residues than in the natural substrates. Our MD simulations suggest it is also possible that the larger P2 residue prevents the catalytically vital His41 from adopting a reactive conformation (**Figure 3.13**).

4. M^{pro}-Ligand Interaction Analysis

Having elucidated how M^{pro} recognises its substrates and our designed peptide inhibitors, we hypothesised that this might be reflected in the extensive small-molecule inhibitor work on M^{pro} and could, in turn, be exploited for the design of novel small-molecule inhibitors and peptidomimetics. We explored whether ligands sharing the same contacts as the natural substrates could lead to better inhibitory activity. We analysed all 91 X-ray structures of small molecule fragments complexed with M^{pro} obtained by high-throughput crystallographic screening at Diamond's XChem facility,[35] as well as the dataset of 798 designed inhibitors and 245 crystal structures obtained from the COVID Moonshot project.[36] We analysed them by investigating their protein-ligand interaction patterns.

Interaction analysis of XChem fragments

We separated fragments into non-active-site binders (25 fragments) and active-site binding/likely-substrate competing molecules (66 fragments; **Figure S3.4.1**). A fingerprint bit-vector was constructed for every active-site binding fragment, with each bit denoting the presence or absence of a given interaction with M^{pro} residues, and used for clustering fragments by their interaction fingerprint Tanimoto similarity,[67] with 1 corresponding to identical contacts, and 0 to no shared contacts (**Figure 3.15**; **Table S3.4.1**; **Figures S3.4.2**, **3**, **5**). All the fragments and ligands in

clusters 1 and 2 (except x0397, x0978 and x0981) are covalently bound to Cys145. As a result, a highly conserved binding mode is observed for the carbonyl-containing covalent warheads (*e.g.*, chloroacetamides), where the carbonyl oxygen binds into the oxyanion hole between residues Gly143 and Cys145, mimicking substrate HBs 8 and 9 (**Figure 3.3**). Cluster 5 stands out as the only major cluster with fragments that bind deeply into S1, one of the main conserved contacts identified in all substrates. Cluster 5 shows a distinct binding motif primarily driven by: (i) hydrogen bonding between a carbonyl oxygen on the fragment and the Glu166 backbone NH-group; and (ii) a strong polar interaction between His163 and the fragment. Notably, the protonation of the imidazole of His163 appears to depend on the fragment.

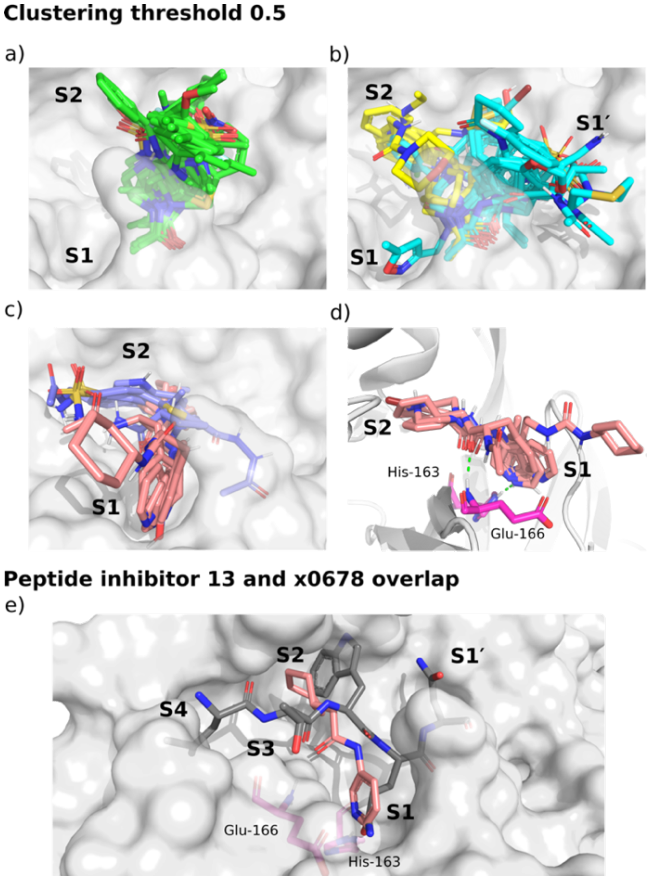


Figure 3.15: Clustering of XChem active site-binding fragments. Surface of the x0830-bound M^{pro} structure (white surface) and the top 5 most populated fragment clusters using a clustering threshold of 0.5. a) Cluster 1 fragments tend to occupy S1′ (green); b) clusters 2 (cyan) and 3 (yellow) tend to span S1′ and S2; c) clusters 4 (lilac) and 5 (pink) tend to occupy S2 and S1. d) Close-up of cluster 5. Green dotted lines indicate the two key HBs between the fragment carbonyl oxygen and the backbone nitrogen of Glu166 (HB 3, **Figure 3.3**), and between the His163 Nε and the heterocyclic nitrogen of the fragment (HB 6, **Figure 3.3**). e) Overlay of the P4-P1′-truncated structure of peptide inhibitor p13 (grey) from an MD snapshot and cluster 5 binder x0678 (pink), with the x0678 co-crystal structure (white surface).

Overall, the primary functionality that facilitates interaction with His163 is the nitrogen-containing heterocycle present in almost all ligands in cluster 5 (**Figure 3.16**); the exception is x0967, which forms the His163 HB via its phenol oxygen. Such heterocycles are well suited to replace the substrate P1 Gln sidechain by mimicking its HB donor/acceptor abilities. In addition, most cluster 5 binders also extend into the hydrophobic S2 pocket, although there is no clear preference in functional group at S2. This agrees with our plasticity analysis, which shows that S2 can accommodate a large variety of functional groups (**Figure S3.4.4**). As seen in the overlap of peptide inhibitor p13 and cluster 5 representative x0678 (**Figure 3.15e**), the binding modes of both inhibitors in the S1 and S2 subsites are very similar, with both forming HBs to His163 (HB 6) and Glu166 (HBs 2 & 3) and binding deep in the S2 pocket. In addition, all cluster 5 ligands (**Figure 3.16**) contain an amide or urea linker between the P1 and P2 binding groups making them interesting building blocks for the development of peptidomimetics.

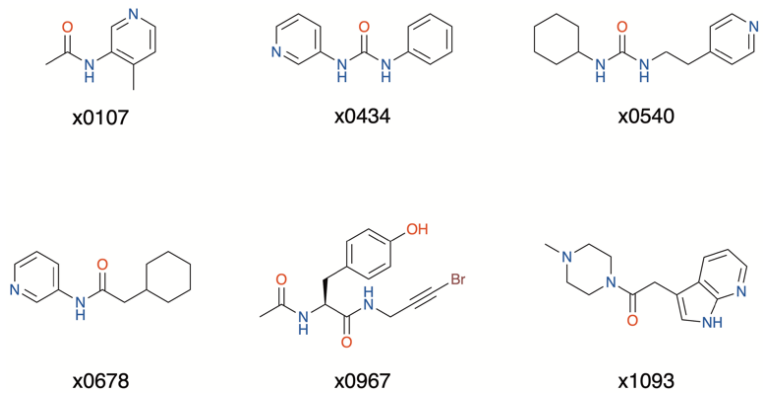


Figure 3.16: Structures of the cluster 5 XChem compounds. Note the prevalence of nitrogen-containing heterocycles, and the phenol-containing derivative x0967.

The interactions between the fragments, substrates, and peptide inhibitors with M^{pro} were analysed by employing linear-scaling DFT. Using short-range (E_{cont}) DFT interactions with M^{pro} as a "descriptor" for clustering, a cluster containing both the substrates and the cluster 5 compounds (x0107, x0434, x0540, x0678, x0967 and x1093) was identified (**Figure S3.4.6**). This cluster also included compounds x0426, x0946, x0195, x0995, x0104, x0874, x1077, x0161 and x0397. This agnostic analysis, based on quantum mechanical descriptors, provides further

confirmation and a powerful alternative to evaluate compounds of differing sizes in biomolecular complexes.

To test whether cluster 5 inhibitors are promising building blocks for optimization, we identified all assayed and crystallised cluster 5 binders in the COVID Moonshot project database[36] as of 11th Jan 2021 and analysed them using Arpeggio. Compounds were deemed cluster 5 inhibitors if they shared at least 70% of the contacts identified in fragment cluster 5. We observed that cluster 5 inhibitors have a significantly higher proportion of "strong" binders, classified as $IC_{50} < 99$ μM (85% of cluster 5 compounds), unlike the rest of the Moonshot project database (67%). Closer analysis can be found in **SI Section S3.4b** and **Figure S3.4.7**. In summary, based on Arpeggio and BigDFT contact analysis and reported assay data, cluster 5 binders are promising building blocks for substrate-competing inhibitor design.

Covalent docking of COVID Moonshot compounds

To accommodate induced fit and create high quality poses of covalent inhibitors for future optimisation, we selected 540 covalently-reacting compounds from 10,001 Moonshot-designed compounds and docked them using AutoDock4 [68] into the M^{pro} structure of the corresponding covalent "inspiration" fragments.[36] We generated an interaction Tanimoto distance matrix as described earlier, and analysed the ability of the procedure to recapitulate the binding mode of the parent fragment. The normalised shape and pharmacophoric overlap (SuCOS[69]) of the lowest energy pose of the highest populated cluster for each Moonshot compound was compared with the inspiration covalent XChem fragment (**Figure S3.4.8**). When controlling for the smallest maximum common substructure (MCS) that encompasses the covalent warhead and one additional atom in the compound, 379 designs remain, 132 (34.8%) of which adopted the binding mode of the inspiration fragment. Given the high similarity between the fragments and docked designs, it is likely that these binding modes are more representative of the actual binding mode. A summary of the workflow is shown in **Figure 3.17**.

In summary, our covalent docking method is more likely to identify the correct binding mode when substantial overlap exists between the inspiration fragment and designed compound beyond the covalent warhead (**SI Section S3.4c** and **Figures S3.4.8-9**). This generated 132 high quality docked poses which serve as inspiration for future inhibitor design and were used in our proposals for compound derivatisation in **SI Section 3.4c**. All poses of the 540 docking runs are available at https://github.com/gmm/SARS-CoV-2-Modelling.

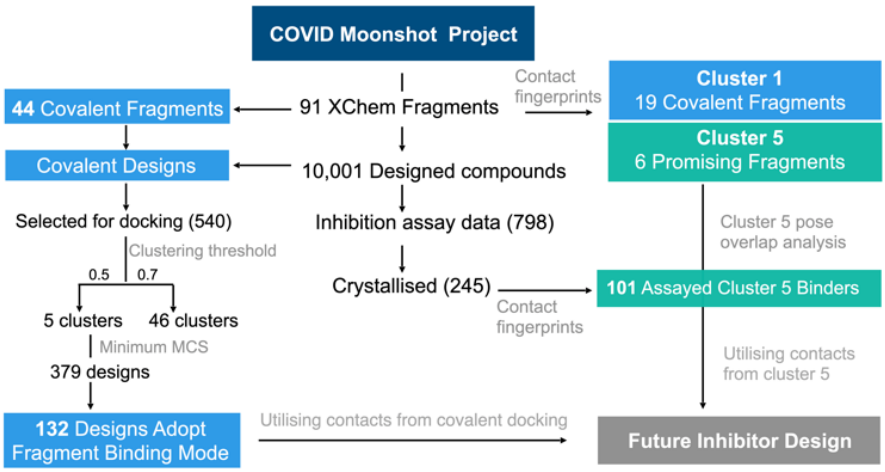


Figure 3.17: Analysis of fragment and designed compounds from the Moonshot project. Workflow used to identify promising fragments and guide novel designs.

Implications for future inhibitor design

We compared the interactions of the cluster 5 binders with those in the substrates, peptide inhibitors, and XChem fragments. Interestingly, unlike the peptides, almost none of the cluster 5 binders interact with the oxyanion hole. The only cluster 5 compounds where this contact is made are a series of covalent inhibitors, none of which showed promising potency (**Figure S3.4.11**). An exhaustive search of Moonshot structures showed that at the time of the analysis, no non-covalent inhibitor has ever been tested that includes both the typical cluster 5 binding mode while also being able to interact with the oxyanion hole.

We compared the structures of the top 10 compounds in cluster 5 (part of the dataset analysed above in "**Interaction analysis of XChem fragments**") to the docked structures of covalent Moonshot designs (above in "**Covalent docking of COVID Moonshot compounds**"). Two compounds—FOC-CAS-e3a94da8-1 and MIH-UNI-e573136b-3—were selected based on their high normalised SuCOS overlap with their inspiration fragments, strongly suggesting that their docked binding modes reflect the actual poses.[70] Both compounds bind into the oxyanion hole as well as into S1 and S2, providing a clear opportunity for extension of the cluster 5 binders **(Figure S3.4.12)**.

Most cluster 5 binders place the aromatic heterocycle into the S1 site and the carbonyl oxygen of the amide linker bonds to Glu166 (**Figure 3.18**). The position of this amide nitrogen overlays perfectly with the ring amine present in the docked compound FOC-CAS-e3a94da8-1. Thus, extension of cluster 5 binders into the oxyanion hole could be achieved by adding a substituent at the amide nitrogen. A promising candidate for extension is x10789, which makes a HB with the backbone oxygen of Glu166 (**Figure 3.18a**) and mimics the non-prime side binding mode of peptide inhibitor p13 (**Figure 3.15e**), even binding into the S4 site via its β-lactam ring (**Figure 3.18a**). Additional expansion into the oxyanion hole and S1′ through the amide linker could yield a powerful peptidomimetic inhibitor, combining protein interactions observed for the substrates, peptide inhibitors and small molecule fragments.

When comparing interactions exhibited by cluster 5 binders (Glu166, His163) or covalent fragments (Gly143, Cys145) with the contacts present in the docked structure of the recently published Phase 1 clinical trial candidate PF-07321332 by Pfizer (**Figure 3.18b**),[11, 12] a nearly identical interaction pattern to the cluster 5 binding motif is observed. However, note that for reacted PF-07321332, AutoDock4 was unable to place the negatively charged azanide nitrogen in the oxyanion hole, which is the expected position given its similarity to related warheads previously docked (**Figure S3.4.9**).

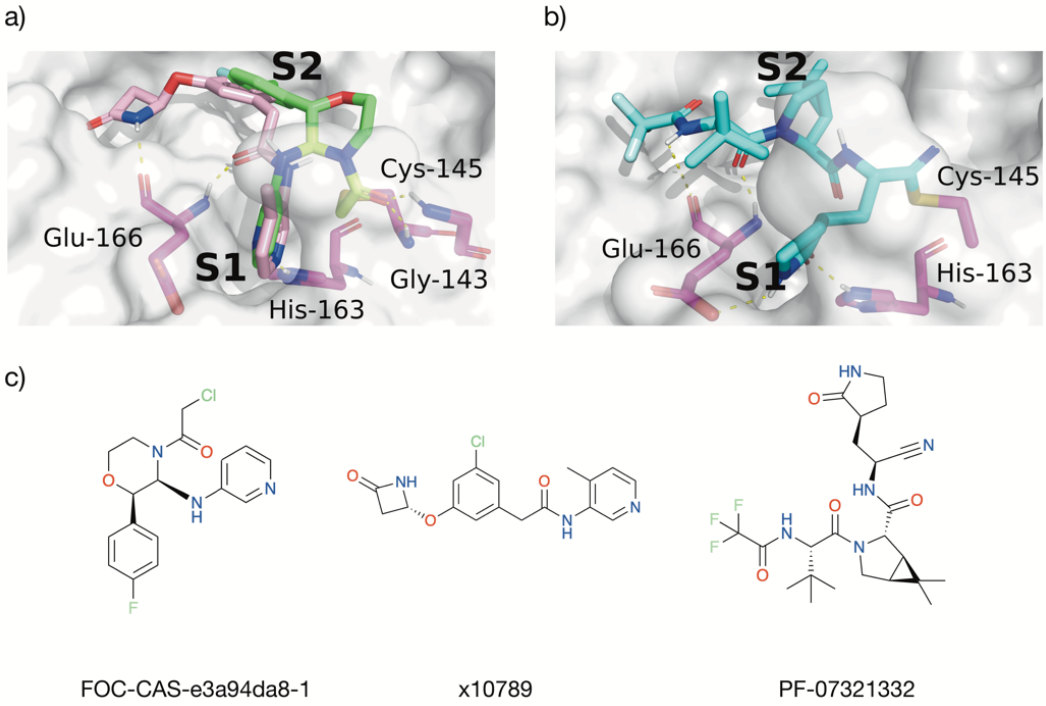


Figure 3.18: Docking informs novel inhibitor design. HBs between M^{pro} (magenta) and the ligands are shown as dotted yellow lines. a) Overlay of the docked pose of FOC-CAS-e3a94da8-1 (green and greenish-yellow) with the crystal structure of x10789 (pink) on the M^{pro} surface (PDB 5RER; 1.88 Å resolution).[35] Derivatisation of x10789 into the oxyanion hole could be achieved by attaching a methylene amide group present in x0830 (highlighted greenish-yellow). b) Docked pose of Pfizer's Phase I covalent inhibitor PF-07321332, covalently docked into M^{pro} (PDB 6XHM; 1.41 Å resolution).[71] PF-07321332 (cyan) is covalently attached to Cys145. The docked PF-07321332 adopts the same major contacts as the 'combination' of x10789 and x0830, namely the double HB to the backbone of Glu166, the HB to His163 in the S1 subsite, and a series of hydrophobic interactions in the S2 subsite. c) Structures of Moonshot designed compound FOC-CAS-e3a94da8-1, crystallographic fragment x10789, and inhibitor PF-07321332.

5. Conclusions

A wealth of crystal structures of SARS-CoV-2 M^{pro} is available, including hundreds with ligands. There is thus the question of how best to use this static information to help develop M^{pro} inhibitors optimised in terms of efficacy and safety for COVID-19 treatment. The dimeric nature of M^{pro}, coupled with its multiple substrates, makes it challenging to understand the structural and dynamic features underpinning selectivity and catalysis, as is the case with many other proteases. Such an understanding is, of course, not essential to develop medicines, as shown by work with other viral proteases. However, it may help improve the quality of such medicines and the efficiency with which they are developed. It will also lay the foundation for tackling anti-COVID-

19 drug resistance — a challenge we will likely encounter as experience with the HIV global pandemic implies. The scale of global efforts on M^{pro} makes this system an excellent model for collaborative efforts linking experimental biophysics, modelling, and drug development (**Figure 3.19**).

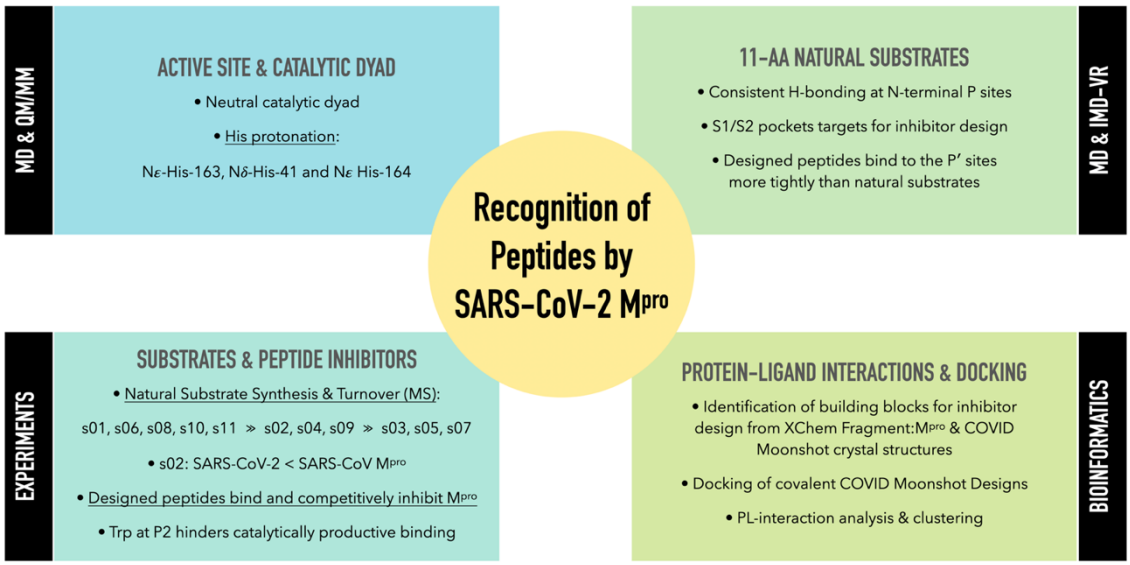


Figure 3.19: Summary of the research and results obtained in this work.

The results of our combined computational studies, employing classical molecular mechanics and quantum mechanical techniques, ranging from automated docking and MD simulations to linear-scaling DFT, QM/MM, and iMD-VR, provide consistent insights into key binding and mechanistic questions. One such question concerns the protonation state of the 'catalytic' His41/Cys145 dyad, an important consideration in the rates of reaction of covalently linked M^{pro} inhibitors which ultimately relates to their selectivity and potency. Our results indicate that a neutral catalytic dyad is thermodynamically preferred in M^{pro} complexed with an unreacted substrate, justifying the neutral state for MD simulations. A more reactive thiolate anion may be deleterious to the virus, as it will be susceptible to reaction with electrophiles. Importantly, analysis of the active site suggests that the precise mechanism of proton transfer in the His41/Cys145 dyad involves dynamic interactions with other residues, including His163, His164,

Asp187, and a water hydrogen bonded to the latter two residues and His41. Proton transfer may be considered a relatively simple part of the overall catalytic cycle — these results thus highlight how M^{pro} catalysis is likely a property of (at least) the entire active site region, with a future challenge being to understand motions during substrate binding, covalent reaction, and product release.

The models we have developed of M^{pro} in complex with its 11 natural substrates provided a basis for analysis of key interactions involved in substrate recognition and for comparison with (potential) inhibitor binding modes. Notably, the P′ (C-terminal) side of substrates appears to be much less tightly bound than the P (N-terminal) side, where there is remarkable consistency in the hydrogen bonding patterns across the substrates. This difference may in part reflect the need for the P′ side to leave (at least from the immediate active site region) after acyl-enzyme complex formation and prior to acyl-enzyme hydrolysis. The tighter binding of the N-terminal P-side residues suggests these are likely more important in substrate recognition by M^{pro}. This is also reflected in potent inhibitors, such as N3 and peptidomimetic ketoamides,[2, 4] which predominantly bind in these non-prime S subsites. The development of S-site-binding inhibitors may also reflect the nature of the substrates used in screens leading to them, which typically comprise an S-site binding peptide with a C-terminal group enabling fluorescence-based measurement. Our results imply that there is considerable scope for developing inhibitors exploiting the S′ subsites, or both S and S′ subsites, though relatively more effort may be required to obtain tight binders compared to targeting the S subsites.

Consistent with prior studies, our work highlights the critical role of the completely conserved P1 Gln residue in productive substrate binding and analogously in inhibitor binding. However, the nature of the P2/S2 interaction is also important in catalysis. In the natural substrates (**Figure 3.1**), the P2 position is Leu in 9 of the 11 substrates, Phe in s02 (which displays medium turnover efficiency), and Val in s03 (which is a poor substrate). Our results show that the S2 subsite plays a critical role in recognition and inhibition. S2 is highly plastic (**Figures 3.7** and **S3.4.4**) and can

accommodate a range of different sidechains, including larger groups, though not necessarily in a productive manner. The observation that substrates with a P2 Leu vary in efficiency reveals that interactions beyond those involving P1 and P2 are important, reinforcing the notion that (likely dynamic) interactions beyond the immediate active site are important in determining selectivity both in terms of binding and rates of reaction of enzyme-substrate complexes.

Notably, the results of computational alanine scanning mutagenesis followed by design, aimed at identifying peptides that would bind more tightly than the natural substrates, led to the finding that substitution of a Trp at P2 ablates hydrolysis creating an inhibitor. The observations with peptide inhibitors of M^{pro} have precedence in studies with other nucleophilic proteases, including the serine protease elastase, showing that substrate substitutions away from the scissile P1/P′ residues can cause inhibition.[72, 73] There is thus scope for the extensive development of tight binding peptidic and peptidomimetic M^{pro} inhibitors for use in inhibition and mechanistic/biophysical studies, with the Trp at P2 of the peptide inhibitors being a good point for SAR exploration, potentially by (i) replacement of the indole hydrogen with suitable alkyl or aryl substituents; (ii) introduction of substituents with different stereoelectronic properties at C-2 or C-5 of the indole ring; or (iii) cyclization by the insertion of a methylene group linking position 2 of the indole ring to the α-nitrogen of Trp itself.[74]

Finally, the combined analysis of interactions involved in substrate binding and extensive structural information on inhibitor/fragment binding to M^{pro} enabled us to identify a cluster of inhibitors whose interactions relate to those conserved in substrate binding (*e.g.*, involving the Glu166 backbone, His163 sidechain, and/or the oxyanion hole formed by the Cys145 and Gly143 backbones). Building out from these 'privileged' interactions (**Figure 3.18**) might be a useful path for inhibitor discovery. Indeed, an M^{pro} inhibitor now in clinical trials[11, 12] exploits the same 'privileged' interactions that we identified. We hope the methods and results that have emerged from our collaborative efforts will help accelerate the development of drugs for treatment of viral infections, and particularly COVID-19.

6. Selected Methods

A detailed description of other experimental and computational methods employed in this work is provided in the Supplementary Information in **Appendix for Chapter 3: Section S3.1**.

Comparative modelling of the SARS-CoV-2 M^{pro}-peptide complexes

A crystal structure of the TSAVLQ↓SGFRK 11-mer peptide substrate bound to the H41A mutant of dimeric SARS-CoV M^{pro} (PDB 2Q6G)[37] was superimposed with one of unmodified dimeric SARS-CoV-2 M^{pro} (PDB 6YB7; 1.25 Å resolution).[14] The substrate was transferred over to the chain A active site of the catalytically-competent SARS-CoV-2 M^{pro} structure. The sequences of the 11 native cleavage sites processed by SARS-CoV-2 M^{pro} (s01-s11) were identified by aligning the sequences of the ORF1ab polyproteins of both SARS-CoV (GenBank accession code NC_004718.3)[75] and SARS-CoV-2 isolate Wuhan-Hu-1 (accession code MN908947.3)[40] using MUSCLE.[76] For each of the 11 cleavage sites, atomic models of an 11-mer peptide matching positions P6 to P5′ and charged N- and C-termini were constructed using the mutagenesis tool of the open source version of PyMOL (v. 2.3.0).[77] For every sidechain from positions P6 to P5′, apart from Gly and Ala, the highest-probability backbone-dependent conformer with the least steric clash and the most chemical complementarity was selected.[78] Using CCG MOE version 2019.0104,[79] each of the resulting 11 models of the SARS-CoV-2 M^{pro} dimer complexed with each 11-mer substrate in the A-chain active site underwent structure preparation protonation using Protonate 3D. Each model was then solvated using 0.1 M NaCl and explicit water and subjected to energy minimization using the AMBER10:EHT force field[80, 81] and periodic boundary conditions, until convergence with an RMS of 0.4184 kcal mol^{-1} per iteration was reached.

Explicit-solvent molecular dynamics

Pre-solvation models of the dimeric M^{pro}-peptide complexes constructed as described above (**Comparative modelling**) were used as starting points for MD simulations. All additives and

crystallographic water molecules were removed from PDB 6YB7, except HOH 644 which provides bridges between His41, His164 and Asp187 (see **main text** and **Figures S3.2.5-6** for details). Protonation and rotameric states of histidines (**Table S3.2.1**) and other titratable residues were assigned at pH 7.4 based on a combination of Reduce (MolProbity, Duke University),[82] H++ (Virginia Tech),[83] PROPKA3 (PDB2PQR),[84] and visual inspection, with a final M^{pro} monomeric charge of -4. His41 (protonated on its δ-nitrogen) and Cys145 were assigned neutral.

MD simulations were performed using GROMACS (v. 2019.2)[85] employing the AMBER99SB-ILDN force field.[86] Each of the constructed complexes was solvated (TIP3P water model)[87] in a rhombic dodecahedral box (1.0 nm buffer), neutralised, and minimised using the steepest descent algorithm until the maximum force was below 1000 kJ mol^{-1} nm^{-1}. For each peptide sequence, three independent simulations were initiated by random velocities at 298.15 K. In each case, the system was equilibrated under NVT (200 ps; 1 fs timestep) and NPT (200 ps; 1 fs timestep) conditions, before being subjected to 200 ns MD simulation (2 fs timestep) at 298.15 K and 1 bar, during which protein-peptide interactions were monitored. All simulations were performed with three-dimensional periodic boundary conditions. Long-range electrostatics was calculated using the smooth particle mesh Ewald method.[88] All bond lengths involving hydrogen atoms were constrained with the LINCS algorithm.[89] Hydrogen bonds between M^{pro} and the peptides were monitored over the course of the simulations, defined using a combined criteria on the donor-acceptor distance ($d_{D\text{-}A} \leq 3.5$ Å) and the proton-donor-acceptor angle ($\angle$(H-D-A) $\leq 30°$).

Models of the designed sequences p12 and p13 complexed with SARS-CoV-2 M^{pro} (PDB 6YB7) were built using a comparative modelling approach similar to that described above (**Comparative modelling**), starting from the previously constructed model of the M^{pro}-s02 complex. Each constructed complex was then solvated, minimised, equilibrated, and subjected to 3 × 200 ns MD as described above, except the retention of a backbone restraint during NPT equilibration to allow longer relaxation of the non-native peptide side chains and the M^{pro} binding pockets.

To generate representative structures of M^{pro}-peptide complexes for interaction analysis, frames extracted every ns from the concatenated 3 × 200 ns MD trajectories were fitted using the M^{pro} backbone, before performing clustering based on the heavy-atom RMSD of the peptide, using the *gromos* algorithm as implemented in GROMACS (v. 2019.2).[90] A cut-off of 2.0 Å (for native substrates) or 2.5 Å (for p12 and p13, due to heavier residues in their terminal regions) was used.

MM/GBSA calculations

The contribution of each residue to protein-peptide binding was evaluated quantitatively using per-residue decomposition of binding energy,[48] estimated using the molecular mechanics/generalised Born surface area (MM/GBSA) method as implemented in MMPBSA.py (v. 14.0)[91] in combination with *sander* (Amber18).[92] The single trajectory protocol was employed, with the M^{pro} dimer defined as the receptor and the 11-mer peptide as the ligand. Snapshots were extracted every 5 ns from the 3 × 200 ns MD trajectories for each substrate (120 frames per substrate). The polar solvation term was calculated using the OBC2 model (igb=5) with mbondi2 radii at 0.15 M salt concentration.[93] Non-polar solvation terms were computed from surface area (recursive approximation from icosahedra)[48] and a surface tension of 0.005 kcal mol^{-1} $Å^{-2}$.[93]

Peptide docking

Docking of substrate and inhibitor peptides was performed using AutoDock CrankPep (ADCP) in the ADFRsuite (v. 1.0) package.[65] For redocking trials, the structure of s01-bound H41A SARS-CoV M^{pro} (PDB 2Q6G)[37] chain A was prepared in ADFRsuite as the receptor. For docking to SARS-CoV-2 M^{pro}, the N3 inhibitor-bound (PDB 7BQY; 1.70 Å resolution)[2] or the C-terminal autocleavage site product-bound C145A (PDB 7JOY; 2.00 Å resolution)[66] M^{pro} structure was used. The dimeric M^{pro} structure was used following processing with MolProbity,[82] correction of histidine states (**Table S3.2.1**), and conversion to pdbqt format with ADFRsuite. The most probable peptide-binding site on the receptor surface was predicted with AutoSite (v1.1).[94] While the peptide-binding site was successfully identified by AutoSite with the chain A monomer

extracted from the dimeric structure in PDB 2Q6G, with the monomeric asymmetric unit structure from PDB 7BQY binding site identification by AutoSite was unsuccessful. When the dimeric structure (PDB 7BQY) was inputted, however, the active site was successfully identified. Hence, subsequent docking with SARS-CoV-2 M^{pro} structures was performed with the dimer. Affinity maps on the receptor were calculated with AutoGridFR (v. 1.2).[95] Each docking run was performed with 100 replicas and 11 million steps, starting from the extended peptide conformation, with solutions internally clustered by a native contact threshold of 0.8. The clustered solutions were then evaluated against the binding mode found in the original structure (SARS-CoV) or in the minimised comparative model (SARS-CoV-2). To allow for flexibility in the less tightly bound terminal regions, and to filter out solutions where the peptide positioning was offset by one residue, solutions were assessed on the criteria of < 2 Å deviation in at least three Cα atoms, after fitting to the M^{pro} backbone with VMD (v. 1.9.4).[96] Out of the top 10 poses, the highest-scoring filtered pose, or if none of the poses passed the filter, the pose with the lowest Cα RMSD, was presented.

Data Availability

Structures, input files, underlying data, and source code are publicly available on GitHub at https://github.com/gmm/SARS-CoV-2-Modelling.

Author Contributions

G.M.M. generated the initial 11 natural substrate-SARS-CoV-2 M^{pro} models. R.M.T. carried out QM/MM studies with subsequent contributions from K.Ś. and V.Mo. H.T.H.C. carried out classical MD simulations and ADCP peptide docking. R.K.W. and H.M.D. performed iMD-VR simulations. M.A.M. performed protein-substrate, protein-peptide inhibitor and protein-ligand contact analysis, fragment clustering, COVID Moonshot covalent docking and future inhibitor suggestions. M.A.M. and L.G. performed bioinformatic study of fragment binding to M^{pro}. L.G. performed linear-scaling QM-DFT calculations, and W.D., T.N. and T.J.W. contributed with its setup and analysis. D.K.S. carried out mutagenesis analysis and designed novel peptides (using software devised by R.B.S.). E.S., P.L., C.S.D., C.D.O. and M.A.W. carried out M^{pro} production and purification. T.R.M. and T.J. synthesised and purified natural and designed peptides. T.R.M. also performed kinetic analyses. V.M. undertook non-denaturing MS analyses. V.Mo. and A.L. contributed to discussions. C.J.S., A.J.M., D.K.S., F.D. and G.M.M. conceptualised and supervised the study. H.T.H.C., R.K.W., M.A.M., T.R.M., R.M.T., V.M., C.J.S., D.K.S., L.G., A.J.M., F.D. and G.M.M. wrote the manuscript.

References

1. Johns Hopkins University of Medicine COVID-19 Map. https://coronavirus.jhu.edu/map.html (accessed 2021).
2. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M., *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
3. Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem* **2020,** *21*, 730.
4. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C., *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368*, 409.
5. Xia, B.; Kang, X. Activation and maturation of SARS-CoV main protease. *Protein Cell* **2011,** *2*, 282.
6. El-Baba, T. J.; Lutomski, C. A.; Kantsadi, A. L.; Malla, T. R.; John, T., *et al.* Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angew. Chem. Int. Ed.* **2020,** *59*, 23544.
7. Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M., *et al.* SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2021,** *17*, 222.
8. Zhu, L.; George, S.; Schmidt, M. F.; Al-Gharabli, S. I.; Rademann, J., *et al.* Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antivir. Res.* **2011,** *92*, 204.
9. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016,** *59*, 6595.
10. Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021,** *9*, 622898.
11. Owen, D., Pfizer Oral COVID-19 Antiviral Clinical Candidate. In *American Chemical Society Spring 2021 Meeting*, 2021.
12. Halford, B. C&EN: Pfizer unveils its oral SARS-CoV-2 inhibitor. https://cen.acs.org/acs-news/acs-meeting-news/Pfizer-unveils-oral-SARS-CoV/99/i13 (accessed April 2021).
13. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004,** *14*, 1188.
14. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J., *et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
15. Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem. Sci.* **2020,** *11*, 10626.
16. Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V., *et al.* Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021,** *12*, 1433.
17. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020,** *10*, 12544.
18. Mondal, D.; Warshel, A. Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of α-Ketoamide Inhibitors of the Main Protease of SARS-CoV-2. *Biochemistry* **2020,** *59*, 4601.
19. Acharya, A.; Agarwal, R.; Baker, M. B.; Baudry, J.; Bhowmik, D., *et al.* Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. *J. Chem. Inf. Model.* **2020,** *60*, 5832.
20. Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **2020,** *12*, 581.
21. Loschwitz, J.; Jäckering, A.; Keutmann, M.; Olagunju, M.; Eberle, R. J., *et al.* Novel inhibitors of the main protease enzyme of SARS-CoV-2 identified via molecular dynamics simulation-guided in vitro assay. *Bioorg. Chem.* **2021**, 104862.
22. Abel, R.; Paredes Ramos, M.; Chen, Q.; Pérez-Sánchez, H.; Coluzzi, F., *et al.* Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front. Chem.* **2020,** *8*, 590263.
23. Ghahremanpour, M. M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J. A.; Zhang, C.-H., *et al.* Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **2020,** *11*, 2526.

24. Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.*, et al.* Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021,** *7*, 467.
25. Wang, L. L.; Lo, K.; Chandrasekhar, Y.; Reas, R.; Yang, J.*, et al.* CORD-19: The Covid-19 Open Research Dataset. *ArXiv* **2020**, arXiv:2004.10706v2.
26. Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.*, et al.* The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 13190.
27. Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L. 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism. *Biochemistry* **2004,** *43*, 4568.
28. Tan, J.; Verschueren, K. H. G.; Anand, K.; Shen, J.; Yang, M.*, et al.* pH-dependent Conformational Flexibility of the SARS-CoV Main Proteinase (Mpro) Dimer: Molecular Dynamics Simulations and Multiple X-ray Structure Analyses. *J. Mol. Biol.* **2005,** *354*, 25.
29. Paasche, A.; Zipper, A.; Schafer, S.; Ziebuhr, J.; Schirmeister, T.*, et al.* Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry* **2014,** *53*, 5930.
30. Pavlova, A.; Lynch, D. L.; Daidone, I.; Zanetti-Polzi, L.; Smith, M. D.*, et al.* Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *Chem. Sci.* **2021,** *12*, 1513.
31. Verma, N.; Henderson, J. A.; Shen, J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *J. Am. Chem. Soc.* **2020,** *142*, 21883.
32. Díaz, N.; Suárez, D. Influence of charge configuration on substrate binding to SARS-CoV-2 main protease. *Chem. Commun.* **2021,** *57*, 5314.
33. Paasche, A. Mechanistic Insights into SARS Coronavirus Main Protease by Computational Chemistry Methods. Julius-Maximilians-Universität Würzburg, Würzburg, 2013.
34. Kneller, D. W.; Phillips, G.; Weiss, K. L.; Pant, S.; Zhang, Q.*, et al.* Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* **2020,** *295*, 17365.
35. Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.*, et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020,** *11*, 5047.
36. Achdout, H.; Aimon, A.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.*, et al.* COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning. *bioRxiv* **2020**, DOI: 10.1101/2020.10.29.339317.
37. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.*, et al.* Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **2008,** *82*, 2515.
38. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.*, et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696.
39. Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013,** *9*, 338.
40. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.*, et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265.
41. O'Connor, M. B.; Bennie, S. J.; Deeks, H. M.; Jamieson-Binnie, A.; Jones, A. J.*, et al.* Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. *J. Chem. Phys.* **2019,** *150*, 220901.
42. Deeks, H. M.; Walters, R. K.; Hare, S. R.; O'Connor, M. B.; Mulholland, A. J.*, et al.* Interactive molecular dynamics in virtual reality for accurate flexible protein-ligand docking. *PLOS ONE* **2020,** *15*, e0228461.
43. Deeks, H. M.; Walters, R. K.; Barnoud, J.; Glowacki, D. R.; Mulholland, A. J. Interactive Molecular Dynamics in Virtual Reality Is an Effective Tool for Flexible Substrate and Inhibitor Docking to the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020,** *60*, 5803.
44. Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.*, et al.* Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J. Biol. Chem.* **2004,** *279*, 1637.
45. Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montaño, B.; Pitt, W. R.; Ascher, D. B.*, et al.* Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J. Mol. Biol.* **2017,** *429*, 365.

46. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401.
47. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.*, et al.* Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889.
48. Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003,** *330*, 891.
49. Metz, A.; Pfleger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K.-H.*, et al.* Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein–Protein Interface. *J. Chem. Inf. Model.* **2012,** *52*, 120.
50. Ratcliff, L. E.; Dawson, W.; Fisicaro, G.; Caliste, D.; Mohr, S.*, et al.* Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *J. Chem. Phys.* **2020,** *152*, 194110.
51. Dawson, W.; Mohr, S.; Ratcliff, L. E.; Nakajima, T.; Genovese, L. Complexity Reduction in Density Functional Theory Calculations of Large Systems: System Partitioning and Fragment Embedding. *J. Chem. Theory Comput.* **2020,** *16*, 2952.
52. Cho, E.; Rosa, M.; Anjum, R.; Mehmood, S.; Soban, M.*, et al.* Dynamic Profiling of β-Coronavirus 3CL Mpro Protease Ligand-Binding Sites. *J. Chem. Inf. Model.* **2021,** *61*, 3058.
53. Diamond Fragalysis. https://fragalysis.diamond.ac.uk/ (accessed January 2021).
54. Silvestrini, L.; Belhaj, N.; Comez, L.; Gerelli, Y.; Lauria, A.*, et al.* The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci. Rep.* **2021,** *11*, 9283.
55. Muramatsu, T.; Takemoto, C.; Kim, Y.-T.; Wang, H.; Nishii, W.*, et al.* SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113*, 12997.
56. Wood, C. W.; Ibarra, A. A.; Bartlett, G. J.; Wilson, A. J.; Woolfson, D. N.*, et al.* BAlaS: fast, interactive and accessible computational alanine-scanning using BudeAlaScan. *Bioinformatics* **2020,** *36*, 2917.
57. Sessions, R. B. BUDE_SM Algorithm. https://github.com/richardbsessions (accessed April 2021).
58. Hetherington, K.; Dutt, S.; Ibarra, A. A.; Cawood, E. E.; Hobor, F.*, et al.* Towards optimizing peptide-based inhibitors of protein–protein interactions: predictive saturation variation scanning (PreSaVS). *RSC Chem. Biol.* **2021,** *2*, 1474.
59. McIntosh-Smith, S.; Price, J.; Sessions, R. B.; Ibarra, A. A. High performance in silico virtual drug screening on many-core processors. *Int. J. High Perform. Comput. Appl.* **2015,** *29*, 119.
60. McIntosh-Smith, S.; Wilson, T.; Ibarra, A. Á.; Crisp, J.; Sessions, R. B. Benchmarking Energy Efficiency, Power Costs and Carbon Emissions on Heterogeneous Systems. *Comput. J.* **2012,** *55*, 192.
61. Pace, C. N.; Scholtz, J. M. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998,** *75*, 422.
62. Ibarra, A. A.; Bartlett, G. J.; Hegedüs, Z.; Dutt, S.; Hobor, F.*, et al.* Predicting and Experimentally Validating Hot-Spot Residues at Protein–Protein Interfaces. *ACS Chem. Biol.* **2019,** *14*, 2252.
63. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
64. Wei, M.; Wynn, R.; Hollis, G.; Liao, B.; Margulis, A.*, et al.* High-throughput determination of mode of inhibition in lead identification and optimization. *J. Biomol. Screen.* **2007,** *12*, 220.
65. Zhang, Y.; Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019,** *35*, 5121.
66. Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.*, et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020,** *11*, 5877.
67. Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytologist* **1912,** *11*, 37.
68. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.*, et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30*, 2785.
69. Leung, S.; Bodkin, M.; von Delft, F.; Brennan, P.; Morris, G. SuCOS is Better than RMSD for Evaluating Fragment Elaboration and Docking Poses. *ChemRxiv* **2019**, DOI: 10.26434/chemrxiv.8100203.v1.
70. Malhotra, S.; Karanicolas, J. When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode? *J. Med. Chem.* **2017,** *60*, 128.

71. Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.*, et al.* Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020,** *63*, 12725.
72. Wright, P. A.; Wilmouth, R. C.; Clifton, I. J.; Schofield, C. J. Kinetic and crystallographic analysis of complexes formed between elastase and peptides from beta-casein. *Eur. J. Biochem.* **2001,** *268*, 2969.
73. Wilmouth, R. C.; Clifton, I. J.; Robinson, C. V.; Roach, P. L.; Aplin, R. T.*, et al.* Structure of a specific acyl-enzyme complex formed between β-casomorphin-7 and porcine pancreatic elastase. *Nat. Struct. Mol. Biol.* **1997,** *4*, 456.
74. Castelli, R.; Tognolini, M.; Vacondio, F.; Incerti, M.; Pala, D.*, et al.* Δ(5)-Cholenoyl-amino acids as selective and orally available antagonists of the Eph-ephrin system. *Eur. J. Med. Chem.* **2015,** *103*, 312.
75. He, R.; Dobie, F.; Ballantine, M.; Leeson, A.; Li, Y.*, et al.* Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem. Biophys. Res. Commun.* **2004,** *316*, 476.
76. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004,** *32*, 1792.
77. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
78. Dunbrack Jr., R. L.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997,** *6*, 1661.
79. *Molecular Operating Environment (MOE)*, 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.
80. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.*, et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179.
81. Gerber, P. R.; Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **1995,** *9*, 251.
82. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.*, et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018,** *27*, 293.
83. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012,** *40*, W537.
84. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525.
85. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
86. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
87. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
88. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.*, et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.
89. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18*, 1463.
90. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.*, et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
91. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.*, et al.* MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314.
92. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T.E. Cheatham, I.*, et al. AMBER 2019*, University of California, San Francisco, 2019.
93. Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004,** *55*, 383.
94. Ravindranath, P. A.; Sanner, M. F. AutoSite: an automated approach for pseudo-ligands prediction—from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics* **2016,** *32*, 3142.
95. Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Comput. Biol.* **2015,** *11*, e1004586.

96. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996,** *14*, 33.

3.3 Statement of authorship

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Discovery of SARS-CoV-2 M^{pro} peptide inhibitors from modelling substrate and ligand binding
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.; John, T.; Deeks, H. M.; Johnston-Wood, T.; Mikhailov, V.; Sessions, R. B.; Dawson, W.; Salah, E.; Lukacik, P.; Strain-Damerell, C.; Owen, C. D.; Nakajima, T.; Świderek, K.; Lodola, A.; Moliner, V.; Glowacki, D. R.; Spencer, J.; Walsh, M. A.; Schofield, C. J.; Genovese, L.; Shoemark, D. K.; Mulholland, A. J.; Duarte, F.; Morris, G. M. Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021**, *12* (41), 13686-13703.

Student Confirmation

Student Name:	Ho Ting Henry CHAN		
Contribution to the Paper	Carried out the classical MD simulations (with hydrogen bond and MM/GBSA analysis) and ADCP peptide docking; wrote the manuscript (as detailed in "**Author Contributions**" section).		
Signature		Date	1 August 2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Fernanda Duarte		
Supervisor comments I confirm the description above regarding authorship and contributions is accurate.		
Signature 	Date	2 August 2023

This completed form should be included in the thesis, at the end of the relevant chapter.

3.4 β-lactam inhibitors

Manuscript VI reports on the development of an SPE-MS based assay to identify inhibitors of SARS-CoV-2 M^{pro}. The assay monitors by MS the depletion of the 11-mer peptide substrate (s01) or the formation of the N- or C-terminal hydrolysed fragments.[1] This approach enabled the identification of M^{pro} inhibitors containing a β-lactam motif, which is found in a range of antibiotics of widespread use with established safety profiles.[2, 3] β-Lactam antibiotics were known to covalently inhibit cysteine proteases[4] and thus could be considered as drug repurposing candidates for targeting SARS-CoV-2 M^{pro}.[5] The most potent β-lactam inhibitors identified in **Manuscript VI** were the two penicillin-derived esters (**Figure 3.20**): penicillin V sulfone with C3-benzyl ester **1** (IC_{50} ~ 1.5 μM) that covalently modifies Cys145 selectively; and penicillin G *R*-sulfoxide with C3-*p*-nitrobenzyl ester and C6-methoxy **2** (IC_{50} ~ 3.5 μM) that does not add covalently to Cys145.[1]

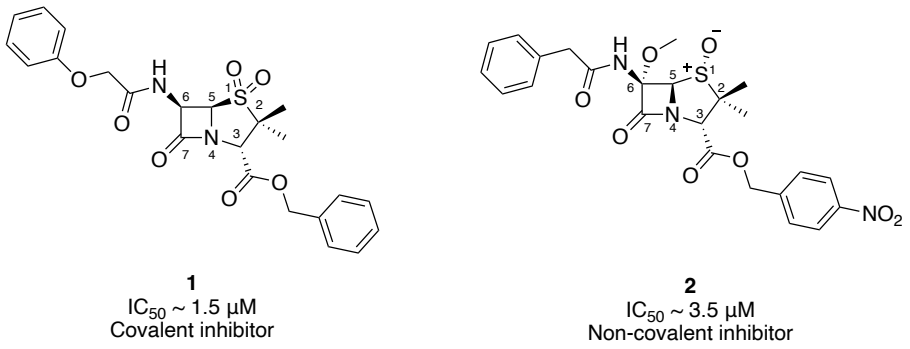


Figure 3.20: The β-lactam inhibitors **1** and **2**.[1] The atom numbering in the bicyclic β-lactam core is used subsequently.

To aid inhibitor design and optimisation efforts, when crystal structures of β-lactam inhibitor bound M^{pro} were not available, I developed a computational workflow to identify the most probable ligand binding modes in the M^{pro} active site. The workflow involves docking, short MD simulations, pose filtering based on binding stability, and identification and calculations of stable M^{pro}-ligand interactions (see **Appendix for Chapter 3: Section S3.6**). Compounds **1** and **2** are presented as case studies in the following.

Interestingly, while only **1** inhibits M^{pro} covalently, docking studies alone suggested that both **1** and **2** can potentially bind in the M^{pro} active site in a manner that enables reaction with Cys145 (**Figure 3.21**), as presented in my contribution to **Manuscript VI**.[1] In these docked poses, named V1 and F2 for both compounds, the β-lactam core sits in the active site with its carbonyl oxygen positioned to hydrogen bond with the backbone NHs of Cys145 and Gly143, which respectively mimic the substrate-conserved HBs 8 and 9 (**Figure 3.3**); these interactions in the oxyanion hole would stabilise the tetrahedral intermediate formed upon addition of Cys145 onto the β-lactam carbonyl C7 atom. The C6 sidechain mimics the P1 glutamine sidechain, with the amide carbonyl oxygen coordinating to His163 as in HB 6 in substrates (**Figure 3.3**). The V1 and F2 poses differ in the positioning of the C3 benzyl or *p*-nitrobenzyl ester group, which occupies S1′ in V1 and S2 in F2.

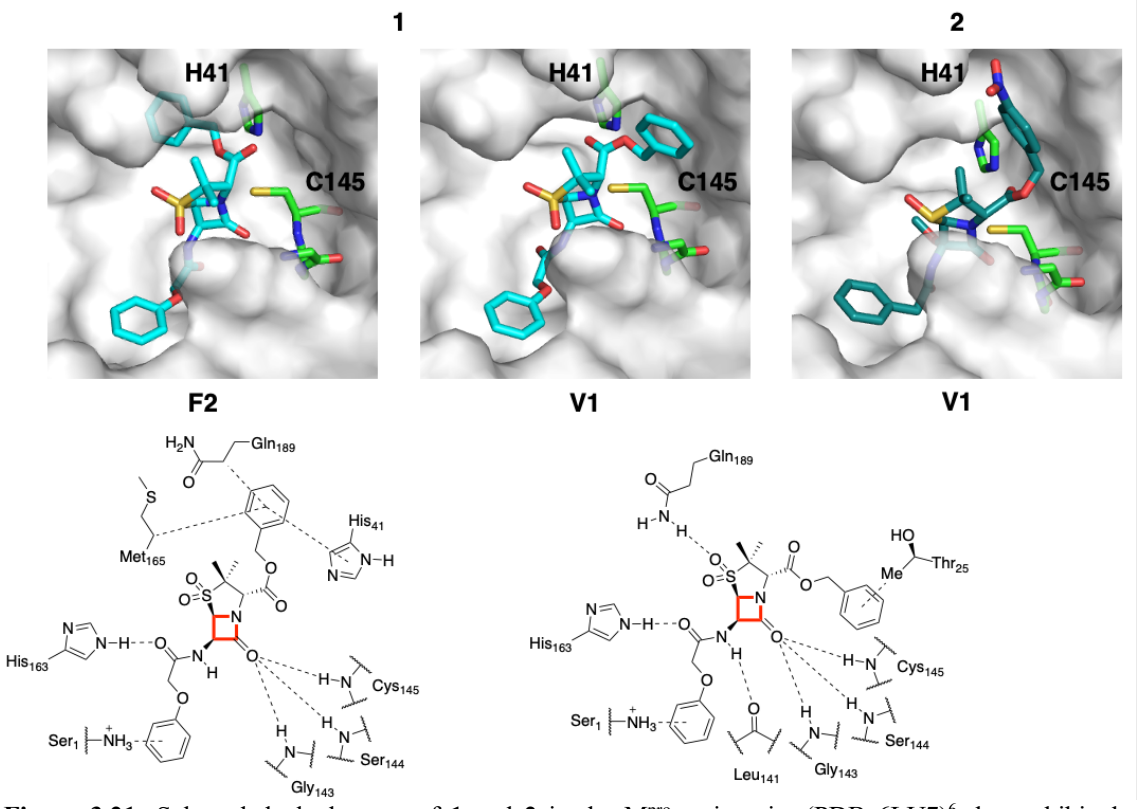


Figure 3.21: Selected docked poses of **1** and **2** in the M^{pro} active site (PDB 6LU7)[6] that exhibit the possibility of reaction with Cys145. Figure adapted from the Supporting Information of **Manuscript VI** (Reference [1]) with permission from the Royal Society of Chemistry.[1]

In subsequent MD simulations, both the F2 and V1 conformations of **1** remained stably bound, maintaining interactions in the oxyanion hole and S1 (**Figure 3.22**). In addition, HBs were observed between the C6 amide NH and the Asn142 sidechain, as well as between the *pro-R*-sulfone oxygen and the Gln189 sidechain and/or the Glu166 backbone, the latter of which mimics the substrate-conserved HB 3. Dispersion interactions were likely contributed by both the C6 phenoxy and the C3 benzyl groups, the latter forming contacts with Thr25 (V1) or Met165 (F2); Met165 forms conserved hydrophobic contacts with the substrate P2 residues (**Figures 3.4, 3.5**).[7] Comparing the two conformations, F2 displayed more favourable vdW interaction energy (**Table S3.7.2**), leading to stronger MM/PBSA estimated binding affinities compared to V1.

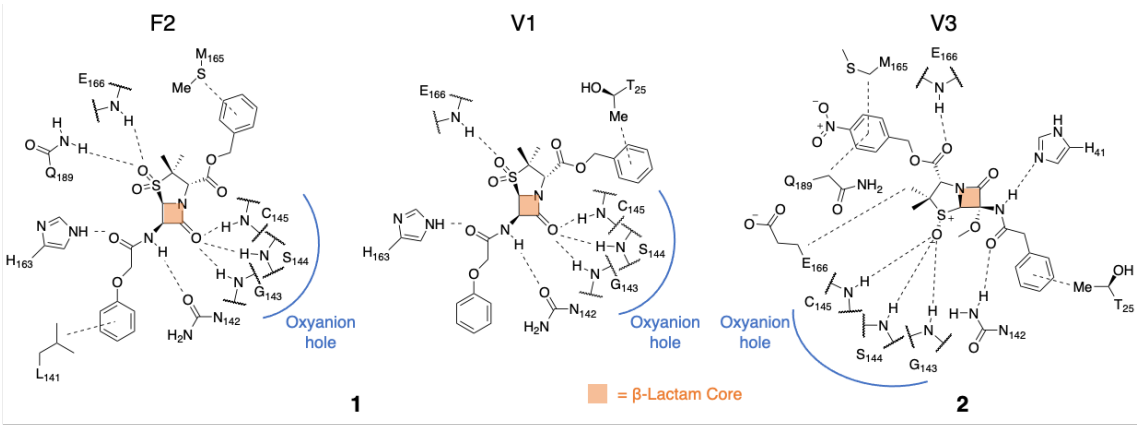


Figure 3.22: Interactions between M^{pro} and compounds (left) **1** and (right) **2**, in their respective most favourable bound conformations according to MD.

By contrast, the equivalent conformations for **2** (**Table 3.7.3**) were not universally stable during MD simulations, possibly due to the C6-methoxy group hindering interaction with His163. Instead, the most stable conformation (V3) spanned S2, the oxyanion hole, and S1′ (**Figure 3.22**). This conformation engaged in multiple HBs with residues in the oxyanion hole as well as His41, and dispersion interactions with Thr25 (via its C6 benzyl group) and Met165 (via its C3 *p*-nitrobenzyl group). While maintaining stability, this conformation precludes covalent adduct formation with Cys145. Overall, despite the differences between the bound conformations of **1**

and **2**, both exhibited binding motifs that remarkably resembled those seen in substrates and potent ligands (**Manuscript I**).[7]

While the two alternative bound conformations of **1** are similar in the positioning of the β-lactam core and the C6 amide sidechain, how the C3 benzyl ester group binds, *i.e.*, in either S1′ or S2, has implications for inhibitor optimisation. Considering the approximate nature of MM/PBSA energies, ΔG_{bind} values of **1** in both V1 and F2 conformations were evaluated using more robust FEP calculations (see **Appendix for Chapter 3: Section S3.8**), with $\Delta G_{lig\text{-}sol}$ in the ABFE thermodynamic cycle (**Figure 2.4**) being the same for F2 and V1. ΔG_{bind} of **1** in F2 and V1 was evaluated to be -5.1 ± 0.4 kcal mol^{-1} and -2.9 ± 0.3 kcal mol^{-1}, respectively (**Table 3.2**). Note that these values correspond to the reversible binding affinity prior to covalent reaction with Cys145,[8] and do not essentially directly correlate with experimental IC_{50} values, where irreversible, covalent irreversible binding of **1** likely occurs.[1, 9] Nevertheless, the values reveal that the binding of the hydrophobic benzyl moiety in the S2 pocket contributes significantly to binding, a pattern also observed in substrates and the designed peptide inhibitors (**Manuscript I**).[7]

Table 3.2: ABFE calculation of ΔG_{bind} of **1** in the F2 and V1 bound conformations.

Energy (kcal mol^{-1})	F2	V1
$-\Delta G_{lig\text{-}sol}$	+273.02 ± 0.09	
$+\Delta G_{rest(sol)}$	+7.03	+7.00
$-(\Delta G_{decoup} + \Delta G_{rest(comp)})$	-285.16 ± 0.39	-282.94 ± 0.32
$=\Delta G_{bind}$	-5.11 ± 0.40	-2.92 ± 0.33

During the course of this project, a crystal structure of M^{pro} covalently complexed with an analogue of the penicillin V sulfone **1**, but with a C3 2,4-difluorobenzyl ester instead of a benzyl ester, was reported (PDB 7Z59).[9] In this structure, Cys145 is acylated, with Sγ attached to the β-lactam C7 as a thioester group, and the thioester carbonyl coordinated to the Cys145 and Gly143 backbone. The β-lactam C7-N4 bond is cleaved, with additional fragmentation of the C5-S1 bond into an enamine and a sulfinic acid (SO_2H) group; the latter coordinates to the Glu166 sidechain and Ser1 of the other M^{pro} protomer at the dimer interface. Contrary to computational predictions,

the C6 phenoxy group binds in S2, while the C3 ester projects outward from the active site, leaving the S1 pocket vacant. Interestingly, the reported structure-activity relationships highlight the importance of the sulfone *pro-R* oxygen for potent inhibition. There is also high selectivity observed for the steric and electronic requirements in both the C3 ester (*e.g.*, *p*-nitrobenzyl, *p*-methoxybenzyl, and naphthylmethyl esters all lead to reduced or abolished inhibition) and the C6 amide sidechain (*e.g.*, both the presence and position of the ether oxygen are essential). These reported structure-activity relationships imply the importance of all these functional groups in interacting with M^{pro} prior to reaction with Cys145 and C5-S1 fragmentation. The necessity of all these interactions for potent inhibition is in accord with the binding patterns exhibited by substrates and XChem fragments, which were observed to form multiple interactions simultaneously in the M^{pro} S2 and S1 pockets and the oxyanion hole (**Manuscript I**).[7] The distinct hydrophobic or hydrophilic nature of each of these pockets (**Manuscript I**) likely governs the specific positioning and electronic requirements for each functional group on the β-lactam inhibitor. While the relevance of the crystal structure to inhibition in solution requires validation, future modelling can involve the acylated and fragmented complex, initiated from the predicted F2 conformation. Complementarily, the reverse reactions of the crystallographic observed fragmented β-lactam back to its initial non-covalently bound conformation can be modelled. These computational investigations will likely offer insights into the exact mechanisms of the non-covalent and covalent inhibition of M^{pro} by β-lactams.

3.5 References: β-lactam inhibitors

1. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
2. Lima, L. M.; Silva, B. N. M. d.; Barbosa, G.; Barreiro, E. J. β-lactam antibiotics: An overview from a medicinal chemistry perspective. *Eur. J. Med. Chem.* **2020,** *208*, 112829.
3. King, D. T.; Sobhanifar, S.; Strynadka, N. C. J. One ring to rule them all: Current trends in combating bacterial resistance to the β-lactams. *Protein Sci.* **2016,** *25*, 787.
4. Li, W.-J.; Li, D.-F.; Hu, Y.-L.; Zhang, X.-E.; Bi, L.-J.*, et al.* Crystal structure of L,D-transpeptidase LdtMt2 in complex with meropenem reveals the mechanism of carbapenem against Mycobacterium tuberculosis. *Cell Res.* **2013,** *23*, 728.
5. Riva, L.; Yuan, S.; Yin, X.; Martin-Sancho, L.; Matsunaga, N.*, et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **2020,** *586*, 113.
6. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.*, et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
7. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
8. Cer, R. Z.; Mudunuri, U.; Stephens, R.; Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **2009,** *37*, W441.
9. Malla, T. R.; Brewitz, L.; Muntean, D.-G.; Aslam, H.; Owen, C. D.*, et al.* Penicillin Derivatives Inhibit the SARS-CoV-2 Main Protease by Reaction with Its Nucleophilic Cysteine. *J. Med. Chem.* **2022,** *65*, 7682.

4. M^{pro} Substrate Binding Dynamics

4.1 Overview

Chapter 3 described the modelling of M^{pro}-substrate complexes that enabled the identification of key conserved interactions, which are also utilised for the binding of peptide and small molecule inhibitors, such as the penicillin-derived sulfone compound **1**. The modelling described thus far focused on the observation of M^{pro}-substrate or M^{pro}-inhibitor complexes under equilibrium conditions. However, the details of how M^{pro} structurally and dynamically adapts to substrate and inhibitor (un)binding have remained elusive from equilibrium MD simulations. In this Chapter, the D-NEMD approach, introduced in **Chapter 2**, was employed to investigate the dynamics of substrate binding of M^{pro}. This was achieved by applying a perturbation where the native peptide substrate s05 was instantaneously removed from the active site, with the subsequently induced M^{pro} response characterised by D-NEMD.

4.2 Manuscript II

The following manuscript (**Manuscript II**) has been published as the following under full open access (https://pubs.acs.org/doi/10.1021/jacsau.3c00185) and is adapted with permission. Copyright 2023 American Chemical Society.

II. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.*; Duarte, F.* Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3* (6), 1767-1774.

I carried out all the work described in this manuscript, with other co-authors contributing to discussions and review of the manuscript. A Statement of Authorship is appended following the manuscript. The Supplementary Information is in the **Appendix for Chapter 4**.

Dynamical nonequilibrium molecular dynamics simulations identify allosteric sites and positions associated with drug resistance in the SARS-CoV-2 main protease

Abstract

The SARS-CoV-2 main protease (M^{pro}) plays an essential role in the coronavirus lifecycle by catalysing hydrolysis of the viral polyproteins at specific sites. M^{pro} is the target of drugs, such as nirmatrelvir, though resistant mutants have emerged that threaten drug efficacy. Despite its importance, questions remain on the mechanism of how M^{pro} binds its substrates. Here, we apply dynamical nonequilibrium molecular dynamics (D-NEMD) simulations to evaluate structural and dynamical responses of M^{pro} to the presence and absence of a substrate. The results highlight communication between the M^{pro} dimer subunits and identify networks, including some far from the active site, that link the active site with a known allosteric inhibition site, or which are associated with nirmatrelvir resistance. They imply that some mutations enable resistance by altering the allosteric behaviour of M^{pro}. More generally, the results show the utility of the D-NEMD technique for identifying functionally relevant allosteric sites and networks including those relevant to resistance.

Keywords

SARS-CoV-2, main protease, coronavirus, molecular dynamics, nonequilibrium, protein dynamics, allosteric, drug resistance

Introduction

The ongoing COVID-19 pandemic has led to worldwide efforts to develop antiviral drugs effective against the SARS-CoV-2 coronavirus. The SARS-CoV-2 main protease (M^{pro}) cleaves viral polyproteins into functional units essential for viral replication and pathogenesis and is a validated drug target.[1] M^{pro} hydrolyses the amide bond between P1 and P1′ residues in the conserved sequence: [P4=Ala/Val/Pro/Thr]-[P3=X]-[P2=Leu/Phe/Val]-[P1=Gln]-[P1′=Ser/Ala/Asn], where X is any residue.[2-4] Due to the essential role of M^{pro}, the apparent absence of human proteases with a similar substrate selectivity, and the high sequence identity (96%) between SARS-CoV and SARS-CoV-2 M^{pro}s, inhibitors targeting M^{pro} enzymes are likely useful for treatment of many coronavirus diseases.[5-8] The covalently acting M^{pro} inhibitor PF-07321332 (nirmatrelvir), the active ingredient in Paxlovid, is the first oral COVID-19 drug that has been authorised for emergency use by the FDA.[9, 10] Another M^{pro} inhibitor in clinical-stage development is S-217622 (ensitrelvir), a non-covalently binding, nonpeptidic inhibitor that was discovered following virtual screening and medicinal chemistry optimisation.[11]

Crystal structures[4, 12-14] and molecular simulations[15-20] have provided insights into M^{pro} structure, dynamics and interactions with its substrates and other ligands. Molecular simulations of M^{pro} have identified cryptic pockets,[21] informed on mechanisms of inhibition,[22-24] and helped to suggest potential lead compounds for M^{pro} inhibitor design.[25, 26] We have identified conserved interactions between M^{pro} and its 11 native substrates, each modelled as an 11-mer P6-P5′ peptide.[20] However, how the binding of substrates or inhibitors induces changes in the M^{pro} structure and its dynamics has been unclear. Changes in dynamics and conformational behaviour have been shown to be the cause of mutation-induced drug resistance for some targets.[27-30] For instance, mutations distal to the active site of the HIV-1 protease have been shown to confer resistance to darunavir through alterations in conformational dynamics.[31] Given the relatively rapid rate at which SARS-CoV-2 mutates, M^{pro} targeting drug resistance is a major concern.[8, 32,

[33] A detailed understanding of how M^{pro} recognises its natural substrates thus may aid the development of next-generation inhibitors.[12, 34]

Here, we describe the application of dynamical-nonequilibrium molecular dynamics (D-NEMD)[35-38] simulations to investigate the response of M^{pro} to the instantaneous removal of a substrate in one of the active sites in the dimer. The D-NEMD approach aims to identify in the protein any structural changes, which refer to conformational rearrangements, as well as dynamic changes, which relate to changes in fluctuations, following a perturbation. D-NEMD simulations are emerging as a useful tool to probe signal propagation and allosteric effects,[38] *e.g.*, they have identified a general mechanism of signal propagation in nicotinic acetylcholine receptors;[39, 40] communication pathways between allosteric and active sites in clinically relevant β-lactamases;[30] and modulation of SARS-CoV-2 spike protein behaviour by pH[41] and ligand occupancy in the fatty acid binding site.[42-44] Notably in this context, the pathways identified in the β-lactamases contain positions that differ between enzymes with different spectrums of antibiotic breakdown activity.[30]

In the D-NEMD approach, the system is sampled first by equilibrium molecular dynamics (MD) simulations. Configurations from these trajectories provide starting points for multiple short nonequilibrium simulations, through which the effect of a perturbation can be determined using the Kubo-Onsager relation.[37, 38] Running a large number (usually tens to hundreds) of nonequilibrium simulations allows the statistical significance of the response of the protein to the perturbation to be tested.[37]

The results here identify structural pathways involved in M^{pro} communication, both within and between subunits of the dimer. We use a substrate sequence representing the nsp8/9 P6-P5′ substrate (Ser-Ala-Val-Lys-Leu-Gln↓Asn-Asn-Glu-Leu-Ser, where "↓" denotes the cleavage site), which we have previously characterised by experiment and simulation, and refer to as s05 according to previous nomenclature.[20] s05 was selected because it is a conserved substrate in

coronaviruses,[4] and its P4-P1 sequence is reported to be optimal for M^{pro} activity amongst natural amino acid sequences.[1]

Results and Discussion

1. Equilibrium simulations of the M^{pro}-substrate complex

Initially, five independent equilibrium MD simulations of 200 ns (an aggregate of 1 μs) were carried out for homodimeric M^{pro} non-covalently complexed with a single 11-mer peptide substrate s05 (**Figure 4.1a, b**), using established protocols.[20] s05 occupies the active site which is formed predominantly by residues of the M^{pro} chain denoted as chain A (ChA), while the other (chain B, ChB) active site was simulated without a substrate. In subsequent descriptions, all residues relate to M^{pro} chain A (Ser1-Gln306), except chain B residues which are indicated by prime (Ser1′-Gln306′).

Analysis of RMSD and RMSF values, protein secondary structure content, and M^{pro}-substrate interactions confirm the stability of the M^{pro} dimer-s05 complex, with the system considered to be well equilibrated from t = 50 ns onwards in all simulations (**Supplementary Information** in **Appendix for Chapter 4: Section S4.1**). The 12 M^{pro}-substrate hydrogen bonds (HBs) previously found to be conserved across 11 native substrates are also observed here (**Figures 4.1b, S4.1.9-11**).[20] These equilibrium trajectories were then used as the basis for D-NEMD simulations.

2. Immediate effect of substrate removal

To analyse the response of the M^{pro} dimer to substrate removal, snapshots were extracted every 5 ns from t = 50 to 195 ns from all five equilibrium simulations (a total of 150 conformations), and the s05 peptide instantaneously annihilated (see **Methods**). The time-dependent response of the

protein to the annihilation of s05 was monitored using the Kubo-Onsager relation,[37, 38] by subtracting the position of each Cα atom between the nonequilibrium and equilibrium trajectories at equivalent time points (**Figure S4.2.1)**. Cα movements were used, employing the same approach as in previous investigations of other systems, as they are less susceptible to thermal fluctuations than those of sidechain atoms, and they normally reflect the most substantial responses in the protein.[38] The subtraction approach enables a clear identification of the response via cancellation of most random fluctuations between the equilibrium and nonequilibrium trajectories.[30, 39] Note that the perturbation (in this case, the deletion of the s05 peptide) is not intended to simulate the physical process of substrate binding or dissociation, but rather to trigger a fast response from M^{pro} as it adapts to s05 removal. Such a perturbation forces the system out of equilibrium, thus creating the driving force for conformational changes to occur. Similarly, the simulation timescale does not reflect the real timescale of conformational changes.[39]

Soon after the perturbation, a clear response from M^{pro} is observed, notably in residues directly involved in substrate binding (**Figures 4.1cA, S4.2.2; Table S4.2.1**). The statistical significance of the response is confirmed by determination of standard errors of the mean (SEMs) (**Figure S4.2.3**), as performed in previous investigations of other systems.[30, 42] At t = 0.25 ps, the most perturbed regions include: Gly23-Thr26, which forms HBs with the substrate P2′ and P4′ residues (**Figure 4.1b**); Asn142-Cys145, which constitutes the oxyanion hole loop that stabilises the scissile amide carbonyl oxygen; Met165, which forms the S2 hydrophobic pocket; Glu166, whose backbone forms two strong HBs with the substrate P3 backbone; and Ala191, which is adjacent to Thr190 that binds the substrate P4 residue via a HB. Hence, as expected, the residues with the largest initial deviations are those originally in direct contact with, or very close to, the substrate, and these residues are likely to rearrange to accommodate the substrate.

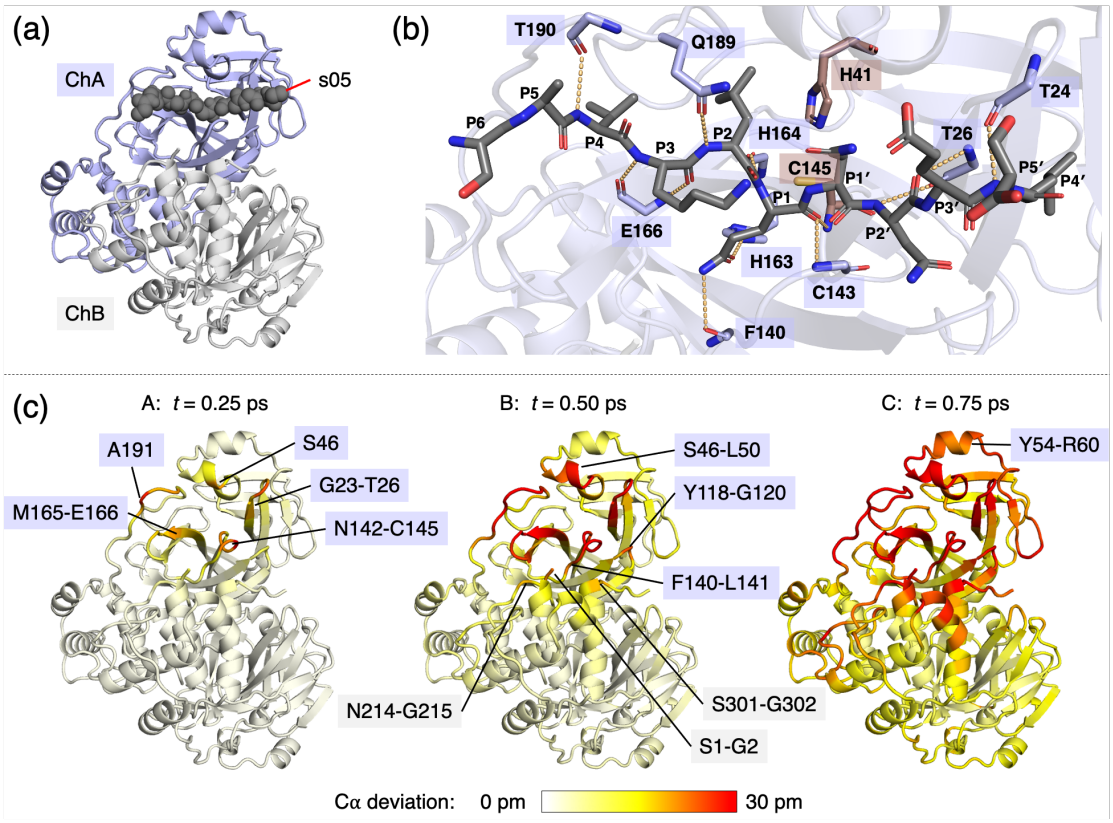


Figure 4.1: The immediate response of M^{pro} to substrate removal from the equilibrated M^{pro}-substrate complex. (a) View of the M^{pro} dimer (chains A and B shown as blue and grey cartoons, respectively) complexed with the s05 peptide substrate (backbone shown as black spheres) prior to MD simulations. (b) View of a MD-derived snapshot, with the 12 conserved HBs (**Figure S4.1.9**) shown as orange dotted lines. s05, the M^{pro} catalytic dyad, and other HB-forming M^{pro} residues are shown as black, brown, and blue sticks respectively (hydrogens omitted). (c) The initial sub-ps response of M^{pro} to s05 removal in terms of Cα deviation represented on a white-yellow-red scale. All figures were generated using PyMOL.[45]

Notable initial responses are also observed in residues belonging to chain B (**Figures 4.1cB, S4.2.2-3**). Chain B residues which show a deviation >15 pm at t = 0.50 ps include: Ser1′, Gly2′, Asn214′, Gly215′, Ser301′ and Gly302′. These residues are in the vicinity of chain A regions that also show significant responses, such as Phe140-Leu141 and Tyr118-Gly120 (**Figure S4.2.4**). These responses highlight the importance of inter-subunit dynamics, where structural and dynamic changes occurring in one are swiftly propagated to the other. For example, the structural importance of the N-terminus (Ser1′-Gly2′) of chain B is consistent with its role of forming part of the S1 pocket that accommodates the conserved P1 Gln substrate residue.[5, 6] Asn214′ and

Gly215′ are further away from the substrate binding site; notably, the N214A variant of SARS-CoV M^{pro} is much less active than the wild type enzyme against a fluorogenic substrate peptide, despite being similar in dimerization affinity and structure, with the inactivation proposed to be driven by changes in protein dynamics.[15, 46] Finally, the observed perturbations in Ser301′ and Gly302′ of the C-terminal tail of chain B are consistent with crystallographic observations that this region adopts distinct conformations upon the binding of substrate or the peptidomimetic inhibitor N3.[13, 47] Overall, the D-NEMD analysis highlights the communication between chains A and B and, in agreement with reported experimental and crystallographic data, identifies three structurally and dynamically important regions of chain B.

At 0.75 ps, the structural responses intensify and propagate further from the substrate binding site into both chains A and B (**Figures 4.1cC, S4.2.3**). These responses include signal transmission in chain A from Ser46 to the whole Ser46-Leu50 helix, and then to the Tyr54-Arg60 helix, above the active site. There is also transmission through the helical regions in chain B away from the dimer interface. These results show communication between the subunits of the dimer.

3. Propagation of signal to allosteric sites

The conformational changes at the dimer interface are of particular interest. While the initial response (**Figure 4.1c**) in chain B is concentrated at dimer interface regions with little secondary structure, such as Asn214′-Gly215′ and Ser301′-Gly302′, notable response in the Gln244′-Thr257′ kinked helix is observed across the ps timescale (**Figure 4.2**). The helix residues Pro252′, Leu253′, and Gln256′; the C-terminal tail residues Cys300′, Ser301′, and Gly302′; Ile213′, which is adjacent to the perturbed Asn214′-Gly215′ region; and the earlier highlighted chain A residues Tyr118, Leu141, and Asn142 all form an allosteric site (coloured orange in **Figure 4.2a**), which binds small molecules as demonstrated by crystallography.[7, 48, 49] One such compound is pelitinib, which has high antiviral activity and which does not bind at the active site.[48] It is striking that the

D-NEMD simulations, which were initiated by substrate removal, identify this allosteric site. Our results show a direct connection between the allosteric site and the active site via conformational motions. Based on these D-NEMD results, it may be proposed that the presence of a ligand at this allosteric site will modulate enzyme dynamics, including those affecting the rate and/or affinity of binding of substrates. Such a ligand may also alter how dynamic and structural changes are communicated within the protein, a likely factor in allosteric inhibition.

As the time following substrate removal increases across the ps and ns timescales, the structural response in residues in terms of Cα deviations increases and spreads across the M^{pro} dimer (**Section S4.3**). Perturbed allosteric regions are found in both M^{pro} chains, and except for the region around Gln189-Ala191, which has higher deviation in chain A than B, the deviations in the two chains are similar (**Figures S4.3.2-4**). Some of the most perturbed residues (*e.g.*, Asn277 and Gly278 in both chains at t = 1 ns) also display high mobility in the equilibrium simulations (**Figure S4.1.3**). This is unsurprising, because more flexible regions require less energy to change their positions.[36, 50]

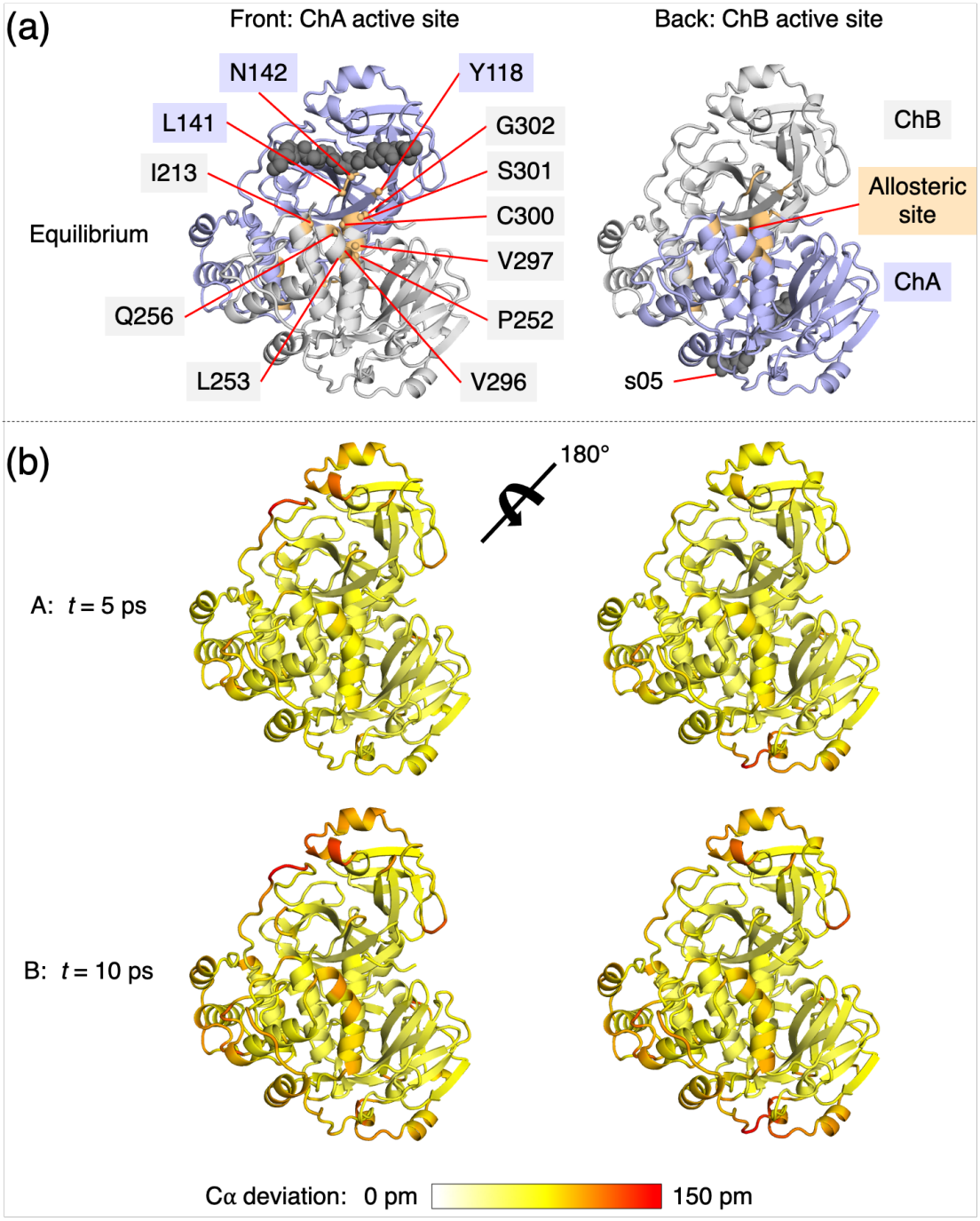


Figure 4.2: D-NEMD identifies connection between the active site and a known allosteric inhibition site. (a) Alternative (180° rotated) views of the M^{pro}-s05 complex prior to MD simulations. Residues constituting the pelitinib binding allosteric site (within 4 Å of pelitinib according to PDB 7AXM: Tyr118, Leu141, Asn142, Ile213′, Pro252′, Leu253′, Gln256′, Val296′, Val297′, Cys300′, Ser301′, Gly302′; see **Figure S4.3.6**)[48] are labelled and shown in orange. (b) D-NEMD Cα responses to s05 peptide removal shown on a white-yellow-red scale, with red indicating large deviation. The figures were generated using PyMOL.[45]

4. Coherent displacements in residues upon substrate removal

Complementarily to the above approach, which averages scalar deviations across replicas, the displacement vector of Cα atoms between the equilibrium and nonequilibrium trajectories at equivalent time points can be averaged. The resulting signals indicate the average direction of the response of the residues upon the perturbation; however, it is important to note that the responses of residues that are dynamically affected in multiple directions or in a disordered manner are not apparent in this analysis.

The response of M^{pro} to substrate removal calculated from Cα displacement vectors is shown in **Figure S4.4.1**, with the statistical significance of the responses assessed by SEM (**Figures S4.4.2-5**). Due to the cancellation of random movements in mobile atoms, the observed responses are more localised than those calculated by averaging scalar deviations (**Figures S4.3.1, S4.4.1**). Upon removal of the substrate peptide, M^{pro} regions that form the substrate binding site show responses with well-defined directions (**Figure 4.3aA; Table 4.1**). These regions include: Thr26-Leu27 in the S1′-S2′ pockets; Asn142-Gly146 in the oxyanion hole loop, with the largest displacement observed in Gly143; His163 and Met165 whose sidechains form parts of the S1 and S2 pockets respectively; and Ala191-Gln192 which lie behind the substrate P5-P4 residues in the substrate complex.

At t = 10 ps (**Figure 4.3aB; Table 4.1**), additional displacements are observed in the Ser46-Leu50 and Tyr54-Arg60 helices and their connecting loop above the active site in chain A. Displacements in residues surrounding Ala191 and Gly143 increase in magnitude, with the Arg188-Gln192 loop moving outwards into the substrate S5 and S4 pockets, and the Asn142-Gly143 loop closing inwards into S1. These motions collectively reflect a closure of the M^{pro} substrate binding site upon substrate removal.

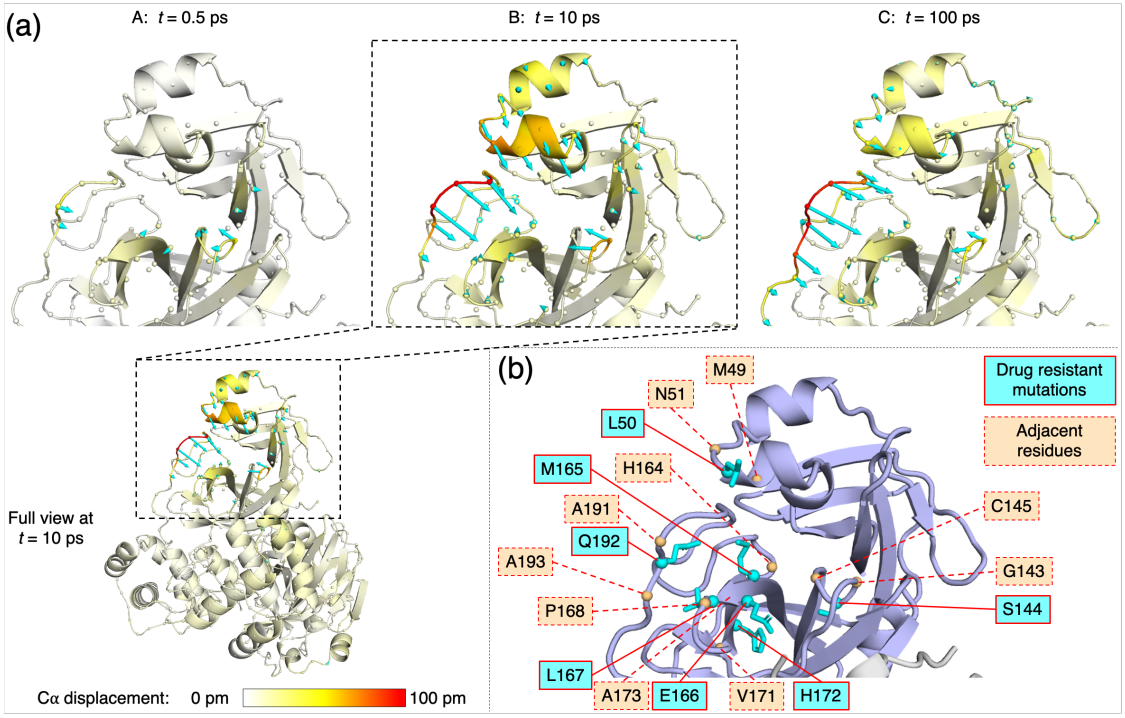


Figure 4.3: D-NEMD analysis of displacement vectors identifies sites associated with nirmatrelvir resistance. (a) Views of the M^{pro}-s05 complex prior to MD simulations, with a focus on the substrate binding site, where significant D-NEMD responses, measured by averaging Cα displacement vectors between the equilibrium and nonequilibrium trajectories, are observed. Displacement magnitudes are represented on a white-yellow-red scale (figures created using PyMOL).[45] Vectors with length ≥15 pm are displayed as cyan arrows with a scale-up factor of 5.[51] (b) M^{pro} residues for which mutations associated with nirmatrelvir resistance have been reported,[52-54] and adjacent residues in protein sequence.

Table 4.1: The most perturbed M^{pro} residues in terms of Cα displacements following substrate removal. Residues corresponding to and adjacent to mutation sites are coloured cyan and orange, respectively (see **Figure 4.3b**).

Rank	M^{pro} residues with the largest Cα displacements (Å)							
	t = 0.5 ps		*t* = 10 ps		*t* = 100 ps		*t* = 1 ns	
1	G143	0.519	A191	1.341	A191	1.433	A193	0.977
2	A191	0.382	T190	1.048	Q192	1.013	A191	0.943
3	C145	0.247	Q189	1.014	T190	0.914	Q192	0.839
4	N142	0.239	L50	0.705	A193	0.867	A194	0.666
5	Q192	0.235	Q192	0.701	Q189	0.742	G143	0.66
6	T26	0.195	G143	0.681	G143	0.524	P168	0.525
7	H163	0.194	N51	0.636	R188	0.502	T190	0.523
8	M165	0.185	E47	0.623	A194	0.461	T169	0.497
9	S144	0.184	N142	0.615	N142	0.436	N142	0.496
10	G146	0.181	R188	0.596	G195	0.392	G195	0.471

From *t* = 10 ps to 100 ps (**Figure 4.3aC**), a reduction in the vector response is observed in certain regions, such as the Ser46-Leu50 helix, suggesting that there is no longer a consistent directional difference in the motions between the equilibrium and nonequilibrium trajectories. By contrast, the displacements in the Gly143 oxyanion hole loop and in the Val186-Gly195 loop persist. In the latter, the response is initially centred at Ala191, but gradually propagates downstream to residues as far as Gly195 (**Figures 4.3aC, S4.4.6; Table 4.1**). The loop in chain A remains the most perturbed region in the dimer at *t* = 5 ns (**Figures S4.4.5-6**), with Ala193 showing the largest displacement of 0.91 Å. These observations likely reflect long-lived structural alterations to accommodate the substrate. The conformational adaptation of the Val186-Gly195 loop observed here echoes structural studies of M^{pro} complexed with substrate peptides: the loop has crystallographically been found to be conformationally variable, which enables the binding of diverse substrate sequences.[12] Overall, the D-NEMD approach provides a statistically meaningful prediction of how each M^{pro} region dynamically adapts to bind its substrate. The collective and coherent motion is not limited to the active site but extends well beyond.

Conclusions

Selection pressure caused by treatments that target M^{pro}, such as nirmatrelvir, drives the emergences of drug resistant mutations in SARS-CoV-2 M^{pro}.[24, 34, 54, 55] Such mutations would be expected to weaken the binding affinity of nirmatrelvir, but still allow M^{pro} to perform its biological function by binding, recognising, and hydrolysing its substrate sequences in the viral polyproteins. The effect of such mutations on dynamics is of particular interest. Indeed, M^{pro} mutations that confer high levels of resistance to nirmatrelvir have been reported, including the combined L50F+E166A+L167F substitutions,[52] L50F+E166V substitutions,[53] and naturally observed variations including S144M/F/A/G/Y, M165T, E166V/G/A, H172Q/F, and Q192T/S/L/A/I/P/H/V/W/C/F (**Figure 4.3b**).[54] Some of these mutated residues, such as Glu166

and Met165, are in direct contact with the substrate P3/P1 and P2 residues respectively, and these interactions are exploited by nirmatrelvir.[9] The effect of mutations that are more distant from the active site, such as those involving Leu50 and Gln192, on substrate and inhibitor binding, however, is less clear. The D-NEMD results here indicate that these residues are involved in allosteric response to active site binding.

Our D-NEMD results (**Figure 4.3a; Table 4.1**) highlight the dynamic behaviour of the oxyanion hole loop around Gly143. The dynamics of this loop will probably be affected by the observed mutation of Ser144, due to alterations in size, polarity, and HB interactions. Further from the catalytic site, the simulations identify Leu50 and its associated helix (Ser46-Leu50), as well as Gln192 and the loop of which it forms part, as being affected by the loss of substrate. While the exact influence of these mutations on binding affinities and reactivities of substrates or inhibitors remains to be investigated, this study has showcased the ability of the D-NEMD approach to provide valuable insights into time-ordered protein structural changes and pinpoint positions of drug-resistant mutations. The evaluation of the perturbation-induced responses also enables the identification, and potentially prediction, of allosteric sites that play major roles in regulating enzyme activity.

The D-NEMD approach is emerging as a useful tool to investigate signal propagation, identify communication networks, and characterise allosteric effects.[38] As demonstrated by this study, D-NEMD simulations will be useful in predicting allosteric sites and mutation sites relevant to drug resistance.

Methods

Equilibrium MD: MD simulations were performed with GROMACS (v 2019.2)[56] using the AMBER99SB-ILDN forcefield.[57] Dimeric M^{pro} (PDB 6YB7;[58] 1.25 Å resolution; chains A and B) in non-covalent complex with the s05 peptide was constructed and prepared as described previously, including the setup of histidine residues (**Table S4.5.1**) and the neutral catalytic dyad.[20] The complex was placed in a rhombic dodecahedral box with at least 1.0 nm separation from box edges, solvated with TIP3P water,[59] neutralised with sodium ions (86,304 atoms in solvated system), and minimised until the maximum force was below 1000 kJ mol^{-1} nm^{-1}. From the minimised system, five replicas were initiated by assigning random velocities at 298.15 K. Each replica was subjected to 200 ps (1 fs step) NVT and 200 ps (1 fs step) NPT equilibration at 298.15 K and 1.0 bar, before a production run of 200 ns (2 fs step), during which coordinates were saved every 100 ps and velocities saved every 1 ns. Temperature coupling to the protein and non-protein at 298.15 K was achieved using a velocity-rescaling thermostat with a stochastic term, with a time constant of 0.1 ps.[60] Pressure was maintained at 1.0 bar with a Parrinello-Rahman barostat with a time constant of 2 ps.[61, 62] Long-range electrostatics were treated using smooth Particle-mesh Ewald with a 1 nm cut-off.[63, 64] For van der Waals interactions, a 1 nm cut-off was used. For these 5 × 200 ns equilibrium simulations, analysis was performed using GROMACS tools (v 2019.2)[56] and DSSP (v 2.0.4).[65, 66] Hydrogen bonds were defined based on a combined distance ($d_{D\text{-}A} \leq 3.5$ Å) and angle ($\angle$(H-D-A) ≤ 30°) criteria.

D-NEMD: From each of the five equilibrium trajectories, coordinates and velocities were extracted every 5 ns from t = 50 ns up to 195 ns, providing a total of 5 × 30 = 150 configurations for D-NEMD initiation. For the nonequilibrium simulations, the s05 peptide was instantaneously removed. No new molecules were added in its place. The nonequilibrium simulations were then started from these conformations. To capture events happening over different timescales, four length scales were chosen for the restarted MD simulations (2 fs step) with different saving frequencies of the protein coordinates: (i) over 2 ps saving every 0.05 ps; (ii) over 200 ps saving

every 0.5 ps; (iii) over 1 ns saving every 10 ps; (iv) over 5 ns saving every 100 ps (only conducted for the nonequilibrium propagation, as the resulting time points coincided with the equilibrium trajectories).

D-NEMD analysis: At equivalent delays after perturbation, the M^{pro} structure from the nonequilibrium propagation was fitted onto the corresponding equilibrium structure based on the 612 M^{pro} Cα atoms, before calculation of the scalar deviation and equilibrium-to-nonequilibrium displacement vector of every Cα atom. The response of each Cα atom was determined using the Kubo-Onsager approach, comparing the equilibrium and nonequilibrium simulations at equivalent points in time.[37,38] For each time, the Cα atom scalar deviation and displacement vector were averaged over the 150 replicas. According to the Kubo-Onsager relation, the averages yield estimates of the macroscopic time-dependent response of the system. The statistical significance was assessed by the standard error of the mean (SEM; N = 150). Average deviations and displacements were visualised using the M^{pro}-s05 complex structure following energy minimisation and prior to any MD simulations, with the aid of PyMOL (v 2.3.0)[45] and the modevectors.py script.[51]

Data Availability

System setup, and detailed analysis of equilibrium and nonequilibrium MD simulations are available in **Appendix for Chapter 4: Supplementary Information to Manuscript II**. All simulation data (including input, structure, and trajectory files) are openly available on GitHub (https://github.com/duartegroup/Mpro-Substrate_D-NEMD).

References

1. Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M., *et al.* SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2021,** *17*, 222.
2. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W., *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265.
3. Koudelka, T.; Boger, J.; Henkel, A.; Schönherr, R.; Krantz, S., *et al.* N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *Proteomics* **2021,** *21*, 2000246.
4. MacDonald, E. A.; Frey, G.; Namchuk, M. N.; Harrison, S. C.; Hinshaw, S. M., *et al.* Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **2021,** *7*, 2591.
5. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M., *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
6. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C., *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368*, 409.
7. Cho, E.; Rosa, M.; Anjum, R.; Mehmood, S.; Soban, M., *et al.* Dynamic Profiling of β-Coronavirus 3CL Mpro Protease Ligand-Binding Sites. *J. Chem. Inf. Model.* **2021,** *61*, 3058.
8. Lv, Z.; Cano, K. E.; Jia, L.; Drag, M.; Huang, T. T., *et al.* Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development. *Front. Chem.* **2022,** *9*, 819165.
9. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M., *et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374*, 1586.
10. US Food & Drug Administration Coronavirus (COVID-19) Update: FDA Authorizes First Oral Antiviral for Treatment of COVID-19. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19 (accessed 2021-12-22).
11. Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y., *et al.* Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022,** *65*, 6499.
12. Shaqra, A. M.; Zvornicanin, S. N.; Huang, Q. Y. J.; Lockbaum, G. J.; Knapp, M., *et al.* Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nat. Commun.* **2022,** *13*, 3556.
13. Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F., *et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020,** *11*, 5877.
14. Kneller, D. W.; Zhang, Q.; Coates, L.; Louis, J. M.; Kovalevsky, A. Michaelis-like complex of SARS-CoV-2 main protease visualized by room-temperature X-ray crystallography. *IUCrJ* **2021,** *8*, 973.
15. Suárez, D.; Díaz, N. SARS-CoV-2 Main Protease: A Molecular Dynamics Study. *J. Chem. Inf. Model.* **2020,** *60*, 5815.
16. Díaz, N.; Suárez, D. Influence of charge configuration on substrate binding to SARS-CoV-2 main protease. *Chem. Commun.* **2021,** *57*, 5314.
17. Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem. Sci.* **2020,** *11*, 10626.
18. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020,** *10*, 12544.
19. Khrenova, M. G.; Tsirelson, V. G.; Nemukhin, A. V. Dynamical properties of enzyme–substrate complexes disclose substrate specificity of the SARS-CoV-2 main protease as characterized by the electron density descriptors. *Phys. Chem. Chem. Phys.* **2020,** *22*, 19069.
20. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M., *et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
21. Sztain, T.; Amaro, R.; McCammon, J. A. Elucidation of Cryptic and Allosteric Pockets within the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2021,** *61*, 3495.
22. Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V., *et al.* Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021,** *12*, 1433.

23. Martí, S.; Arafet, K.; Lodola, A.; Mulholland, A. J.; Świderek, K.*, et al.* Impact of Warhead Modulations on the Covalent Inhibition of SARS-CoV-2 Mpro Explored by QM/MM Simulations. *ACS Catal.* **2022,** *12*, 698.
24. Ramos-Guzmán, C. A.; Andjelkovic, M.; Zinovjev, K.; Ruiz-Pernía, J. J.; Tuñón, I. The impact of SARS-CoV-2 3CL protease mutations on nirmatrelvir inhibitory efficiency. Computational insights into potential resistance mechanisms. *Chem. Sci.* **2023,** *14*, 2686.
25. El Khoury, L.; Jing, Z.; Cuzzolin, A.; Deplano, A.; Loco, D.*, et al.* Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation. *Chem. Sci.* **2022,** *13*, 3674.
26. Thakur, A.; Sharma, G.; Badavath, V. N.; Jayaprakash, V.; Merz, K. M., Jr.*, et al.* Primer for Designing Main Protease (Mpro) Inhibitors of SARS-CoV-2. *J. Phys. Chem. Lett.* **2022,** *13*, 5776.
27. Matthew, A. N.; Leidner, F.; Lockbaum, G. J.; Henes, M.; Zephyr, J.*, et al.* Drug Design Strategies to Avoid Resistance in Direct-Acting Antivirals and Beyond. *Chem. Rev.* **2021,** *121*, 3238.
28. King, N. M.; Prabu-Jeyabalan, M.; Bandaranayake, R. M.; Nalam, M. N.; Nalivaika, E. A.*, et al.* Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease. *ACS Chem. Biol.* **2012,** *7*, 1536.
29. Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.*, et al.* L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chem. Sci.* **2018,** *9*, 2740.
30. Galdadas, I.; Qu, S.; Oliveira, A. S. F.; Olehnovics, E.; Mack, A. R.*, et al.* Allosteric communication in class A β-lactamases occurs via cooperative coupling of loop dynamics. *eLife* **2021,** *10*, e66567.
31. Henes, M.; Lockbaum, G. J.; Kosovrasti, K.; Leidner, F.; Nachum, G. S.*, et al.* Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance. *ACS Chem. Biol.* **2019,** *14*, 2441.
32. Moghadasi, S. A.; Esler, M. A.; Otsuka, Y.; Becker, J. T.; Moraes, S. N.*, et al.* Gain-of-Signal Assays for Probing Inhibition of SARS-CoV-2 Mpro/3CLpro in Living Cells. *mBio* **2022,** *13*, e00784.
33. Telenti, A.; Hodcroft, E. B.; Robertson, D. L. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harb. Perspect. Med.* **2022,** *12*, a041390.
34. Flynn, J. M.; Samant, N.; Schneider-Nachum, G.; Barkan, D. T.; Yilmaz, N. K.*, et al.* Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms. *eLife* **2022,** *11*, e77433.
35. Ciccotti, G.; Jacucci, G. Direct Computation of Dynamical Response by Molecular Dynamics: The Mobility of a Charged Lennard-Jones Particle. *Phys. Rev. Lett.* **1975,** *35*, 789.
36. Ciccotti, G.; Jacucci, G.; McDonald, I. R. "Thought-experiments" by molecular dynamics. *J. Stat. Phys.* **1979,** *21*, 1.
37. Ciccotti, G.; Ferrario, M. Non-equilibrium by molecular dynamics: a dynamical approach. *Mol. Simul.* **2016,** *42*, 1385.
38. Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur. Phys. J. B* **2021,** *94*, 144.
39. Oliveira, A. S. F.; Shoemark, D. K.; Campello, H. R.; Wonnacott, S.; Gallagher, T.*, et al.* Identification of the Initial Steps in Signal Transduction in the α4β2 Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. *Structure* **2019,** *27*, 1171.
40. Oliveira, A. S. F.; Edsall, C. J.; Woods, C. J.; Bates, P.; Nunez, G. V.*, et al.* A General Mechanism for Signal Propagation in the Nicotinic Acetylcholine Receptor Family. *J. Am. Chem. Soc.* **2019,** *141*, 19953.
41. Dommer, A.; Casalino, L.; Kearns, F.; Rosenfeld, M.; Wauer, N.*, et al.* #COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol. *Int. J. High Perform. Comput. Appl.* **2022,** *37*, 28.
42. Oliveira, A. S. F.; Shoemark, D. K.; Avila Ibarra, A.; Davidson, A. D.; Berger, I.*, et al.* The fatty acid site is coupled to functional motifs in the SARS-CoV-2 spike protein and modulates spike allosteric behaviour. *Comput. Struct. Biotechnol. J.* **2022,** *20*, 139.
43. Oliveira, A. S. F.; Shoemark, D. K.; Davidson, A. D.; Berger, I.; Schaffitzel, C.*, et al.* SARS-CoV-2 spike variants differ in their allosteric response to linoleic acid. *J. Mol. Cell. Biol.* **2023**, mjad021.
44. Gupta, K.; Toelzer, C.; Williamson, M. K.; Shoemark, D. K.; Oliveira, A. S. F.*, et al.* Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2. *Nat. Commun.* **2022,** *13*, 222.

45. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
46. Shi, J.; Han, N.; Lim, L.; Lua, S.; Sivaraman, J.*, et al.* Dynamically-Driven Inactivation of the Catalytic Machinery of the SARS 3C-Like Protease by the N214A Mutation on the Extra Domain. *PLOS Comput. Biol.* **2011,** *7*, e1001084.
47. Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.*, et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020,** *11*, 3202.
48. Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.*, et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021,** *372*, 642.
49. Alzyoud, L.; Ghattas, M. A.; Atatreh, N. Allosteric Binding Sites of the SARS-CoV-2 Main Protease: Potential Targets for Broad-Spectrum Anti-Coronavirus Agents. *Drug. Des. Devel. Ther.* **2022,** *16*, 2463.
50. Abreu, B.; Lopes, E. F.; Oliveira, A. S. F.; Soares, C. M. F508del disturbs the dynamics of the nucleotide binding domains of CFTR before and after ATP hydrolysis. *Proteins* **2020,** *88*, 113.
51. Law, S. M. modevectors.py. https://pymolwiki.org/index.php/Modevectors (accessed 2022-06-30).
52. Jochmans, D.; Liu, C.; Donckers, K.; Stoycheva, A.; Boland, S.*, et al.* The Substitutions L50F, E166A, and L167F in SARS-CoV-2 3CLpro Are Selected by a Protease Inhibitor In Vitro and Confer Resistance To Nirmatrelvir. *mBio* **2023,** *14*, e02815.
53. Zhou, Y.; Gammeltoft, K. A.; Ryberg, L. A.; Pham, L. V.; Tjørnelund, H. D.*, et al.* Nirmatrelvir-resistant SARS-CoV-2 variants with high fitness in an infectious cell culture system. *Sci. Adv.* **2022,** *8*, eadd7197.
54. Hu, Y.; Lewandowski, E. M.; Tan, H.; Zhang, X.; Morgan, R. T.*, et al.* Naturally Occurring Mutations of SARS-CoV-2 Main Protease Confer Drug Resistance to Nirmatrelvir. *ACS Cent. Sci.* **2023**, DOI: 10.1021/acscentsci.3c00538.
55. Moghadasi, S. A.; Heilmann, E.; Khalil, A. M.; Nnabuife, C.; Kearns, F. L.*, et al.* Transmissible SARS-CoV-2 variants with resistance to clinical protease inhibitors. *Sci. Adv.* **2023,** *9*, eade8778.
56. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
57. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
58. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
59. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
60. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126*, 014101.
61. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52*, 7182.
62. Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50*, 1055.
63. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089.
64. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.*, et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.
65. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.*, et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015,** *43*, D364.
66. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577.

4.3 Statement of authorship

Statement of Authorship for joint/multi-authored papers for PGR thesis
To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3* (6), 1767-1774.

Student Confirmation

Student Name:	Ho Ting Henry CHAN		
Contribution to the Paper	Entire work.		
Signature		Date	1 August 2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Fernanda Duarte			
Supervisor comments I confirm the description above regarding authorship and contributions is accurate.			
Signature		Date	2 August 2023

This completed form should be included in the thesis, at the end of the relevant chapter.

5. M^{pro} Substrate Recognition in each Subsite

5.1 Overview

Chapter 4 described D-NEMD analysis of equilibrium and nonequilibrium MD simulations that enabled characterisation of the response of M^{pro} to the presence or absence of substrate, in particular the observation of the strikingly dynamic behaviour of the oxyanion hole loop, the Ser46-Leu50 helix, and the Gln192-containing loop. Further dissection of the origins of these responses, that is, whether they arise from the binding of the sidechain of any particular substrate residue, was of interest. This is because M^{pro} is required, by nature, to exhibit high substrate selectivity, such as at the specified positions of P1 (Gln) and P2 (Leu/Phe/Val) (**Chapter 1**), to cleave at the correct positions in the long polyproteins and release nsps of correct lengths. As described in the following, D-NEMD was employed to investigate how each substrate residue is recognised in the corresponding M^{pro} subsite, by introducing alanine substitutions. Furthermore, D-NEMD was utilised to identify any differences between the recognition of the P2 residue of native substrates and of inhibitory peptides such as p12, where the P2-Trp was proposed to hinder efficient hydrolysis by M^{pro} (**Chapter 3**).

5.2 Manuscript III

At the time of writing this thesis, the following manuscript (**Manuscript III**) is in preparation and has yet to be submitted to a journal for consideration.

III. Chan, H. T. H.; Oliveira, A. S. F.; Mulholland, A. J.; Schofield, C. J.; Duarte, F. Substrate specificity and recognition in each SARS-CoV-2 main protease subsite informed by dynamical nonequilibrium molecular dynamics simulations. *In preparation* **2023**.

I carried out all the work described in this manuscript, with other co-authors contributing to discussions and review of the manuscript. A Statement of Authorship is appended following the manuscript. The Supplementary Information is in the **Appendix for Chapter 5**.

Substrate specificity and recognition in each SARS-CoV-2 main protease subsite informed by dynamical nonequilibrium molecular dynamics simulations

Abstract

The main protease (M^{pro}) of the SARS-CoV-2 coronavirus employs a cysteine-histidine dyad in its active site to hydrolyse the viral polyproteins at specific sites, specifically recognising a glutamine residue at the P1 position prior to the cleavage site. While the favourable binding of the P1-Gln in the M^{pro} active site has been utilised for substrate and inhibitor design, how M^{pro} dynamically recognises and responds to its binding remains difficult to probe by experimental methods. In this study, we employed the dynamical nonequilibrium molecular dynamics (D-NEMD) approach to probe the response of M^{pro} to a mutation of the substrate P1-Gln into alanine. This mutation perturbs the reactive conformations of the active site residues Cys145 and His41. We observed different responses in M^{pro} triggered by mutations of other residues along the substrate sequences, which differ in location and intensity. Our results provide insights into the dynamic and coordinated response of M^{pro} to the binding of substrate residues in each of its subsites, highlighting the importance of recognising the P1-Gln residue for initiating catalysis, and the value of the D-NEMD approach in investigating the dynamics of substrate recognition and specificity.

Keywords

SARS-CoV-2, main protease, coronavirus, nonequilibrium, molecular dynamics, protein dynamics, substrate recognition, specificity

Introduction

The SARS-CoV-2 coronavirus main protease (M^{pro}) has attracted considerable attention as a potential drug target for COVID-19 since the onset of the global pandemic. Encoded as the non-structural protein 5 (nsp5) in the viral polyproteins, M^{pro}, along with the papain-like protease (PL^{pro}) encoded within nsp3, are the two viral nucleophilic cysteine proteases responsible for cleaving the polyproteins at designated sites to release the individual nsps.[1-4] As these nsps are essential to viral maturation and replication, inhibition of M^{pro} and/or PL^{pro} is an effective antiviral strategy.[2-4] M^{pro} is particularly interesting due to its conserved nature among coronaviruses, with 96% sequence identity compared to SARS-CoV M^{pro}, and minimal similarity to human proteases, thus likely reducing the risk of side effects caused by inhibitors.[5, 6] Indeed, current COVID-19 treatments such as nirmatrelvir (PF-07321332 in Paxlovid)[7] and ensitrelvir (S-217622 in Xocova)[8] rely on M^{pro} inhibition.

Mechanistically, the predominantly homodimeric M^{pro} enzyme employs its active site catalytic dyad, composed of the nucleophilic Cys145 and general base His41,[9-13] to hydrolyse the amide bond at 11 sites in the viral polyproteins. The cleavage sites are between the P1 and P1′ residues in the consensus sequence: [P4=Ala/Val/Pro/Thr]-[P3=X]-[P2=Leu/Phe/Val]-[P1=Gln]↓[P1′=Ser/Ala/Asn], where "↓" denotes the cleavage site and X apparently any proteinogenic amino acid.[12, 14-19] The P4-P1′ residues bind in the corresponding S4-S1′ subsites on M^{pro}. Notably, the recognition of Gln in S1 is a common feature among M^{pro} substrates of SARS-CoV-2 and other coronaviruses,[18, 20] although a His residue is also tolerated.[14] The S1 pocket is a binding hotspot for potent M^{pro} inhibitors,[21-23] and has been targeted using various compounds, including those containing an unmodified Gln,[17, 21, 24] Gln-mimicking cycloglutamine (γ-lactam)[5-7, 25, 26] and δ-lactam,[27] His,[28] and heterocycles like 1,2,4-triazole.[8] Some inhibitors also have hydrophobic sidechains in the P1 position, such as Met,[29] cyclobutylmethyl in boceprevir, propyl in telaprevir, and butyl in narlaprevir,[30-33] although in these cases the higher reactivity of the aldehyde or α-ketoamide warheads likely compensates the less favourable interactions in S1.[29]

Despite this apparent promiscuity of S1 binders, M^{pro} selectively hydrolyses C-terminal to a Gln during polyprotein processing. How M^{pro} dynamically recognises P1-Gln, and how it would respond in the absence of the Gln sidechain, is crucial for understanding substrate selectivity and activity mechanism in M^{pro}.[34] However, exploring these aspects using traditional experimental methods remains challenging.

In this study, we employ dynamical-nonequilibrium molecular dynamics (D-NEMD)[35-38] to investigate the dynamic responses of M^{pro} to an instantaneous mutation of substrate residues of interest to alanine. D-NEMD has been successfully employed to investigate various biological systems, including signal propagation and allosteric effects in nicotinic acetylcholine receptors,[39, 40] β-lactamases,[41] and SARS-CoV-2 spike protein.[42-45] Recently, we reported the use of D-NEMD to identify the structural and dynamic changes in SARS-CoV-2 M^{pro} following substrate removal.[46] In D-NEMD, the system of interest is firstly simulated using equilibrium molecular dynamics (MD), from where configurations are used to initiate multiple short nonequilibrium MD simulations where a perturbation is applied. The response to the perturbation is quantified by comparing the equilibrium and nonequilibrium propagations. Using the Kubo-Onsager relation,[37, 38] the response to the perturbation is evaluated by averaging over the responses obtained from multiple (tens to hundreds) pairs of equilibrium and nonequilibrium simulations. Such averaging enables one to assess the significance of the response by statistical measures,[37] such as the standard error of the mean (SEM).[46]

Our combined D-NEMD/mutation study, akin to traditional alanine scanning in experimental[47-49] and computational studies,[50-53] provides insights into the functional role of the original amino acid while minimising disruption to the peptide fold and backbone conformation. Out of the 11 native substrate sequences, we selected two representatives for our investigations. The first sequence was the nsp4/5 P6-P5′ 11-mer substrate (TSAVLQ↓SGFRK, referred to as "s01"), which exhibits the most efficient hydrolysis by M^{pro}.[21] The second sequence was the nsp8/9 substrate (SAVKLQ↓NNELS, referred to as "s05"), which is conserved among coronaviruses[18] and whose

P4-P1 sequence demonstrates optimal activity amongst proteinogenic amino acids screened.[17] Alanine mutations were performed at each residue in the P4-P2′ region, including residues towards which M^{pro} demonstrates apparent selectivity (P4, P2, P1, and P1′) and residues displaying no apparent selectivity (P3 and P2′). Additionally, we performed mutation at P2 for the nsp5/6 substrate "s02" (SGVTFQ↓SAVKR), and the reported inhibitory peptide "p12" (KYTFWQ↓YSQFY), which we examined previously.[21] Analysis of s02 and p12 enables us to probe any difference in the structural responses when an alternative amino acid (Phe or Trp) is present in the S2 subsite. The dynamical response of M^{pro} towards mutation at each subsite reveals the structural adaptation required to accommodate each substrate/inhibitor residue, highlighting the relevance of the P1-Gln residue in substrate recognition by M^{pro}.

Results and Discussion

1. Equilibrium simulations of M^{pro}-peptide complexes

The models of dimeric M^{pro} non-covalently complexed with the three selected substrate peptides (s01, s02, s05) and inhibitor peptide (p12) employed for MD simulations were prepared as described previously.[21] In each case, the peptide was present at the substrate binding site formed predominantly by residues of the M^{pro} chain denoted as chain A (ChA), including the ChA His41-Cys145 catalytic dyad. The substrate binding site in chain B (ChB) was left unoccupied. Unless otherwise specified, for subsequent discussions we will refer to ChA residues. Each M^{pro}-peptide complex was subjected to five equilibrium MD simulations of 200 ns, totalling 1 μs per complex. Results from equilibrium MD simulations of M^{pro}-s05 have been reported in our previous study.[46]

The peptides s01, s02, and p12 demonstrated stable binding over the course of the simulations, with all systems considered equilibrated from t = 50 ns onwards, based on analyses of root mean square deviation (RMSD) and fluctuation (RMSF), protein secondary structure content, and M^{pro}-

peptide interactions, including 12 conserved hydrogen bonds (HBs 1-12) across the P4-P4′ positions[21] (**Supplementary Information** in **Appendix for Chapter 5: Section S5.1**). To investigate if these complexes represent the non-covalent pre-reaction complexes (sometimes referred to as Michaelis complexes), we assessed the binding of the P1 scissile amide in the active site by monitoring five relevant distances (**Figures 5.1b and S5.1.8**). They include the distances describing deprotonation of Cys145 by His41 *d*(His41_Nε2 – Cys145_Hγ) ("d_{Deprot}"), nucleophilic addition of Cys145 onto the scissile amide carbonyl carbon *d*(Cys145_Sγ – P1_C) ("d_{Nuc}"), and stabilisation of the carbonyl oxygen in the oxyanion hole by the backbone NHs of Cys145, Ser144, and Gly143 *d*(P1_O – Cys145/Ser144/Gly143_H) ("d_{OH145}/d_{OH144}/d_{OH143}", respectively). In all cases, while His41 was poised for deprotonating Cys145 during most of the simulations, and the scissile amide carbonyl bound stably in the oxyanion hole, d_{Nuc} showed a bimodal distribution, as observed in previous studies;[10, 54] this resulted from alternate conformations of the Cys145 sidechain, with the *gauche* and *trans* conformations corresponding respectively to shorter and longer d_{Nuc} distances (**Figures 5.1b and S5.1.8-11**).

Equilibrium MD simulations were also performed for s01 and s05 with the P1-Gln replaced by Ala, designated as s01-QP1A and s05-QP1A. The mutant peptides remained stably bound throughout the MD simulations, forming all conserved HBs except for HBs 6-7 which require a Gln sidechain (**Figure S5.1.12**). In both peptides, the P1 scissile amide group was stably bound in the oxyanion hole (**Figure S5.1.8**). However, analysis of d_{Nuc} distances indicated the predominance of the unproductive *trans* Cys145 conformation relative to the productive *gauche* conformation, when compared to the corresponding wild type (wt) peptides (**Figures S5.1.8-11**). This raised the questions of whether and how the amino acid identity at every position of the peptide affects the M^{pro} active site configuration. These questions were addressed using the D-NEMD approach, by "mutating" the residues of interest on the peptide to alanine. Initial configurations of M^{pro} complexed with wt peptides were taken from the five equilibrium MD simulations from t = 50 ns up to 195 ns in 5 ns intervals, yielding 150 pairs of short equilibrium

and nonequilibrium propagations in each case (N = 150). The subtraction technique in vector evaluation helps distinguish significant coherent displacements caused by the perturbation from random fluctuations of mobile atoms.[39, 41, 46]

Given the lack of conserved amino acid identity, ordered peptide structure, and well-defined M^{pro} subsites outside of P4-P2′ of substrate peptides,[12, 14, 18, 20, 21] mutation was only considered in the peptide P4-P2′ positions, excluding cases where the original amino acid was Ala or Gly (*e.g.*, P4 and P2′ of s01). In all cases, the D-NEMD responses were measured up to 1 ns as in previous studies.[46] The responses were quantified by the following approaches. Firstly, to characterise any statistically significant changes in the conformation of and peptide binding in the active site, differences in the five active-site distances between nonequilibrium (d_{neq}) and equilibrium (d_{eqm}) propagations at equivalent timepoints, $\Delta d = d_{neq} - d_{eqm}$, were evaluated. Secondly, to consider the overall effect on M^{pro}-peptide binding, similarly the differences in the donor-acceptor distances ($\Delta d_{D\text{-}A}$) of the conserved HBs 1-5 and 8-12, all of which involve the peptide backbone, were calculated. HBs 6-7, which are dependent on the P1-Gln sidechain, were excluded. Thirdly, the displacement vector, $\boldsymbol{v} = \boldsymbol{r}_{neq} - \boldsymbol{r}_{eqm}$ where $\boldsymbol{r}$ is atomic position, of each of the 612 M^{pro} and 11 peptide Cα atoms, was measured and averaged across the 150 D-NEMD replicas, with Cα atoms chosen to represent the most substantial structural changes in the protein.[38, 46] In regions of interest according to the above analyses, average displacement vectors were evaluated additionally for non-hydrogen atoms.

2. Effect of P1 Gln-to-Ala mutation

The role of the P1-Gln, a residue conserved across all native M^{pro} substrates, was investigated by mutating P1-Gln to Ala in s01 and s05, with the response analysed using the D-NEMD approach (**Figure 5.1a**; **Section S5.2**). Analysis of the HB interactions between the P1 scissile amide carbonyl and the oxyanion hole ($\Delta d_{OH145}/\Delta d_{OH144}/\Delta d_{OH143}$, **Figures 5.1b,c and S5.2.1**; **Table**

S5.2.1) showed narrow distributions, indicating stability of these interactions upon mutation. The distance describing deprotonation of Cys145 by His41 ("d_{Deprot}") showed a wider spread but was overall centred around zero. The largest differences were observed in the Cys145_Sγ – P1_C (d_{Nuc}) distance, with predominantly positive differences for both s01 and s05, especially after 10 ps post perturbation. For example, the average Δd_{Nuc} at t = 100 ps was (0.37 ± 0.06) Å for s01 and (0.19 ± 0.05) Å for s05 (error bar = SEM), indicating a statistically significant increase. The increase of this distance was caused by the movement of Cys145 Sγ away from the scissile amide carbonyl, which remained rigid according to the unaffected d_{OH} distances; in this case, the Cys145 sidechain switched from the productive *gauche* to the unproductive *trans* conformation.

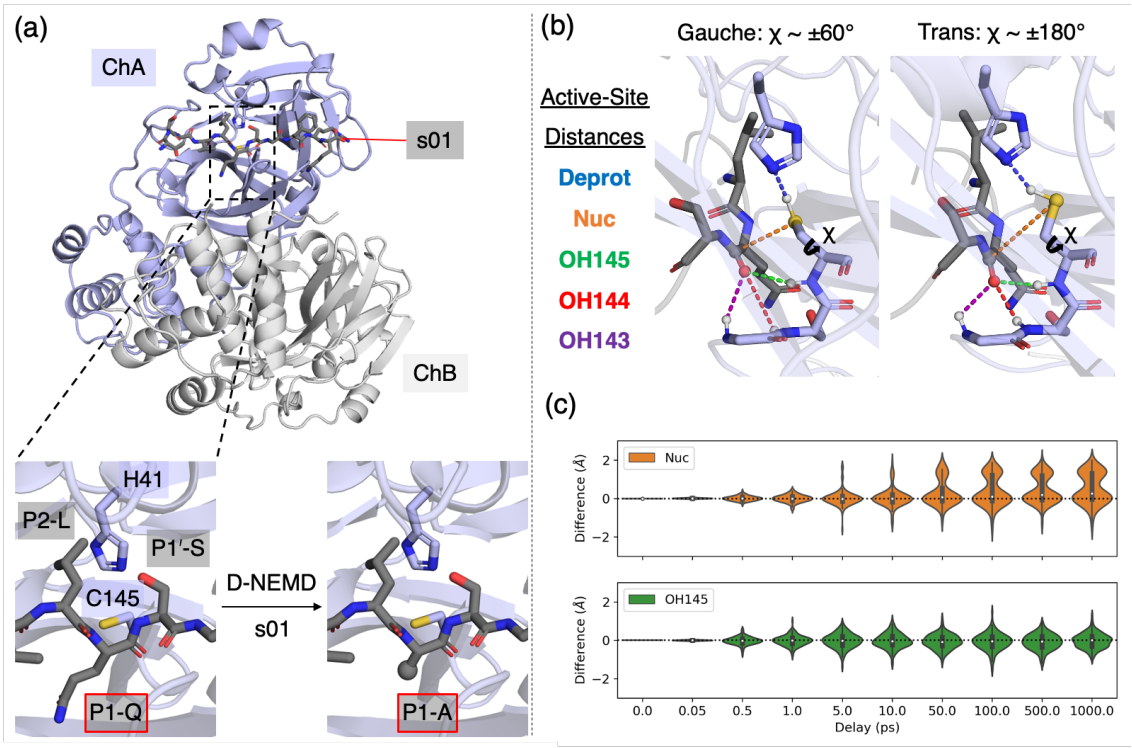


Figure 5.1: D-NEMD P1 Gln-to-Ala mutation. (a) View of the dimeric M^{pro}-s01 complex prior to MD simulations, with the P1 Gln-to-Ala mutation perturbation studied by the D-NEMD method. (b) Illustration of the five active-site distances using two equilibrium MD-derived snapshots, corresponding to the *gauche* and *trans* Cys145 conformations, which show distinct χ dihedrals. (c) Distributions of the differences in d_{Nuc} and d_{OH145} following the P1 mutation in s01.

Analysis of HBs 1-3 and 8-12 (**Table S5.2.2-3**; **Figure S5.2.2**; note that HBs 6-7 require a Gln sidechain and hence do not exist here) showed little variation with $\Delta d_{D\text{-}A}$ values around zero. A wider distribution was observed for HB4, likely due to its less conserved nature and its involvement of the flexible Gln189 sidechain. However, there was a consistent and statistically significant decrease in the HB5 $d_{D\text{-}A}$ distance between the peptide P1 backbone N and His164 backbone O (**Table S5.2.2**), with $\Delta d_{D\text{-}A}$(HB5) = (-0.23 ± 0.03) Å for s01 and (-0.16 ± 0.02) Å for s05 at t = 100 ps. Surprisingly, this suggests a tighter binding of the P1 backbone following the removal of the P1-Gln sidechain.

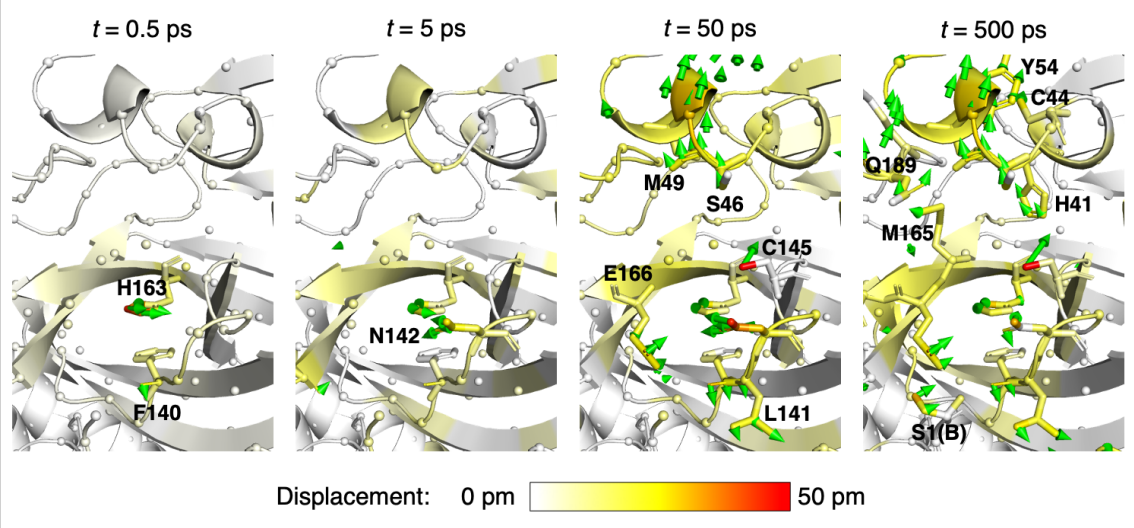


Figure 5.2: The response of the M^{pro} S1 subsite residues to P1 Gln-to-Ala mutation in s01. The response was measured by averaging the non-hydrogen displacement vectors between the equilibrium and nonequilibrium propagations at selected timepoints following the mutation. Displacement magnitudes are displayed on a white-yellow-red scale, using the structure prior to MD. Significant vectors with length ≥20 pm are shown as green arrows with a scale-up factor of 5.[55, 56] Residues that show such significant displacements located within 10 Å of the substrate P1 residue are visualised as sticks.

The propagation of signal arising from the P1 Gln-to-Ala mutation was visualised using the non-hydrogen atomic displacement vectors averaged across 150 D-NEMD replicas, with a focus on the M^{pro} S1 pocket (**Figure 5.2**). Initially, at t = 0.5 ps following the mutation, the largest displacements were observed in the M^{pro} His163 sidechain moving outwards into the space originally occupied by the P1-Gln sidechain, and in the Phe140 backbone which moved

downwards, likely due to the loss of HB7. At t = 5 ps, the signal propagated outwards, with displacements observed in neighbouring residues such as the Asn142 sidechain. The P1-Ala backbone and Cβ atoms moved inwards (**Figure S5.2.7**), resulting in the observed decrease in $d_{D\text{-}A}$(HB5). At t = 50 ps, the Cys145 Sγ atom showed an upward displacement away from the P1 scissile amide carbonyl, accompanied by the movement of the Glu166 sidechain closing into S1, the displacement of the Leu141 sidechain, and the upward displacement of residues in the Ser46-Leu50 helix above the active site. The retreat of the Cys145 Sγ atom from the P1 carbonyl, caused by a change in the Cys145 sidechain conformation from *gauche* to *trans*, continued to later timepoints including t = 500 ps, when the His41 residue was also impacted and pulled away from the active site. Similar effects on Cys145 were observed in the case of s05 from t = 50 ps onwards (**Figure S5.2.6**).

The results demonstrate that the recognition of the conserved P1-Gln in substrates is communicated to M^{pro} residues beyond those directly binding to its sidechain. Structural adaptations are observed in residues composing the S1 pocket and the helices above the active site (**Figure S5.2.6**). Notably, the mutation causes a change in Cys145 from the productive *gauche* to the unproductive *trans* conformation, along with a potential retreat of the His41 sidechain, thus disabling reaction initiation. Conversely, the binding of the P1-Gln sidechain, and its associated interactions with His163 and Phe140 in the S1 pocket, would help establish the positioning of the P1 backbone in the active site, and activates the catalytic dyad residues His41 and Cys145 into their productive conformations. This recognition and communication process helps ensure that M^{pro} specifically targets the backbone carbonyl of a Gln residue.

3. Effect of P2 mutation

To investigate if there is any coordinated dynamic behaviour of M^{pro} regions upon binding of the P2 residue, which corresponds to Leu in 9 of the 11 native substrates, except for s02 and s03, which have Phe and Val respectively, we performed D-NEMD on s01 and s05 by mutating P2-

Leu to Ala (**Section S5.3**). The active-site distances and the HB $d_{D\text{-}A}$ distances remained mostly unaffected (**Tables S5.3.1-3**; **Figures S5.3.1-2**). According to displacement vectors, M^{pro} remained mostly unperturbed relative to equilibrium following the mutation at the P2 site (**Figures 5.3 and S5.3.5-6**). The immediate response of M^{pro} was mostly localised to S2 residues, such as the sidechains of Met49 and Met165. This remained to be the case in the tens of picoseconds following the perturbation. These findings suggest that minimal structural rearrangement of M^{pro} residues is required to accommodate the P2-Leu.

In our previous study, we reported novel peptide sequences that competitively inhibit M^{pro} and are resistant to hydrolysis, which we attributed to the binding of a P2-Trp in the S2 pocket.[21] To determine if there are differences in M^{pro} dynamics when binding an aromatic ring in S2, we mutated the P2-Phe residue in s02 and the P2-Trp residue in p12 to Ala, with the response analysed using D-NEMD. While the mutations did not significantly affect the active-site distances, the flexible HB4 $d_{D\text{-}A}$ decreased in both systems, with average $\Delta d_{D\text{-}A}$ = (-0.37 ± 0.12) Å and (-0.23 ± 0.11) Å at t = 100 ps for s02 and p12, respectively. For s02, $d_{D\text{-}A}$(HB1) was also reduced. As the $d_{D\text{-}A}$ of HBs 2-3 that bind the peptide P3 backbone remained unaffected, the decreases in $d_{D\text{-}A}$ of HB4, and in the s02 case also HB1, suggested an impact of the P2-Phe/Trp on the conformations of Thr190 and/or Gln189. This impact was visualised with displacement vectors (**Figure 5.3**). Compared to s01 and s05 where signals were mostly localised on Met49 and Met165, the mutations in s02 and p12 resulted in significant downward movements of residues in the Ser46-Leu50 helix above the active site, the Asn53-Arg60 helix further above, and the connecting loop in between. Outward movements of the Phe185-Gln192 loop, whose residues line the S4 and S2 pockets, were observed. Among the loop residues, both the backbone and sidechain atoms of Asp187 were significantly displaced. Asp187 sidechain carboxylate coordinates a conserved water molecule,[21] and this water is bridged also with His164 and the catalytic His41 residue, playing a role in stabilising the positioning and positive charge of His41 during the catalytic cycle.[57-59] Therefore, displacement in Asp187 may be transmitted to His41

and affect catalysis. Indeed, at timepoints in the hundreds of picoseconds timescale (*e.g.*, t = 500 ps) for p12, small but statistically significant displacements were observed in the His41 sidechain, which was propelled forwards towards the P1′-Tyr of p12. This demonstrates the possibility of communication between the P2-Trp and His41, as we previously proposed.[21] The distortion required in M^{pro} residues in the Phe185-Gln192 loop, including Asp187, to accommodate the larger P2-Phe/Trp in s02/p12, may explain the poorer or lack of hydrolytic efficiency towards these peptides.[21]

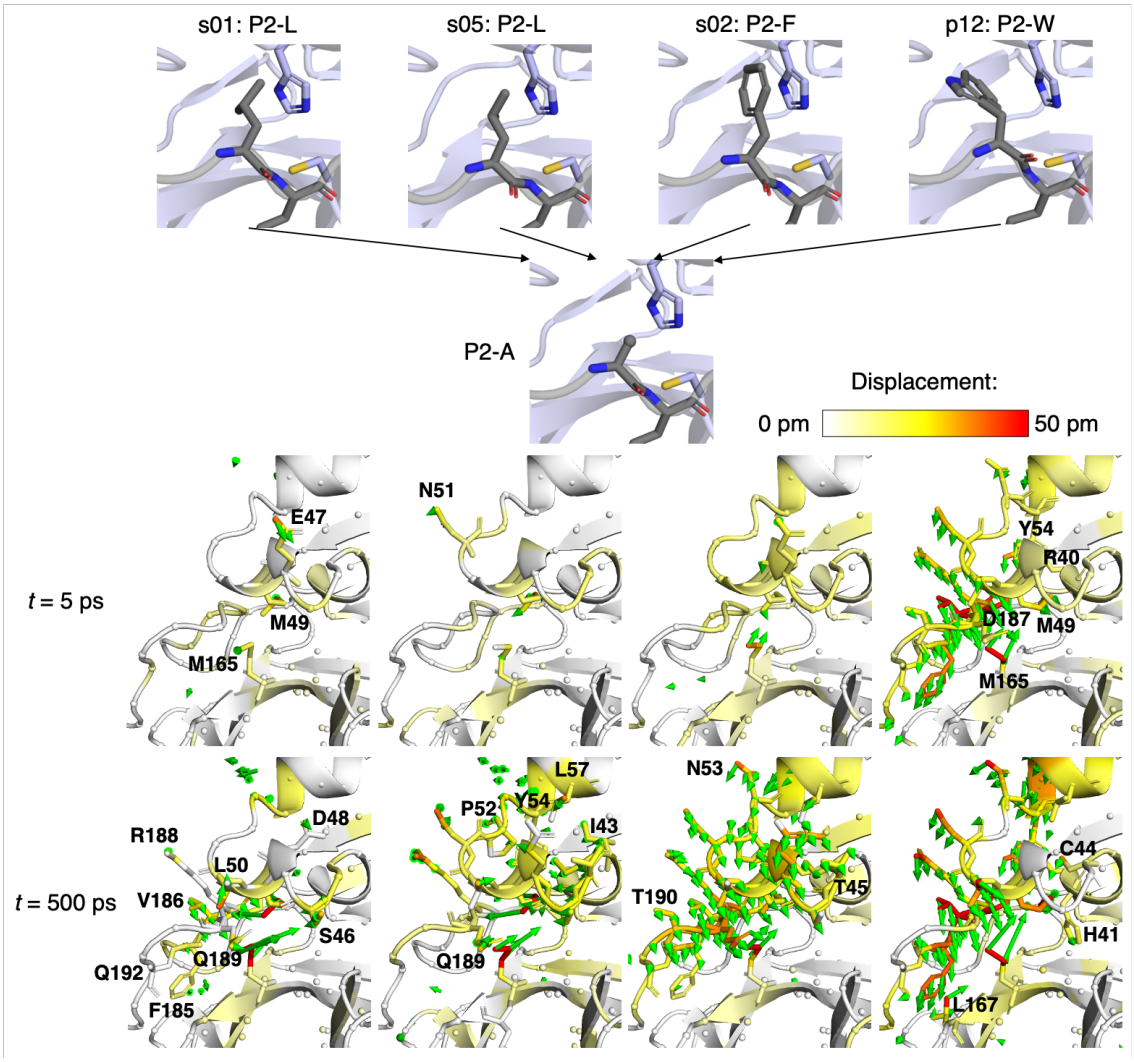


Figure 5.3: M^{pro} response to P2 mutation in substrate and inhibitor peptides. The structural response was measured by averaging the non-hydrogen displacement vectors between the equilibrium and nonequilibrium propagations at selected timepoints following the mutation. Displacement magnitudes are displayed on a white-yellow-red scale, using the structure prior to MD. Significant vectors with length ≥20 pm are shown as green arrows with a scale-up factor of 5.[55, 56] Residues that show such significant displacements located within 10 Å of the peptide P2 residue are visualised as sticks.

4. Mutation of P3 and P4 residues

Further upstream of the scissile amide in the substrate sequence, the P3 position is apparently tolerant of investigated proteinogenic amino acids. While its backbone atoms are coordinated to the Glu166 backbone in HBs 2-3, the P3 sidechain is solvent exposed and demonstrates flexibility relative to the neighbouring P4 and P2 residues (**Figure S5.1.7**). Upon mutation of the respective P3-Val and P3-Lys in s01 and s05 to Ala and analysis of the system response using D-NEMD (**Section S5.4A**), no significant deviations in the active-site distances were observed. The only consistently significant deviation in HB $d_{D\text{-}A}$ was observed in HB4 caused by the P3 Lys-to-Ala mutation in s05 (**Table S5.4A.2**), where $d_{D\text{-}A}$ was on average increased from t = 50 ps onwards, with Δd = (0.74 ± 0.13) Å at t = 500 ps. During the equilibrium MD simulations, the P3-Lys sidechain occasionally formed a loop around the peptide P2 backbone, with the simultaneous coordination of Gln189 sidechain carbonyl oxygen to the P2 backbone NH (HB4) and the P3-Lys sidechain amino group. This coordination and shielding from solvent likely contributed to positioning of Gln189,[18] but these effects were lost upon the mutation of the P3-Lys sidechain to Ala. In terms of displacement, the effect of mutating P3 was minimal and less consistent over the observed timescales, compared to mutations of P1 or P2, which was expected given the lack of a well-defined pocket on M^{pro} that encloses the P3 sidechain and the associated lack of substrate specificity (**Figures S5.4A.5-6**).

Mutation of the P4 residue was only conducted on s05, as P4 was already Ala in s01. Mutating P4-Val to Ala in s05 did not significantly affect the active-site distances (**Section S5.4B**). However, there was an average decrease in $d_{D\text{-}A}$ of HBs 1 and 4 (**Table S5.4B.2**). Analysis of the displacement vectors revealed that in the picosecond timescale following the mutation (*e.g.*, t = 5 ps), the M^{pro} loop residues Val186-Ala193 responded by closing around the S4 pocket, with movements observed in both the backbone and sidechain atoms (**Figures 5.4 and S5.4B.5-6**). The signal propagated downstream along the loop to residues including Thr196 at t = 500 ps, as the loop adapted to the mutation of the P4-Val sidechain. This dynamic response is in line with

previous crystallographic observations that the Val186-Gly195 loop is structurally variable depending on the bound substrate sequence,[20] and echoes the coherent D-NEMD response observed when the full s05 substrate peptide is removed.[46] In comparison to the M^{pro} response when the P2-Leu was mutated to Ala in the same s05 substrate, here the S4 pocket needed to undergo more significant structural adaptation to accommodate the substrate P4-Val, compared to the relatively rigid S2 pocket that accommodates the predominantly conserved P2-Leu residue.

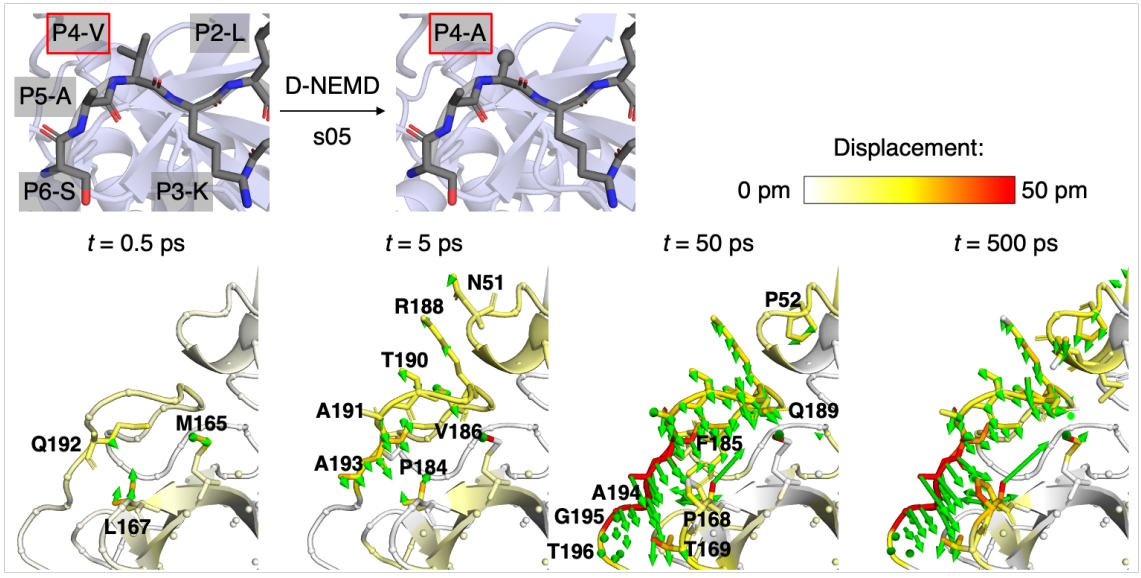


Figure 5.4: M^{pro} response to P4 mutation in s05. The response was measured by averaging the non-hydrogen displacement vectors between the equilibrium and nonequilibrium propagations at selected timepoints following the mutation. Displacement magnitudes are displayed on a white-yellow-red scale, using the structure prior to MD. Significant vectors with length ≥20 pm are shown as green arrows with a scale-up factor of 5.[55, 56] Residues that show such significant displacements located within 10 Å of the substrate P4 residue are visualised as sticks.

5. Mutation of P1′ and P2′ residues

The S subsites on M^{pro} that bind the substrate P residues are more well-defined and form more HB interactions than the S′ subsites which bind residues C-terminal of the scissile amide.[21] In the P′ side, substrate specificity has only been reported at the P1′ position, which requires a small, neutral amino acid, with Ser, Ala, and Asn observed among SARS-CoV-2 M^{pro} native substrates,[18] and additionally Gly in the nsp5/6 sequence of SARS-CoV.[60] We analysed the response of M^{pro} to

Ala mutation at the P1′-Ser/Asn site in s01/s05, respectively, using D-NEMD (**Section S5.5A**). Among the active-site distances analysed, we observed an average decrease in the d_{Deprot} distance, with s05 exhibiting a consistently negative Δd_{Deprot} from t = 0.25 ps onwards (**Table S5.5A.1**). Analysis of the displacement vectors for s05 revealed a downward movement of the His41 sidechain to form a closer association with Cys145 and the scissile amide carbonyl, *e.g.*, at t = 5 ps and 500 ps (**Figure 5.5**). In the equilibrium MD, His41 was observed to engage in HB with the P1′-Asn sidechain in s05 (25% of the frames), and to a lesser extent with the P1′-Ser sidechain in s01 (18%). These HBs may compete with the coordination of His41 to Cys145 for initiating deprotonation, potentially contributing to the experimentally observed slow hydrolysis of s05[21] despite its optimal P4-P1 sequence.[17] Indeed, a P1′ Asn-to-Ala substitution in a similar coronavirus nsp8/9-derived substrate was reported to result in approximately three times more efficient hydrolysis than the wt peptide with P1′-Asn.[18] However, experimental validation is required to confirm this hypothesis, since s05 is the only native substrate possessing a P1′-Asn.[21] Among the M^{pro}-peptide HBs (**Tables S5.5A.2-3**), an average distance decrease was observed in HB4 in s01 after t = 300 ps, accompanied by a displacement of the Gln189 sidechain towards the P2 backbone (**Figure S5.5A.6**).

No mutation was performed at P2′ for s01 which has P2′-Gly, while for s05, P2′-Asn was mutated to Ala (**Section S5.5B**). Overall, the mutation did not cause significant deviations in the active-site or HB distances, including HBs 10-11 involving the P2′ backbone. Additionally, there were no significant displacements in M^{pro} residues. This is consistent with the lack of substrate specificity at P2′, and the solvent exposed and flexible nature of its sidechain.

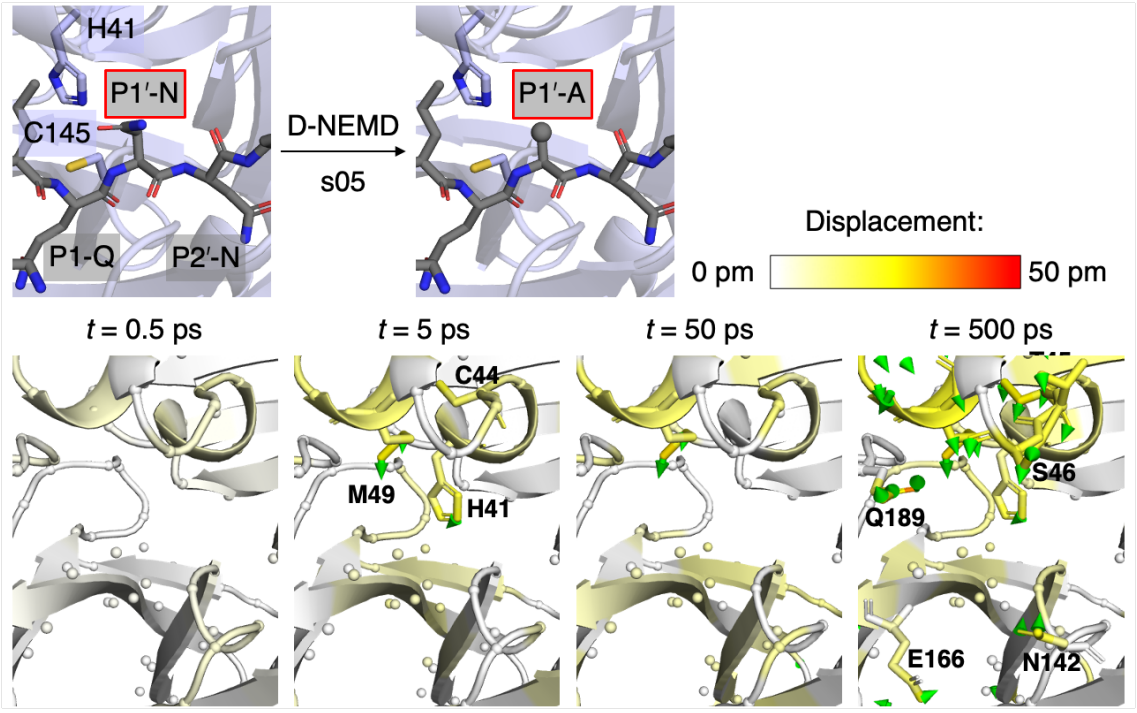


Figure 5.5: M^{pro} response to P1′ mutation in s05. The response was measured by averaging the non-hydrogen displacement vectors between the equilibrium and nonequilibrium propagations at selected timepoints following the mutation. Displacement magnitudes are displayed on a white-yellow-red scale, using the structure prior to MD. Significant vectors with length ≥20 pm are shown as green arrows with a scale-up factor of 5.[55, 56] Residues that show such significant displacements located within 10 Å of the substrate P1′ residue are visualised as sticks.

Conclusions

The D-NEMD method is an emerging strategy for probing signal propagation and communication networks upon a perturbation. It has been employed in various studies to examine dynamic processes of proteins, including ligand and substrate binding, as well as changes in pH.[38-46] By combining equilibrium and nonequilibrium simulations, the D-NEMD approach enables the identification of statistically significant changes by averaging the responses across hundreds of replicas. In a recent study, we utilized this method to identify communication pathways from the SARS-CoV-2 M^{pro} active site to a known allosteric site, and to sites relevant to drug resistant mutations, by entirely annihilating the native substrate s05.[46] In this work, we used an alanine mutation strategy commonly employed in experimental and computational studies, to investigate

how M^{pro} structurally responds to binding of substrate residues in each of its subsites. This analysis is crucial for understanding substrate specificity and recognition.

Our findings reveal significant differences in the nature and magnitude of the response of M^{pro} residues to the removal of sidechain moieties at each subsite, carried out by mutating each residue to alanine. As expected, mutation in residues with no specificity requirement, such as the P3 and P2′ positions, had limited effect, serving as control experiments. In contrast, mutations of P1′ and P1 residues induced changes in the active site. In P1, this mutation resulted in significant conformational changes in the catalytic dyad, by switching the Cys145 from an active to an inactive conformation, thereby inhibiting the chemical reaction. The binding of P1-Gln, which is conserved among all native substrates of M^{pro}, would thus help activate the enzyme and ensure that M^{pro} only cleaves at the designated positions in the viral polyproteins.

We also explored the dynamic behaviour of the S4 and S2 subsites in accommodating their corresponding P4 and P2 substrate residues. The S4 pocket is known to accommodate residues of various sizes, including Ala, Pro, Val, and Thr, enabled by the dynamic adaptation of the S4 pocket lined by the Val186-Gln192 loop, and possibly involving collective movements of downstream residues (Ala193-Thr196). Accommodation of the P4 residue thus requires M^{pro} to undergo an induced fit. By contrast, the S2 pocket appears to be minimally affected by the truncation of P2-Leu, perhaps reflecting a binding mode more closely resembling a lock-and-key model, where the pocket is optimised in size and shape for accommodating the P2-Leu, present in 9 of the 11 polyprotein cleavage site sequences. Indeed, the rigid, Leu-mimicking 6,6-dimethyl-3-azabicyclo[3.1.0]hexane group in nirmatrelvir was observed to fully occupy this pocket.[7] However, accommodation of alternative residues, such as P2-Phe of s02 and P2-Trp of the inhibitor peptide p12, which is not efficiently hydrolysed by M^{pro}, led to more significant adaptations in M^{pro} residues, including those related to the catalytic mechanism. These adaptations might hinder peptide hydrolysis by M^{pro}.

To characterize the distinct response by M^{pro} following each substrate mutation, we analysed changes in key distances reflecting active site conformation and protein-peptide binding. We also visualised statistically significant displacement vectors and presented them as plots of the displacement magnitudes of each protein Cα or non-hydrogen atom (**Figures S5.2.3-4, S5.3.3-4, S5.4A.3-4, S5.4B.3-4, S5.5A.3-4, S5.5B.3-4**). Every plot maps the response to a specific perturbation, in this case, a mutation in the bound peptide. These plots act as dynamic fingerprints, with peaks corresponding to residues or atoms displaying the largest structural adaptations. It is of interest to investigate whether such dynamic fingerprint mapping could be applied to cases involving modifications to bound ligands, in a similar manner as free energy perturbation methods informing on binding free energy differences, to characterise whether and how a ligand modification is tolerated and might trigger a response in the structure and dynamics of the target receptor. Some drug targets, such as the HIV-1 protease, have been shown to develop drug resistance through distal mutations altering protein dynamic behaviour.[61] Therefore, the dynamic mapping offered by the D-NEMD approach has the potential to provide useful insights into residues susceptible to drug resistant mutations due to changes in receptor dynamics.

Methods

Equilibrium molecular dynamics (MD). The SARS-CoV-2 M^{pro} dimer (PDB 6YB7;[62] 1.25 Å resolution) non-covalently complexed with a polyprotein-derived[19] native substrate peptide (s01, s02, s05) or an inhibitor peptide (p12)[21] in the chain A (ChA) active site was prepared as described previously.[21, 46] The substituted s01-QP1A and s05-QP1A peptides were modelled by truncating the P1-Gln sidechains to Ala in s01 and s05 respectively. Cys145 and His41 were both neutral, and the setup of titratable and histidine residues (**Table S5.6.1**) was identical to previous setups.[21, 46] Using the GROMACS (v 2019.2/2020.3/2020.4)[63] MD package and the AMBER99SB-ILDN forcefield,[64] and following previously reported procedures,[21, 46] each M^{pro}-peptide complex was solvated with TIP3P water[65] in a rhombic dodecahedral box with edges at least 1 nm away from the complex and neutralised by sodium ions, with the total number of atoms being 86,305, 86,326, 86,304, 86,241, 86,301, and 86,297 for the s01, s02, s05, p12, s01-QP1A, and s05-QP1A complexes respectively. The system was subjected to energy minimisation until convergence with maximum force <1000 kJ mol^{-1} nm^{-1}. Starting from random velocities from the Maxwell-Boltzmann distribution at 298.15 K, five replicas were initiated by a 200 ps (1 fs step) NVT and 200 ps (1 fs step) NPT equilibration, followed by a 200 ns (2 fs step) production MD run. A temperature of 298.15 K was maintained by velocity rescaling with a stochastic term, with separate coupling to the protein and non-protein (time constant = 0.1 ps).[66] A pressure of 1.0 bar was maintained by a Parrinello-Rahman barostat (time constant = 2 ps).[67, 68] Van der Waals interactions were cut off at 1 nm. Long-range electrostatic interactions beyond 1 nm were calculated using smooth particle mesh Ewald.[69, 70] Frames were saved every 100 ps and velocities were saved every 1 ns. Trajectories were analysed using GROMACS (v 2019.2)[63] and DSSP (v 2.0.4).[71, 72] Hydrogen bonds were defined according to the default criteria in gmx hbond, when donor-acceptor distance $d_{D\text{-}A} \leq 3.5$ Å and hydrogen-donor-acceptor angle $\angle_{H\text{-}D\text{-}A} \leq 30°$.

Dynamical nonequilibrium MD (D-NEMD) setup: D-NEMD simulations were set up and analysed similarly to previously reported procedures.[46] For every M^{pro}-peptide complex system,

frames were extracted from the five equilibrium MD simulations in 5 ns intervals from t = 50 ns to 195 ns, resulting in 5 × 30 = 150 configurations. To initiate the nonequilibrium simulations, in the coordinates and velocities of these configurations, the residue of interest was instantaneously mutated to alanine by truncation up to the Cβ atom and its bonded protons (Hβ1-3), with the Cγ atom modified to Hβ3, and the simulations were run using the Ala topology. No new atoms were added, except in the s05 P3 Lys-to-Ala mutation where a water molecule was replaced by a sodium ion to preserve overall charge neutrality. The corresponding equilibrium simulations were run using the original unmodified topology. To obtain the responses over different timescales, the pair of equilibrium and nonequilibrium simulations were conducted over two length scales: (i) over 10 ps saving frames every 50 fs; (ii) over 1 ns saving frames every 10 ps. For the s05 P1 Gln-to-Ala mutation, the length scales of simulations were as previously reported.[46]

D-NEMD analysis: The M^{pro}-peptide complex structure at a certain post-mutation timepoint was extracted from the nonequilibrium trajectory and superimposed onto the structure from the equilibrium trajectory at the equivalent timepoint, fitted based on the 612 M^{pro} Cα atoms. The two structures were compared by evaluating the equilibrium-to-nonequilibrium displacement vector (x, y, z) of every non-hydrogen atom, and the displacement vectors were averaged across the 150 replicas to yield the average vector, which, according to the Kubo-Onsager relation,[37, 38] estimates the macroscopic time-dependent response of the M^{pro}-peptide complex to the mutation. The magnitude of the average vector v was evaluated by:

$$v = \sqrt{(\bar{x})^2 + (\bar{y})^2 + (\bar{z})^2} \tag{5.1}$$

where $\bar{x}$, $\bar{y}$, $\bar{z}$ are the displacements in the x, y, and z directions, respectively, of the average vector $\boldsymbol{v}$. Based on error propagation, the standard deviation in v, s_v, can be evaluated from the standard deviations in x, y, and z, denoted respectively by s_x, s_y, and s_z:

$$s_v = \frac{\sqrt{\left(\bar{x} \cdot s_x\right)^2 + \left(\bar{y} \cdot s_y\right)^2 + \left(\bar{z} \cdot s_z\right)^2}}{\bar{v}} \tag{5.2}$$

The standard error of the mean (SEM; N = 150) in the magnitude of the average vector was evaluated to assess statistical significance:

$$\text{SEM} = \frac{s_v}{\sqrt{N}} \tag{5.3}$$

A response was considered significant only if the magnitude was greater than zero by at least 2 × SEM. Insignificant responses were treated as zero. The deviation vectors were visualised on the pre-MD M^{pro}-peptide complex with the aid of PyMOL (v 2.3.0)[55] and the modevectors.py script.[56] The five active-site distances and ten HB $d_{D\text{-}A}$ distances were outputted from both the equilibrium d_{eqm} and nonequilibrium d_{neq} propagations. The distance differences $\Delta d = d_{neq} - d_{eqm}$ at equivalent timepoints were calculated, plotted as a violin plot to visualise the distributions, and averaged across the 150 replicas with statistical significance assessed by SEM.

Data Availability

Detailed setup and analyses of the equilibrium and nonequilibrium MD simulations are included in the **Appendix for Chapter 5 (Supplementary Information to Manuscript III)**.

References

1. Moustaqil, M.; Ollivier, E.; Chiu, H. P.; Van Tol, S.; Rudolffi-Soto, P.*, et al.* SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg. Microbes Infect.* **2021,** *10*, 178.
2. Lv, Z.; Cano, K. E.; Jia, L.; Drag, M.; Huang, T. T.*, et al.* Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development. *Front. Chem.* **2022,** *9*, 819165.
3. Yan, F.; Gao, F. An overview of potential inhibitors targeting non-structural proteins 3 (PLpro and Mac1) and 5 (3CLpro/Mpro) of SARS-CoV-2. *Comput. Struct. Biotechnol. J.* **2021,** *19*, 4868.
4. Amin, S. A.; Banerjee, S.; Ghosh, K.; Gayen, S.; Jha, T. Protease targeted COVID-19 drug discovery and its challenges: Insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors. *Bioorg. Med. Chem.* **2021,** *29*, 115860.
5. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.*, et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
6. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.*, et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368*, 409.
7. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.*, et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374*, 1586.
8. Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.*, et al.* Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022,** *65*, 6499.
9. Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem. Sci.* **2020,** *11*, 10626.
10. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020,** *10*, 12544.
11. Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.*, et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020,** *11*, 5877.
12. Zhao, Y.; Zhu, Y.; Liu, X.; Jin, Z.; Duan, Y.*, et al.* Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* **2022,** *119*, e2117142119.
13. Kneller, D. W.; Zhang, Q.; Coates, L.; Louis, J. M.; Kovalevsky, A. Michaelis-like complex of SARS-CoV-2 main protease visualized by room-temperature X-ray crystallography. *IUCrJ* **2021,** *8*, 973.
14. Koudelka, T.; Boger, J.; Henkel, A.; Schönherr, R.; Krantz, S.*, et al.* N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *Proteomics* **2021,** *21*, 2000246.
15. Gallo, G.; Barcick, U.; Coelho, C.; Salardani, M.; Camacho, M. F.*, et al.* A proteomics-MM/PBSA dual approach for the analysis of SARS-CoV-2 main protease substrate peptide specificity. *Peptides* **2022,** *154*, 170814.
16. Cho, E.; Rosa, M.; Anjum, R.; Mehmood, S.; Soban, M.*, et al.* Dynamic Profiling of β-Coronavirus 3CL Mpro Protease Ligand-Binding Sites. *J. Chem. Inf. Model.* **2021,** *61*, 3058.
17. Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.*, et al.* SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2021,** *17*, 222.
18. MacDonald, E. A.; Frey, G.; Namchuk, M. N.; Harrison, S. C.; Hinshaw, S. M.*, et al.* Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **2021,** *7*, 2591.
19. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.*, et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265.
20. Shaqra, A. M.; Zvornicanin, S. N.; Huang, Q. Y. J.; Lockbaum, G. J.; Knapp, M.*, et al.* Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nat. Commun.* **2022,** *13*, 3556.
21. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
22. Nepali, K.; Sharma, R.; Sharma, S.; Thakur, A.; Liou, J.-P. Beyond the vaccines: a glance at the small molecule and peptide-based anti-COVID19 arsenal. *J. Biomed. Sci.* **2022,** *29*, 65.
23. Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021,** *9*, 622898.

24. Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.*, et al.* Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021,** *12*, 1433.
25. Brewitz, L.; Dumjahn, L.; Zhao, Y.; Owen, C. D.; Laidlaw, S. M.*, et al.* Alkyne Derivatives of SARS-CoV-2 Main Protease Inhibitors Including Nirmatrelvir Inhibit by Reacting Covalently with the Nucleophilic Cysteine. *J. Med. Chem.* **2023,** *66*, 2663.
26. Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.*, et al.* Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020,** *30*, 678.
27. Bai, B.; Belovodskiy, A.; Hena, M.; Kandadai, A. S.; Joyce, M. A.*, et al.* Peptidomimetic α-Acyloxymethylketone Warheads with Six-Membered Lactam P1 Glutamine Mimic: SARS-CoV-2 3CL Protease Inhibition, Coronavirus Antiviral Activity, and in Vitro Biological Stability. *J. Med. Chem.* **2022,** *65*, 2905.
28. Miura, T.; Malla, T. R.; Owen, C. D.; Tumber, A.; Brewitz, L.*, et al.* In vitro selection of macrocyclic peptide inhibitors containing cyclic γ2,4-amino acids targeting the SARS-CoV-2 main protease. *Nat. Chem.* **2023,** *15*, 998.
29. Sacco, M. D.; Ma, C.; Lagarias, P.; Gao, A.; Townsend, J. A.*, et al.* Structure and inhibition of the SARS-CoV-2 main protease reveals strategy for developing dual inhibitors against Mpro and cathepsin L. *Sci. Adv.* **2020**, eabe0751.
30. Kneller, D. W.; Galanie, S.; Phillips, G.; O'Neill, H. M.; Coates, L.*, et al.* Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals. *Structure* **2020,** *28*, 1313.
31. Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q.*, et al.* Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020,** *11*, 4417.
32. Oerlemans, R.; Ruiz-Moreno, A. J.; Cong, Y.; Dinesh Kumar, N.; Velasco-Velazquez, M. A.*, et al.* Repurposing the HCV NS3–4A protease drug boceprevir as COVID-19 therapeutics. *RSC Med. Chem.* **2021,** *12*, 370.
33. Andi, B.; Kumaran, D.; Kreitler, D. F.; Soares, A. S.; Keereetaweep, J.*, et al.* Hepatitis C virus NS3/4A inhibitors and other drug-like compounds as covalent binders of SARS-CoV-2 main protease. *Sci. Rep.* **2022,** *12*, 12197.
34. Flynn, J. M.; Samant, N.; Schneider-Nachum, G.; Barkan, D. T.; Yilmaz, N. K.*, et al.* Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms. *eLife* **2022,** *11*, e77433.
35. Ciccotti, G.; Jacucci, G. Direct Computation of Dynamical Response by Molecular Dynamics: The Mobility of a Charged Lennard-Jones Particle. *Phys. Rev. Lett.* **1975,** *35*, 789.
36. Ciccotti, G.; Jacucci, G.; McDonald, I. R. "Thought-experiments" by molecular dynamics. *J. Stat. Phys.* **1979,** *21*, 1.
37. Ciccotti, G.; Ferrario, M. Non-equilibrium by molecular dynamics: a dynamical approach. *Mol. Simul.* **2016,** *42*, 1385.
38. Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur. Phys. J. B* **2021,** *94*, 144.
39. Oliveira, A. S. F.; Shoemark, D. K.; Campello, H. R.; Wonnacott, S.; Gallagher, T.*, et al.* Identification of the Initial Steps in Signal Transduction in the α4β2 Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. *Structure* **2019,** *27*, 1171.
40. Oliveira, A. S. F.; Edsall, C. J.; Woods, C. J.; Bates, P.; Nunez, G. V.*, et al.* A General Mechanism for Signal Propagation in the Nicotinic Acetylcholine Receptor Family. *J. Am. Chem. Soc.* **2019,** *141*, 19953.
41. Galdadas, I.; Qu, S.; Oliveira, A. S. F.; Olehnovics, E.; Mack, A. R.*, et al.* Allosteric communication in class A β-lactamases occurs via cooperative coupling of loop dynamics. *eLife* **2021,** *10*, e66567.
42. Dommer, A.; Casalino, L.; Kearns, F.; Rosenfeld, M.; Wauer, N.*, et al.* #COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol. *Int. J. High Perform. Comput. Appl.* **2022,** *37*, 28.
43. Oliveira, A. S. F.; Shoemark, D. K.; Avila Ibarra, A.; Davidson, A. D.; Berger, I.*, et al.* The fatty acid site is coupled to functional motifs in the SARS-CoV-2 spike protein and modulates spike allosteric behaviour. *Comput. Struct. Biotechnol. J.* **2022,** *20*, 139.
44. Oliveira, A. S. F.; Shoemark, D. K.; Davidson, A. D.; Berger, I.; Schaffitzel, C.*, et al.* SARS-CoV-2 spike variants differ in their allosteric response to linoleic acid. *J. Mol. Cell. Biol.* **2023**, mjad021.

45. Gupta, K.; Toelzer, C.; Williamson, M. K.; Shoemark, D. K.; Oliveira, A. S. F.*, et al.* Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2. *Nat. Commun.* **2022,** *13*, 222.
46. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3*, 1767.
47. Cunningham, B. C.; Wells, J. A. High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science* **1989,** *244*, 1081.
48. Wells, J. A., [18] Systematic mutational analyses of protein-protein interfaces. In *Methods in Enzymology*, Academic Press: 1991; Vol. 202, pp 390.
49. Morrison, K. L.; Weiss, G. A. Combinatorial alanine-scanning. *Curr. Opin. Chem. Biol.* **2001,** *5*, 302.
50. Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999,** *121*, 8133.
51. Kortemme, T.; Kim, D. E.; Baker, D. Computational Alanine Scanning of Protein-Protein Interfaces. *Science's STKE* **2004,** *2004*, pl2.
52. Wood, C. W.; Ibarra, A. A.; Bartlett, G. J.; Wilson, A. J.; Woolfson, D. N.*, et al.* BAlaS: fast, interactive and accessible computational alanine-scanning using BudeAlaScan. *Bioinformatics* **2020,** *36*, 2917.
53. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.*, et al.* MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314.
54. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Inhibition Mechanism of SARS-CoV-2 Main Protease with Ketone-Based Inhibitors Unveiled by Multiscale Simulations: Insights for Improved Designs. *Angew. Chem. Int. Ed.* **2021,** *60*, 25933.
55. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
56. Law, S. M. modevectors.py. https://pymolwiki.org/index.php/Modevectors (accessed 2022-06-30).
57. Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.*, et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020,** *11*, 3202.
58. Paasche, A.; Zipper, A.; Schafer, S.; Ziebuhr, J.; Schirmeister, T.*, et al.* Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry* **2014,** *53*, 5930.
59. Pavlova, A.; Lynch, D. L.; Daidone, I.; Zanetti-Polzi, L.; Smith, M. D.*, et al.* Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *Chem. Sci.* **2021,** *12*, 1513.
60. Muramatsu, T.; Takemoto, C.; Kim, Y.-T.; Wang, H.; Nishii, W.*, et al.* SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113*, 12997.
61. Henes, M.; Lockbaum, G. J.; Kosovrasti, K.; Leidner, F.; Nachum, G. S.*, et al.* Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance. *ACS Chem. Biol.* **2019,** *14*, 2441.
62. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
63. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
64. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
65. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
66. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126*, 014101.
67. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52*, 7182.
68. Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50*, 1055.
69. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089.

70. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.*, et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.
71. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.*, et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015,** *43*, D364.
72. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577.

5.3 Statement of authorship

Statement of Authorship for joint/multi-authored papers for PGR thesis
To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Substrate specificity and recognition in each SARS-CoV-2 main protease subsite informed by dynamical nonequilibrium molecular dynamics simulations
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Manuscript in preparation.

Student Confirmation

Student Name:	Ho Ting Henry CHAN		
Contribution to the Paper	Entire work.		
Signature		Date	1 August 2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Fernanda Duarte		
Supervisor comments I confirm the description above regarding authorship and contributions is accurate.		
Signature 	Date	2 August 2023

This completed form should be included in the thesis, at the end of the relevant chapter.

6. PL^{pro} Selectivity on Polyprotein Substrates

6.1 Overview

Chapters 3-5 describe computational modelling studies that enabled the prediction of M^{pro}-substrate complexation modes at equilibrium and provided insight into M^{pro} protein dynamics involved in substrate accommodation. SARS-CoV-2 PL^{pro} is another cysteine protease that is an attractive drug target (**Chapter 1**). While crystal structures of inhibitor-bound PL^{pro} have informed on binding in the S4-S1 subsites,[1] how substrates bind outside this core region remained unknown. An SPE-MS based assay developed by Brewitz *et al.* (2022)[2] revealed significant differences in the PL^{pro} hydrolytic efficiency towards oligopeptides representative of its native polyprotein substrates, despite similarity in their P4-P1 sequences, thus suggesting the potential importance of recognition outside of the P4-P1 binding regions. Using similar computational approaches as used for M^{pro}, this Chapter describes the modelling of PL^{pro}-substrate complexes for understanding interactions and PL^{pro} selectivity. The hypotheses proposed based on these models were tested by SPE-MS in this collaborative work.

6.2 Manuscript IV

This manuscript (**Manuscript IV**) has been deposited as a bioRxiv preprint (2023.07.11.548309) and submitted to RSC Chemical Biology; it is under review at the time of writing.

IV. Chan, H. T. H.; Brewitz, L.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A.; Schofield, C. J.*; Duarte, F.* Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2' proline of the viral polyprotein. *bioRxiv* **2023**, DOI: 10.1101/2023.07.11.548309. Submitted to *RSC Chem. Biol.*

I carried out all computational work and wrote the manuscript, with experimental data and the associated sections contributed by co-authors. A Statement of Authorship is appended following the manuscript. The Supplementary Information is in the **Appendix for Chapter 6**.

Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2′ proline of the viral polyprotein

Abstract

The SARS-CoV-2 papain-like protease (PL^{pro}) is an antiviral drug target that catalyses the hydrolysis of the viral polyproteins pp1a/1ab, releasing the non-structural proteins (nsps) 1-3 that are essential for the coronavirus lifecycle. The LXGG↓X motif found in pp1a/1ab is crucial for recognition and cleavage by PL^{pro}. We describe molecular dynamics, docking, and quantum mechanics/molecular mechanics (QM/MM) calculations to investigate how oligopeptide substrates derived from the viral polyprotein bind to PL^{pro}. The results reveal how the substrate sequence affects the efficiency of PL^{pro}-catalysed hydrolysis. In particular, a proline at the P2′ position promotes catalysis, as validated by residue substitutions and mass spectrometry-based analyses. Analysis of PL^{pro} catalysed hydrolysis of LXGG motif-containing oligopeptides derived from human proteins suggests that factors beyond the LXGG motif and the presence of a proline residue at P2′ contribute to catalytic efficiency, possibly reflecting the promiscuity of PL^{pro}. The results will help in identifying PL^{pro} substrates and guiding inhibitor design.

Keywords

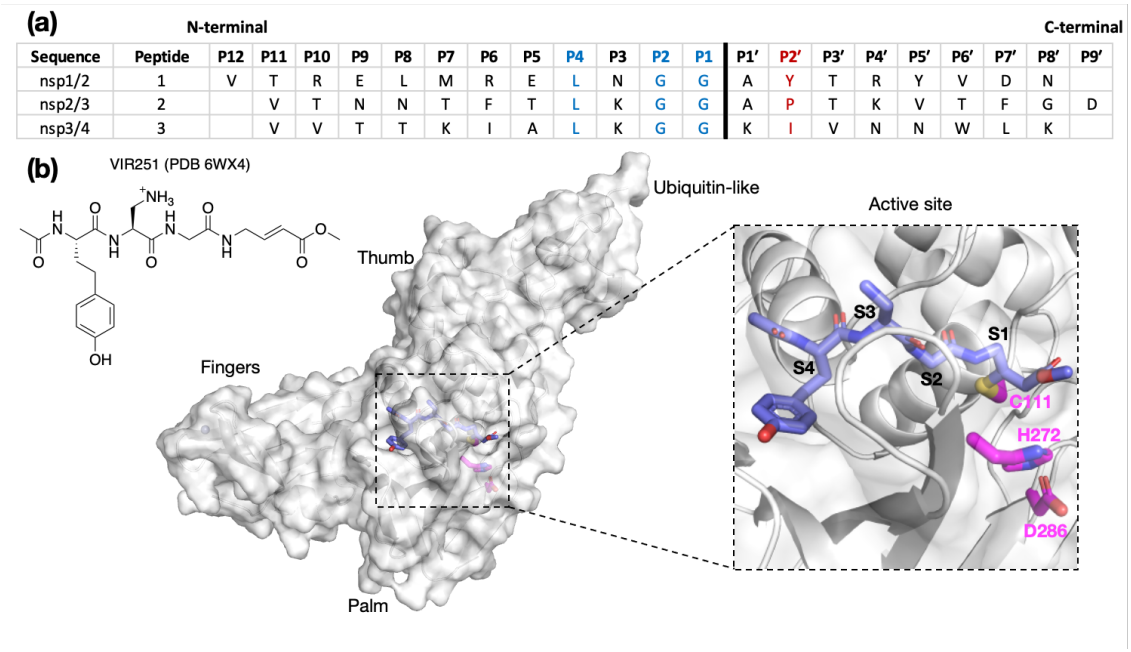
SARS-CoV-2, coronavirus, papain-like protease PL^{pro}, main protease M^{pro}, computational modelling of enzymes, substrate selectivity, nucleophilic cysteine protease

Introduction

The SARS-CoV-2 polyproteins ORF1a and ORF1ab are precursors of the 16 non-structural proteins (nsp1-16) that are essential for virus maturation and replication.[3] SARS-CoV-2 employs two nucleophilic cysteine proteases to release the nsps: the nsp5 main protease (M^{pro}) and the papain-like protease (PL^{pro}), the latter of which is a domain within nsp3.[4] Inhibition of M^{pro} and/or PL^{pro} hinders nsp release, leading to the stalling of viral maturation and replication.[5] The M^{pro} inhibitors PF-07321332 (nirmatrelvir), the active pharmaceutical ingredient (API) in Paxlovid, and ensitrelvir, the API in Xocova, are approved for treating COVID-19 infections.[6-8] However, clinical use of PL^{pro} inhibitors has not yet been approved.[5, 9, 10]

Although M^{pro} and PL^{pro} are both nucleophilic cysteine proteases, they differ in their structures, catalytic mechanisms, and substrate selectivities.[9] M^{pro} is predominantly homodimeric and employs an active site Cys-His dyad to catalyse hydrolysis of the SARS-CoV-2 polyproteins at 11 sites, which have [P4:A/V/P/T]-[P3:X]-[P2:L/F/V]-[P1:Q]↓[P1′:S/A/N] motifs, where "↓" denotes the scissile amide bond and X corresponds to any proteinogenic amino acid.[5, 11, 12] By contrast, PL^{pro} is likely monomeric and employs an active site Cys-His-Asp triad to catalyse the hydrolysis of the polyprotein at the nsp1/2, nsp2/3, and nsp3/4 sites (**Figure 6.1**).[13] PL^{pro} catalyses the hydrolysis of the polyprotein C-terminal to [P4:L]-[P3:X]-[P2:G]-[P1:G] sequences; the requirements for the P′ positions of PL^{pro} substrates have not yet been identified ([P1′:X]).[1, 14-16]

The apparent lack of selectivity for residues binding to the S′ subsites of PL^{pro} may contribute to its reported promiscuity. PL^{pro} catalyses the hydrolysis of peptide bonds C-terminal to LXGG motifs in various host proteins,[16, 17] including human interferon regulatory factor 3 (IRF3),[3] protein S (PROS1),[17] and the serine/threonine unc-51-like kinase (ULK1).[18] PL^{pro} also hydrolyses isopeptide bonds C-terminal to LXGG motifs in proteins that have been post-translationally modified by conjugation of ubiquitin or ubiquitin-like modifiers.[3, 19-22]



Sequence	Peptide	P12	P11	P10	P9	P8	P7	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′	P7′	P8′	P9′
nsp1/2	1	V	T	R	E	L	M	R	E	L	N	G	G	A	Y	T	R	Y	V	D	N	
nsp2/3	2		V	T	N	N	T	F	T	L	K	G	G	A	P	T	K	V	T	F	G	D
nsp3/4	3		V	V	T	T	K	I	A	L	K	G	G	K	I	V	N	N	W	L	K	

Figure 6.1: (a) Sequences of the reported oligopeptides used in SARS-CoV-2 PL^{pro} mass spectrometry-based assays.[2] The consensus LXGG sequence residues are in blue; the P2′ residues are in red. **(b)** View from a reported PL^{pro} structure (PDB 6WX4)[1] in complex with the covalently bound inhibitor VIR251 (carbon atoms in violet), which binds to the S1-S4 subsites; the catalytic triad residues are shown (carbon atoms in magenta). Non-carbon atoms are coloured: N in blue; O in red; S in yellow; Zn in grey. Hydrogens are omitted for clarity.

Assays using isolated recombinant PL^{pro} have been developed for activity and inhibition studies. Most of these are fluorescence-based and monitor the release of a C-terminal linked fluorescent group.[13, 21-28] We have previously reported mass spectrometry (MS)-based assays for PL^{pro} and M^{pro} utilising solid phase extraction coupled to MS (SPE-MS).[2, 29] The PL^{pro} SPE-MS assay employs a 20-mer peptide, containing the nsp2/3 cleavage site (V808-G818/A819-D827; **Figure 6.1a**) as a substrate. Surprisingly, little to no hydrolysis was observed for peptides corresponding to the nsp1/2 (V169-G180/A181-N188) and nsp3/4 (V2753-G2763/K2764-K2771) cleavage sites. Similarly, cell-based assays employing oligopeptide linkers with P5-P3′ cleavage site sequences, and liquid chromatography (LC)-based assays employing shorter polyprotein fragments, indicate relatively slow PL^{pro} catalysed hydrolysis at the nsp1/2 and nsp3/4 sites.[30, 31] Preferential hydrolysis of the nsp2/3 site has also been observed for SARS-CoV PL^{pro} using three analogous SARS-CoV polyprotein-derived peptide substrates as measured by LCMS assays.[32]

Collectively, these results indicate that the presence of the consensus LXGG P4-P1 residue motif is an insufficient criterion for, at least efficient, PL^{pro} catalysis.

Comparison of the P4 to P1′ residues of the three SARS-CoV-2 substrates shows that P3 is the first non-conserved substrate position N-terminal to the scissile amide bond (**Figure 6.1a**). nsp1/2 has an Asn at P3. While both nsp2/3 and nsp3/4 have a Lys residue at P3, only the former exhibited considerable activity. C-Terminal to the scissile amide, both nsp1/2 and 2/3 have an Ala at P1′, but hydrolysis of nsp1/2 was not observed by SPE-MS. Therefore, the observed difference in activities is likely a result of differences in binding affinity rather than catalytic mechanism.

The binding modes of oligopeptide substrates or peptidomimetic inhibitors in the S4-S1 subsites of PL^{pro} have been characterised crystallographically (*e.g.*, **Figure 6.1b**).[1, 14, 19, 23, 33] However, unlike M^{pro}, where structures with polyprotein-derived oligopeptides provide insights into S and S′ subsite binding,[34, 35] there is limited information on binding to the S′ subsites in PL^{pro}. Such information may help understanding of the selectivity of PL^{pro} catalysis and aid in inhibitor design. We thus constructed and assessed computational models of SARS-CoV-2 PL^{pro} complexed with its nsp1/2, nsp2/3, and nsp3/4 oligopeptide substrates. Our models presented here suggest that the P2′ proline in the nsp2/3 sequence plays a key role in efficient PL^{pro} catalysis, a proposal validated by residue substitutions and SPE-MS analyses (*vide infra*). The difference in hydrolytic activities between the tested oligopeptides and full-length (poly)proteins suggests that the overall polypeptide conformation is important for substrate recognition. This finding is relevant for future mechanistic and inhibition studies.

Results and Discussion

Predicted conformations of the PL^{pro}-bound nsp1-3 peptides

To investigate how the SARS-CoV-2 polyprotein nsp1/2, 2/3, and 3/4 cleavage sites bind to PL^{pro}, AutoDock CrankPep (ADCP)[36] was used to dock the peptides previously used in the SPE-MS assay,[2] *i.e.* **1**, **2**, and **3**, with PL^{pro}, using a crystal structure of PL^{pro} covalently linked with the peptidomimetic inhibitor VIR251 (PDB 6WX4, 1.66 Å resolution; see **Methods** for details).[1] The top 100 ranked peptide poses were analysed and denoted by "dX_YY", where "X"=1-3 represents the peptide number according to **Figure 6.1a** and "YY" indicates the ranking, with dX_01 corresponding to the pose predicted to have the highest binding affinity. In all the highest ranked solutions (d1_01, d2_01, and d3_01), the S2-S1 channel was observed to be occupied (**Supplementary Figure S6.1** in **Appendix for Chapter 6**), suggesting it may be a binding hotspot. However, only with d2_01 was the S2-S1 channel observed to be occupied by the P2-P1 Gly-Gly residues in an apparently catalytically relevant manner. Thus, consideration of only the highest ranked solution was considered to be insufficient for generating catalytically relevant protein-substrate complexes.

Further analysis of the top 100 poses involved filtering based on the position of the P4-P1 residues relative to the binding mode of the inhibitor, VIR251.[1] A pose was deemed to pass the filter if all four Cα atoms of P4-P1 were within 2 Å of the corresponding VIR251 Cα atoms (**Table S6.1**). For peptide **1**, only one docked solution passed the filter (d1_16). For peptide **2**, 10 solutions fulfilled the criteria, including the top ranked pose (d2_01, 02, 06, 09, 38, 62, 66, 69, 75, 97); these solutions had a similar conformation to that of d2_01, except for d2_09, which adopts an extended conformation (**Figure S6.2**). Thus, both d2_01 and d2_09 were selected as representative modes for further modelling. Finally, none of the 100 poses passed the filter for peptide **3**. However, closer inspection of these results revealed multiple solutions in which the P2-P1 Gly-Gly residues were placed in the S2-S1 channel, but in the reverse direction, with P2-

Gly in S1 and the P1-Gly in S2. In these conformations, the P1′-Lys was in S3 and the P2′-Ile in S4 (*e.g.*, d3_07; see **Figures 6.2 and S6.3**). Although no docked solutions passed the filter for **3**, it does not rule out the possibility of **3** forming a pre-reaction complex with PL^{pro}. A model of such a pre-reaction complex was later constructed based on one of the poses that passed the filter for peptides **1** and **2** (d1_16, d2_01, d2_09) (**Figure 6.2**).

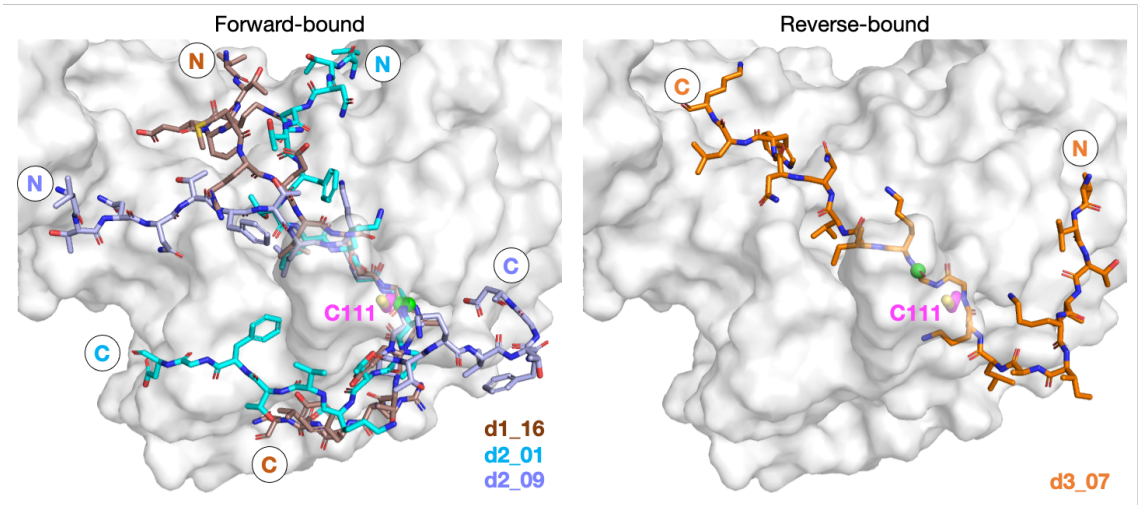


Figure 6.2: Predicted bound conformations of linear oligopeptides in complex with PL^{pro}. AutoDock CrankPep (ADCP)[36]-predicted peptide conformations when non-covalently bound with PL^{pro} (white surface), with PL^{pro} Cys111 residue (magenta sticks) and the peptide N- and C-terminals labelled. The P1 scissile amide carbonyl carbon is shown as a green sphere. The poses are denoted by "dX_YY", where "X" is the peptide number **1**, **2**, or **3** for the nsp1/2, 2/3, or 3/4 cleavage sites respectively, and "YY" the ranking out of 100 docked poses.

To assess the relative stability of the selected representative peptide poses for **1** and **2**, namely d1_16, d2_01, and d2_09, 3 × 100 ns molecular dynamics (MD) simulations were performed for each system using the GROMACS package with the AMBER99SB-ILDN forcefield and explicit (TIP3P) water solvation,[37-39] as reported in our previous study on M^{pro}-peptide complexes (see **Methods** for details).[40] Since protonation states are fixed in conventional MD, both the neutral ("N") and zwitterionic ("Z") states of the active site Cys111-His272 pair were simulated.

Simulating peptides **1** and **2** in either N- or Z-states of the Cys111-His272 dyad revealed significant deviation in the peptide backbone root mean square deviation (RMSD) and root mean

square fluctuation (RMSF) at the N- and C-terminal regions. Notably, the core P5-P1′ region around the consensus LXGG sequence exhibited a more rigid structure (**Figures S6.4-5**). Analysis of non-hydrogen atom RMSF (**Figure S6.6**) indicates stable binding of the P4-Leu residue sidechain compared to its neighbour residues, reflecting its conserved nature in PL^{pro} substrates (**Figure 6.1**). Among the three poses studied, d2_01 was the most stable (**Figures 6.2 and S6.5**), with relative rigidity of the N-terminal region in the Z-state and of the C-terminal region in the N-state. The d2_01 pose prior to MD was used for comparative modelling of **1** and **3** to explore if all three peptides could bind similarly. However, these MD simulations did not inform on whether the N- or Z-state was preferred in PL^{pro}-substrate complexes.

QM/MM-US calculations indicate a preference for the zwitterionic Cys-His pair

To further explore the preferred ionic state of the Cys111-His272-Asp286 catalytic triad, quantum mechanics/molecular mechanics (QM/MM) umbrella sampling (US) calculations were performed using the Amber18 and Gaussian16 packages (see **Methods**).[41, 42] Given that the thermodynamic preference may be affected by the presence of substrate, as reported for other cysteine proteases,[43, 44] both *apo* PL^{pro} and PL^{pro} non-covalently complexed with **2** were considered (**Figures S6.7-11**). In both cases, the results reveal a preference for the Z-state of Cys111-His272 by >5.0 kcal/mol (**Figure S6.10**). Subsequent proton transfer from His272 to Asp286 leads to a state that is >1.8 kcal/mol higher in energy than when His272 is positively charged and Asp286 is deprotonated (**Figure S6.11**). Hence, the most thermodynamically favourable state for PL^{pro} in both the *apo* and substrate-bound cases, at least with peptide **2** in the d2_01 conformation, involves a negatively charged Cys111, a positively charged His272, and a negatively charged Asp286. The thermally accessible activation barriers of the proton transfer steps, which range between 2.6 and 11.9 kcal/mol relative to the Z-state, suggest fast interconversion between N- and Z-states under solution conditions (**Figures S6.10-11**). These findings align with reported QM/MM calculations for the papain-like protease cathepsin B[45] and from constant pH MD simulations of *apo* coronavirus PL^{pro} enzymes.[46] However, they differ from a QM/MM study of SARS-CoV-2 PL^{pro}

complexed with VIR251,[47] which suggests that the N-state is thermodynamically favoured over the Z-state. Differences in the substrate and the inclusion of Asp286 in our QM region may account for this apparent discrepancy. The preference of the catalytic triad for different ionic states likely depends on the surrounding environment, which is affected by the conformation of the bound peptide and the interactions it establishes with the protein residues, especially those involving nearby peptide P′ residues immediately C-terminal to the scissile amide. While subsequent discussions primarily focus on simulations conducted in the more favoured Z-state, the N-state is discussed where necessary, particularly in relation to the P′ interactions.

Hydrogen bond and dispersion interactions are present both in the core P5-P1′ region and around the peptide N- and C-termini

To investigate if peptides **1** and **3** can bind stably in the d2_01 conformation of **2**, models were constructed (**Figure S6.12**) considering both the Z- and N-states. For each setup, 3 × 200 ns MD simulations were performed (**Figures S6.13-15**). Over the course of MD simulations, the P5-P1′ residues remain stable as indicated by RMSF trends (**Figures S6.14-15**). Binding is conferred by a series of hydrogen bond (HB) interactions (**Figure 6.3a**; **Table S6.2**). However, these interactions are more poorly maintained around P5-P3 for **3** compared to **1** and **2**. Towards the N-terminus, peptide **3** shows higher flexibility than **1** and **2**, where residues are held more rigidly by interactions with PL^{pro} residues located between the thumb and fingers domains (**Figure S6.16**).

Towards the C-terminus of all three peptides, the P′ residues appear to be more flexible in the Z-state compared to the N-state. The backbone RMSF increases with distance from the active site and reaches values >5 Å after P3′-P5′, except for **2** in the N-state where the RMSF is >5 Å only at P8′ (**Figure S6.14**). Clustering analysis was performed in the N-state to visualise interactions which might confer stability, using a 3 Å RMSD cut-off on the peptide backbone (**Figure 6.3b**). For **2** in the N-state, the three most populated clusters (27%, 16%, 8%) have the P′ residues adopting an extended conformation adjacent to the PL^{pro} palm domain (**Figure 6.3c**). These

clusters reveal the presence of four significant HBs, including HBs with Gln250, Thr265, and Gly266, and dispersion interactions with Trp106, Asn109, Cys111, Leu162, Pro248-Gln250, Tyr264-Tyr268, Cys270-His272, and Pro299 (**Figure 6.3b-c; Table S6.2**). Notably, the P2′ proline residue in **2** emerges as a crucial residue for inducing and stabilising a bend centred at the P1′ residue adjacent to the P2′ proline (**Figure S6.18**). The P2′ proline thus influences the overall peptide conformation and facilitates binding of residues in the P′ region (**Figures 6.3c and S6.19-20**). In **1** and **3**, such the bend centred at P1′ is not as well preserved, and residues C-terminal of P2′ explore wider conformational space in terms of their backbone ϕ and ψ dihedrals compared to **2** (**Figures S6.18-20**).

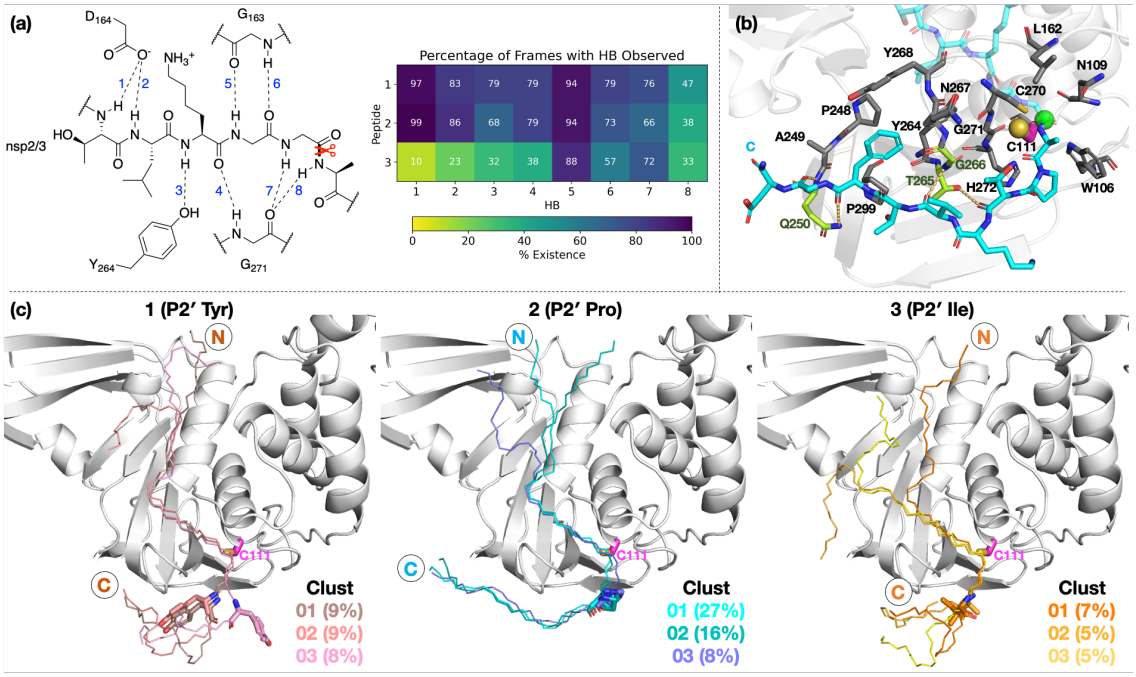


Figure 6.3: Interactions of PL^{pro} with the oligopeptides 1, 2, and 3. (a) Conserved HBs (occurrence ≥ 25% for at least two peptides) in the P5-P1′ region, observed in 3 × 200 ns MD. (b) Representative structure of N-state PL^{pro} complexed with **2** (cyan; P1 scissile amide carbonyl carbon as a green sphere) from cluster analysis (peptide backbone RMSD cut-off of 3 Å) highlighting interactions in the P′ region. PL^{pro} residues involved in HBs are shown in lime. Residues within 4 Å of P′ residues are in grey, except for Cys111 which is in magenta (its sulfur as a yellow sphere). (c) View of representative structures of the three most populated clusters, from the combined 3 × 200 ns MD of each of the three PL^{pro}-peptide (N state) complexes. The peptide backbone is shown as thin sticks and the P2′ residue as thick sticks.

A reverse non-reactive binding pose is possible

Our initial docking studies suggested a possible reverse binding mode for **3**, wherein the P2-Gly is positioned in the PL^{pro} S1 subsite and the P1-Gly is positioned in the S2 subsite (**Figures 6.2 and S6.3**). To evaluate this binding pose for **3** and investigate if such a reverse binding conformation is possible for **1** and **2**, the d3_07 pose was subjected to initial MD simulations (**Figures S6.21-22**). A stable peptide conformation was obtained from these simulations (**Figures S6.23-26**) and was used to analogously generate models with **1** and **2** (**Figure S6.27**). Each system was subjected to 3 × 200 ns MD, with Cys111-His272 in the Z-state.

All of **1-3** bind stably in the reverse mode in the core region of PL^{pro} as indicated by peptide RMSF values (**Figures S6.28-30**). Similar to the forward-bound peptides (**Figure 6.3a**), several HBs are observed between the peptide and PL^{pro} residues, notably, Gly271 and Gly163 in the S3-S1 subsites (**Figure 6.4**; **Table S6.3**). However, considering the distance between His272 and the leaving group scissile amide derived nitrogen, it is unlikely that the amide bond at S1 is in a catalytically productive position (**Figure S6.31**). In the hydrophobic S4 pocket, the original P2′ residues (Tyr, Pro, Ile) of **1**, **2**, and **3**, respectively, are stable according to RMSF analysis (**Figure S6.30**). Outside of the core S4-S1 region, the peptides manifest high flexibility, except for **3** where the N-terminal residues are stably bound in the PL^{pro} S′ region. Closer inspection reveals that the N-terminal residues of **3** are positioned to form well-maintained HBs and dispersion interactions with PL^{pro} adjacent to the PL^{pro} palm domain (**Figure S6.32; Table S6.3**). The simulations suggest the possibility of a reverse, non-productive binding mode for the nsp peptides **1-3.** Note, however, that the biological relevance of this observation for protein substrates is unclear and **3** does not inhibit the PL^{pro}-catalysed hydrolysis of **2**.[2] Previous studies have shown that non-productive binding of peptides, including D-enantiomers of substrates, with protease substrate binding clefts are possible.[48] This is perhaps unsurprising given the nature of the interactions between proteases and their substrates, which often involve multiple hydrogen bonds forming β-sheet type

structures. In an *in vivo* context, most proteases are likely to encounter several potential substrates, including themselves; hence further work on understanding how selectivity is achieved via non-productive binding, as we observe with PL^{pro}, is of interest.

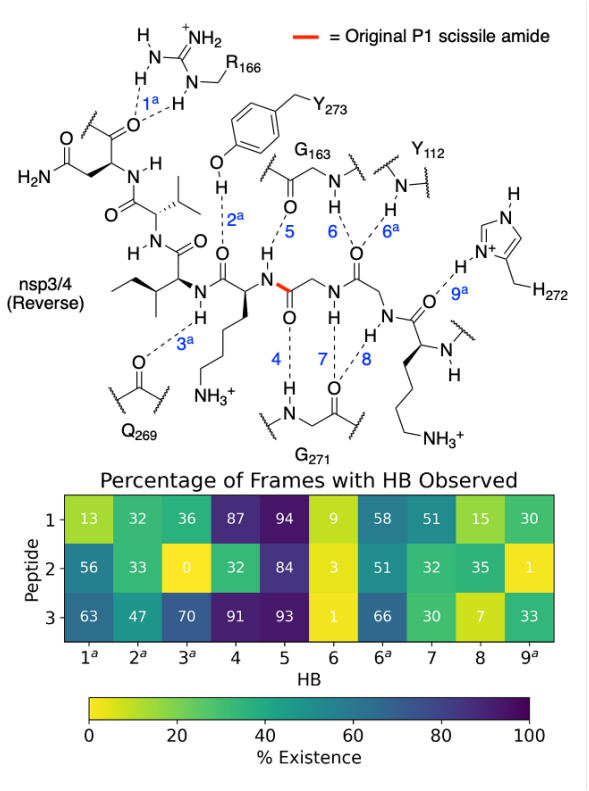


Figure 6.4: Possibility of reverse binding conformations for the oligopeptides 1, 2, and 3. Occurrence of conserved HBs during 3 × 200 ns MD of PL^{pro} complexed with the three nsp oligopeptides **1**, **2**, and **3** in the reverse binding conformations. The HBs are numbered according to the equivalent peptide position in **Figure 6.3a**, with additional HBs carrying the "a" superscript.

Dynamics of the loop around Tyr268 distinguishes *apo* and substrate-bound PL^{pro}

The dynamics of the PL^{pro} loop around Tyr268 (Gly266-Gly271), also known as the blocking loop 2 (BL2) and which is reported to regulate active site access, [14, 22, 23, 46, 49, 50] was analysed for the *apo* and PL^{pro} complexes with forward- and reverse-bound peptides in each set of combined 3 × 200 ns MD. This analysis considers the observed proximity and interactions of BL2 residues with forward-bound peptide P′ residues (**Figure 6.3b**) and reverse-bound peptide P residues (**Figure**

S6.32). The opening and closing states of BL2 were quantified using the backbone RMSF values of BL2 residues and the Pro248–Tyr268 Cα–Cα distance (**Figure 6.5a**).

Backbone RMSF analysis indicates that BL2 is more flexible in the *apo* state compared to the peptide-bound state (**Figure S6.33-34**). In the *apo* PL^{pro}, the RMSF of the most flexible residue in BL2, Tyr268, is 2.5 ± 0.1 Å, whereas in the substrate peptide **2**-bound PL^{pro} it is 0.98 ± 0.03 Å (**Figure S6.35**). For the *apo* system, multiple BL2 opening and closing events were observed to occur, an observation which may be related to substrate capture processes (**Figures 6.5 and S6.36**).

In the forward-bound peptide-PL^{pro} complexes, BL2 becomes more rigid compared to the *apo*-form, although BL2 opening still occurred when bound to peptide **3**, and to a lesser extent when bound to peptide **1** (**Figures 6.5b and S6.36**). Such opening events with **1** or **3** bound may reflect the tendency of PL^{pro} to revert to its *apo* form, indicating binding instability and a propensity towards peptide dissociation. By contrast, complexation with peptide **2** results in a rigidly closed BL2 throughout the simulations, possibly as a result of the stronger interactions between the P′ residues of **2** and PL^{pro} residues close to BL2 (**Figure 6.3b**).

The reverse-bound complexes show the opposite trend to the forward-bound complexes (**Figures 6.5c and S6.34-36**). BL2 remains rigidly closed over the course of MD with peptides **1** and **3**, but opens multiple times with **2**. Previous studies have reported that ligand occupation at the S4 and S3 subsites induces BL2 closure.[21, 23, 24, 49] However, here both **1** and **2** have a P1′ Ala residue which binds in S3, and yet BL2 is more flexible when bound with **2** than with **1**. The difference in BL2 flexibility probably arises from the different residues binding in the S4 pocket. In **1** and **3**, the S4 pocket is occupied by the hydrophobic sidechains of P2′ Tyr and Ile respectively, whereas it remains mostly vacant in **2** where the P2′ Pro lacks an extended hydrophobic sidechain (**Figure S6.37**).

The combined evidence from our modelling studies suggests that the more efficient reaction of PL^{pro} with peptide **2** compared to peptides **1** or **3** can be attributed to the presence of a Pro residue at the P2′ position. This results in a bound pre-reaction complex conformation where the P′ residues favourably interact with PL^{pro} and promote the closure of BL2, thus reducing the probability of peptide dissociation. Additionally, compared to **1** or **3**, peptide **2** shows a greater preference for the productive forward binding mode relative to the competing non-productive reverse binding mode.

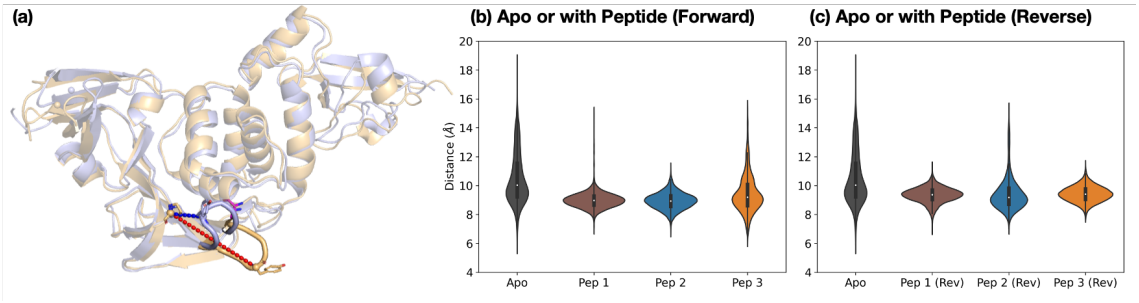


Figure 6.5: Dynamics of the Tyr268-containing loop PL^{pro} BL2. (a) Views of MD-observed structures of *apo* PL^{pro} when BL2 is closed (light blue) and open (light orange), which can be quantified by the Pro248–Tyr268 Cα–Cα distance (blue and red dotted lines respectively). The distribution of this distance over combined 3 × 200 ns MD is plotted for PL^{pro} in the *apo* state and (b) when complexed with the nsp peptides in the forward binding mode, and (c) when complexed with the nsp peptides in the reverse binding mode.

A proline residue at the substrate P2′ position enhances PL^{pro}-catalysed hydrolysis of peptides based on the viral polyprotein

We next investigated the proposed role of the P2′ proline residue of the nsp2/3 (**2**) oligopeptide for SARS-CoV-2 PL^{pro} catalysis. Using solid phase peptide synthesis (SPPS), we synthesised nsp1/2 (**1**)- and nsp3/4 (**3**)-based peptides where the wildtype (wt) P2′ residue of the Wuhan-Hu-1 strain of SARS-CoV-2[51] was substituted for a proline, as in the wt nsp2/3 peptide sequence (**2**), *i.e.* the nsp1/2(Y182P) (**4**) and nsp3/4(I2765P) (**5**) variant peptides (**Figure 6.6a**).

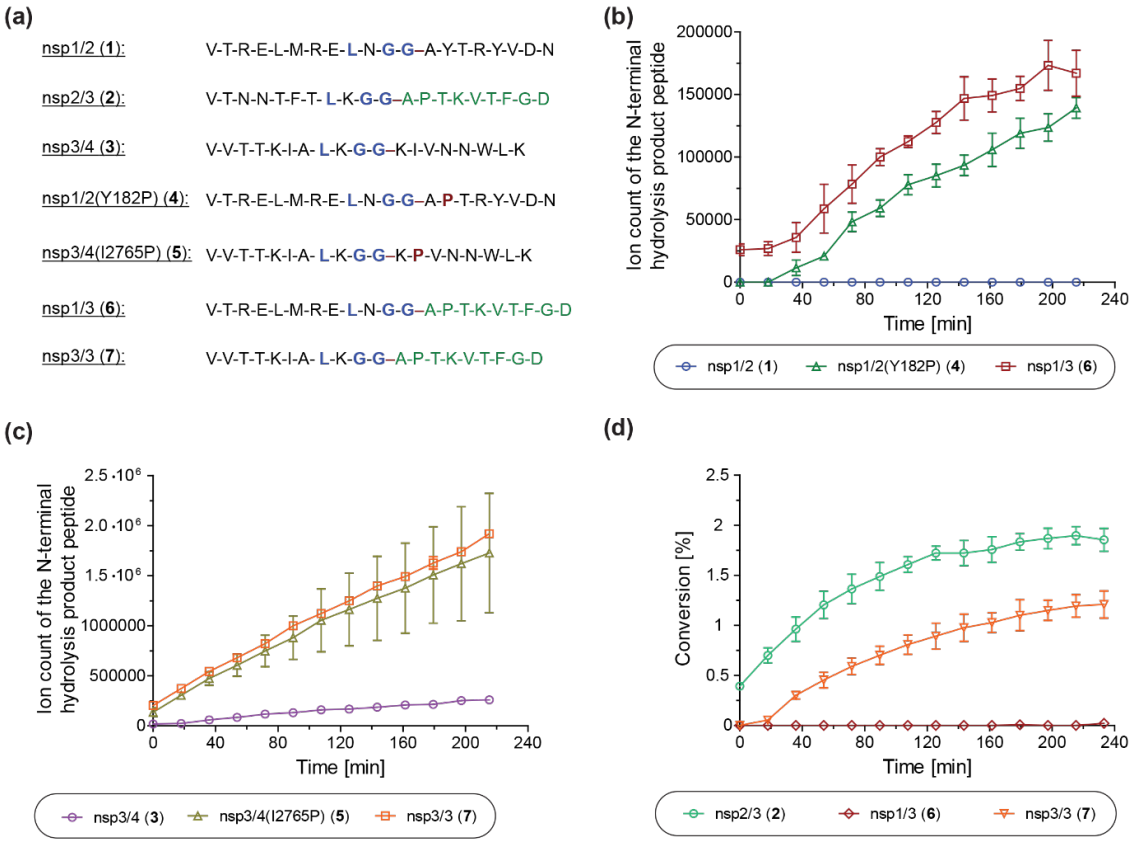


Figure 6.6: A proline at the substrate P2′ position enhances PL^{pro} catalysis. (**a**) Sequences of the employed SARS-CoV-2 polyprotein-derived oligopeptides used; (**b**) ion counts of the N-terminal hydrolysis product peptides of wt nsp1/2 (**1**; blue circles), nsp1/2(Y182P) (**4**; green triangles), and nsp1/3 (**6**; red boxes) incubated with SARS-CoV-2 PL^{pro}; (**c**) ion counts of the N-terminal hydrolysis product peptides of wt nsp3/4 (**3**; lavender circles), nsp3/4(I2765P) (**5**; ochre triangles), and nsp3/3 (**7**; orange boxes) incubated with SARS-CoV-2 PL^{pro}; (**d**) conversion of the PL^{pro}-catalysed hydrolysis of wt nsp2/3 (**2**; light green circles), nsp1/3 (**6**; red diamonds), and nsp3/3 (**7**; orange inverse triangles), determined using the reported N-terminal acetylated C-terminal hydrolysis product peptide of **2**[2] as an internal standard (0.2 µM). Reactions were performed using peptide (2.0 µM) and SARS-CoV-2 PL^{pro} (0.2 µM) in buffer (50 mM Tris, pH 8.0, 20 ºC) and analysed using SPE-MS under reported conditions.[2] Results are means of three independent runs ($N = 3$; mean ± standard deviation, SD).

The nsp1/2(Y182P) (**4**) and nsp3/4(I2765P) (**5**) variant peptides were incubated with isolated recombinant SARS-CoV-2 PL^{pro} under reported conditions (50 mM Tris, pH 8.0, 20 ºC)[2] and peptide hydrolysis was monitored by SPE-MS. The results reveal that PL^{pro}-catalysed hydrolysis of **4** and **5** is substantially more efficient compared to wt nsp1/2 and nsp3/4 peptides **1** and **3**, respectively, which were used as controls. PL^{pro} catalysed the hydrolysis of **4**, albeit at low levels, whereas it did not catalyse the hydrolysis of the wt nsp1/2 peptide (**1**) under the tested conditions (**Figure 6.6b**). The PL^{pro}-catalysed hydrolysis of **5** was ~10-fold more efficient than that of wt **3**

(**Figure 6.6c**). Note that the ion counts of the N-terminal product peptides were used for comparison of the hydrolysis efficiencies of the wt and variant peptides.

To investigate whether residues C-terminal to the LXGG motif of the nsp2/3 peptide, other than the P2′ proline residue, contribute to enhanced PL^{pro} catalysis, chimeric oligopeptides of nsp1/2 (**1**) and nsp3/4 (**3**) containing the entire fragment C-terminal to the LXGG motif of the nsp2/3 peptide (**2**) were synthesised, *i.e.* the nsp1/3 (V169-G180/A819-D827) (**6**) and nsp3/3 (V2753-G2763/A819-D827) (**7**) hybrid peptides (**Figure 6.6a**). These chimeric peptides were incubated with isolated recombinant SARS-CoV-2 PL^{pro} and hydrolysis was monitored using SPE-MS.[2] The results revealed that the efficiency of the PL^{pro}-catalysed hydrolysis of the hybrid peptides **6** and **7** was similar, within experimental error, to those of the nsp1/2(Y182P) and nsp3/4(I2765P) variant peptides **4** and **5**, respectively, indicating that residues at the P′ positions other than the P2′ proline did not have pronounced effects on PL^{pro} catalysis, within the context of the tested peptides (**Figure 6.6b-c**).

Reactions with the nsp1/3 and nsp3/3 hybrid peptides **6** and **7** were performed in the presence of the reported N-terminal acetylated C-terminal hydrolysis product peptide of the wt nsp2/3 peptide (**2**),[2] which was used as an internal standard to quantify peptide hydrolysis. By comparing the PL^{pro}-catalysed hydrolysis of the wt nsp2/3 peptide (**2**) with peptides **6** and **7**, it became evident that **2** was a substantially more efficient substrate of PL^{pro} than both **6** and **7**, as well as, by implication, than the nsp1/2(Y182P) and nsp3/4(I2765P) variant peptides **4** and **5**. The PL^{pro}-catalysed hydrolysis of **2** was ~2-fold more efficient than that of **7**, while the hydrolysis levels of **6** were marginal (**Figure 6.6d**). This observation suggests that substrate residues which bind to the S subsites are, in general, more important for efficient PL^{pro} catalysis than those binding to the S′ subsites, in accord with previous studies on SARS-CoV and SARS-CoV-2 PL^{pro}, which have shown a preference for lysine over asparagine residues at the P3 position,[1, 31, 32, 52] as well as with other cysteine proteases, such as SARS-CoV-2 M^{pro}.[40] Note that the low absolute levels of

hydrolysis of the wt nsp2/3 peptide (**2**) may be a result of performing SPE-MS assays at 20 ºC rather than at 37 °C and, potentially, of using relatively short oligopeptides instead of fully folded polyproteins as substrates, which are likely unable to bind to PL^{pro} at allosteric positions; substrate binding to pockets distal to the PL^{pro} active site has been reported to be essential for efficient catalysis.[53]

The primary sequence of LXGG-motif containing human proteins affects their ability to be substrates of SARS-CoV-2 PL^{pro}

Human proteins containing an LXGG motif in their sequence have been identified as (potential) substrates of SARS-CoV-2 PL^{pro}. This conclusion is based on *in vitro* and cell-based studies,[16] including with interferon regulatory factor 3 (IRF3),[3] protein S (PROS1),[17] and serine/threonine-protein kinase ULK1.[18] To investigate these potential substrates further, we synthesised SPE-MS compatible peptides based on the sequences of IRF3 (**8**), PROS1 (**9**), ULK1 (**10**), and the ubiquitin-like modifier-activating enzyme ATG7 (ATG7, **11**). ATG7 was also investigated as a potential human PL^{pro} substrate, because, like ULK1, it is involved in autophagy induction and bears a LGGG motif; in addition, it has a Pro residue at the P2′ corresponding position.[54, 55]

The peptides were incubated with isolated recombinant SARS-CoV-2 PL^{pro} under standard conditions (50 mM Tris, pH 8.0), though at 37 °C, because of the low levels of PL^{pro} catalysed hydrolysis initially observed at 20 ºC. Analysis of the peptide mixtures after 12 h of incubation with PL^{pro} revealed that PL^{pro} catalyses hydrolysis of PROS1 (**9**) and ULK1 (**10**) more efficiently than IRF3 (**8**). However, PL^{pro} did not catalyse the hydrolysis of ATG7 (**11**) under the tested conditions (**Table 6.1**). PL^{pro}-catalysed hydrolysis of PROS1 (**9**) and ULK1 (**10**) was, however, substantially less efficient than for nsp2/3 (**2**) (**Table 6.1**, entry ii), though was more efficient than for nsp1/2 (**1**) and nsp3/4 (**3**). This suggests that PL^{pro} may catalyse the hydrolysis of certain human proteins more efficiently than of, at least some, of its natural viral substrates. Note that

conversions were estimated by comparing starting material depletion with no-enzyme controls; the identity of the N- and C-terminal product peptides was confirmed by MS.

Table 6.1: Comparison of the SARS-CoV-2 PL^{pro}-catalysed hydrolysis of LXGG-motif containing peptides derived from the primary sequence of viral and human proteins as estimated by SPE-MS analysis.

	Peptide	[a]**Sequence**	**Protein**	[b]**%-hydrolysis**
i	nsp1/2 (**1**)	VTRELMRELNGG/AYTRYVDN	SARS-CoV-2 polyprotein	not detected
ii	nsp2/3 (**2**)	VTNNTFTLKGG/APTKVTFGD	SARS-CoV-2 polyprotein	>90%
iii	nsp3/4 (**3**)	VVTTKIALKGG/KIVNNWLK	SARS-CoV-2 polyprotein	<5%
iv	IRF3 (**8**)	VRHVLSCLGGG/LALWRA	interferon regulatory factor 3 (IRF3)	<5%
v	PROS1 (**9**)	LLIALRGG/KIEVQLKN	protein S (PROS)	27 ± 4%
vi	ULK1 (**10**)	LARKMSLGGG/RPYTPSPQVG	serine/threonine-protein kinase ULK1	23 ± 4%
vii	ATG7 (**11**)	FEDCLGGG/KPKALAAAD	ubiquitin-like modifier-activating enzyme ATG7	not detected

a) '/' indicates the predicted site of SARS-CoV-2 PL^{pro} catalysed hydrolysis; the peptide identity was confirmed by SPE-MS; b) determined using SPE-MS by analysing substrate (2.0 μM) depletion after 12 h incubation with SARS-CoV-2 PL^{pro} (0.2 μM) at 37 °C in buffer (50 mM Tris, pH 8.0); product formation was confirmed by SPE-MS analysis using reported conditions.[2] Results are means of two independent runs (n = 2; mean ± SD).

Interestingly, the results indicate that a proline residue at the substrate P2′ position does not necessarily result in improved PL^{pro} catalysis. For example, the ULK1-derived peptide **10**, which has a P2′ proline residue, shows ~23% hydrolysis after 12 h, whereas hydrolysis products were not observed for ATG7 peptide (**11**), which also contains a proline residue at P2′ (**Table 6.1**, entries vi and vii). By contrast, the PL^{pro}-catalysed hydrolysis of PROS1 (**9**) proceeded with similar efficiency as that of ULK1 (**10**), *i.e.* ~27% hydrolysis after 12 h, though it does not bear a proline residue at P2′ (**Table 6.1**, entry v). These combined results suggest that factors other than the presence of an LXGG motif and a proline residue at P2′ determine the efficiency of PL^{pro}

catalysis, potentially including the protein/peptide fold. In addition, other, yet unidentified, interactions of primary sequence residues with PL^{pro} which bind near the active site may be of importance. For instance, it appears that efficient substrates bear an alcohol-containing side chain at P5 (Ser/Thr).

Predicted binding modes of human protein-derived peptides substrates to PL^{pro}

To investigate how peptides **8**, **9**, **10**, and **11**, derived respectively from IRF3, PROS1, ULK1, and ATG7, may bind to SARS-CoV-2 PL^{pro}, the same docking procedures used for the nsp peptides **1-3** were applied, with assessment of the poses relative to the Cα atoms of the peptidomimetic inhibitor VIR251 in complex with PL^{pro} as observed by crystallography (PDB 6WX4)[1] (**Tables S6.4-5**). Among these peptides, only PROS1 (**9**) returned conformations that passed the filter (**Figure S6.38**). Despite the lack of a P2′ Pro, the PROS1 P′ residues are predicted to adopt a similar conformation as d2_01 for **2**. In the PROS1 poses, the hydrophobic sidechains of P2′-Ile, P4′-Val, and P6′-Leu point towards BL2, while the hydrophilic P1′-Lys, P3′-Glu, P5′-Gln, and P7′-Lys sidechains point in the opposite direction and/or outwards into solvent. These interactions in the P′ region with PL^{pro} might contribute to the formation of a productive pre-reaction complex as investigated with **2**. For ULK1 (**10**), despite the observation of efficient hydrolysis and the presence of P2′-Pro, none of the 100 predicted conformations passed the filter. Multiple poses nearly passed the filter, placing the P3-P1 residues in the corresponding S3-S1 subsites but not the P4-Leu in S4, and the P′ residues interacting with PL^{pro} underneath BL2 (**Table S6.4** and **Figure S6.39**). For IRF3 (**8**), for which efficient hydrolysis was not observed, none of the 100 docked conformations passed the filter. Several of the IRF3 poses were observed to bind in reverse, with the P1′-Leu in S4, P1-Gly in S3, P2-Gly in S2, and P3-Gly in S1, which might compete with productive binding (**Figure S6.40**). For ATG7 (**11**), hydrolysis of which was not detected by SPE-MS, none of the 100 docked conformations passed the filter. The only poses binding within the S2-S1 channel place the consecutive P6′-P8′ Ala-Ala-Ala in the S2-S4 subsites

respectively, with the peptide bound in the reverse direction (**Figure S6.41**). Overall, these results suggest that whilst crystallography-based docking predictions can be helpful in suggesting possible substrate binding poses, they are unlikely to predict whether a peptide is hydrolysed efficiently by PL^{pro}, which is regulated by a complex mixture of factors. This complexity reflects the relatively high substrate promiscuity of PL^{pro}, including its ability to accept both peptide and isopeptide substrates.

Conclusions

Several studies have provided evidence that PL^{pro} catalysis of nsp2/3 cleavage is more efficient than that of nsp1/2 or nsp3/4, indicating that the presence of an LXGG motif alone is insufficient to predict catalysis efficiency.[2, 30-32, 56] The interaction of viral proteases with the (human) host proteome is well-documented,[57, 58] including for SARS-CoV-2 PL^{pro}, which is reported to possess deubiquitinase activity and catalyse the removal of (poly)ubiquitin and ubiquitin-like modifiers bearing a LXGG motif from post-translationally modified human proteins.[19, 22] However, factors that govern the ability of SARS-CoV-2 PL^{pro} to catalyse peptide bond hydrolysis C-terminal to LXGG motifs in human proteins are not well understood.

Our combined modelling and experimental results reveal that the primary sequence of potential substrate proteins apart from LXGG motifs affect the efficiency of PL^{pro}-catalysed substrate hydrolysis. Studies on the bound conformations and the substrate-PL^{pro} dynamics suggest that having a proline at the P2′ position can, at least in some cases, promote catalysis of isolated recombinant PL^{pro}. This proposal is validated by residue substitutions and *in vitro* turnover assays. The results suggest that the activity enhancement with a proline at the P2′ position is likely due to interactions between PL^{pro} and the P′ residues, which are enhanced by the P2′ proline induced

bend in peptide fold, and, possibly, unfavourable binding of a proline in the hydrophobic S4 pocket which may limit unreactive reverse direction binding modes in the active site.

Our results also suggest that whilst docking predictions can inform potentially productive binding modes, they do not confidently predict whether a potential sequence will be efficiently hydrolysed by PL^{pro}. Studies on the PL^{pro}-catalysed hydrolysis of LXGG motif-containing oligopeptides derived from human proteins*, i.e.*, IRF3, PROS1, ULK1, and ATG7, suggest that, in addition to the presence of an LXGG motif and a proline residue at P2′, other factors are important in regulating the efficiency of PL^{pro} catalysis; efficient PL^{pro}-catalysed hydrolysis can occur without a P2′ proline in the substrate.

Our data shows that primary sequences that may *per se* appear unfavourable for PL^{pro} catalysis, such as those lacking a proline at P2′, can still be efficient substrates provided that *e.g.* the protein fold favours binding to PL^{pro}, as is likely the case for the viral polyprotein nsp1/2 cleavage site. Conversely, primary sequences that *per se* appear to favour PL^{pro} catalysis might in fact not be efficient PL^{pro} substrates in cells. Thus, care needs to be taken when investigating human proteins for substrate activity with PL^{pro}; studies aimed at identifying novel human SARS-CoV-2 PL^{pro} substrates, and by extension substrates of other viral proteases, should involve work with folded proteins, ideally combined with proteomic studies of infected cells to correlate *in vitro* and *in silico* results to real biology.

The importance of non-covalent interactions at the P′ substrate regions in enhancing PL^{pro} proteolytic activity likely extends well beyond the consensus LXGG motif, as demonstrated with other cysteine proteases such as hepacivirus NS2-3 proteases.[59-63] From a basic enzymology perspective, it is of interest to explore whether the different substrate binding mechanisms used by different proteases reflect their biological roles, including hydrolysis efficiency and selectivity. PL^{pro} and M^{pro} are excellent models to explore this question given their medicinal interest and apparently different substrate selectivities.

Methods

Peptide docking

The structure of SARS-CoV-2 PL^{pro} in which the active site Cys111 was covalently attached to the peptidomimetic inhibitor VIR251 was taken from the RCSB Protein Data Bank (PDB 6WX4, 1.66 Å resolution, https://www.rcsb.org/structure/6WX4).[1] Protein affinity maps covering the entire protein (centre = (-0.785, -19.447, -32.879); box dimensions (Å) = (65.25, 50.25, 99.75)) were calculated with AutoGridFR (v. 1.2).[64] Starting from an extended conformation and with the search space covering the entire protein, each oligopeptide was docked with AutoDock CrankPep (ADCP; in ADFRsuite v. 1.0),[36] with 100 replicas and 3M energy evaluations per peptide amino acid, and the solutions outputted following clustering with a native contact cut-off of 0.8. The 100 outputted solutions were then filtered based on a comparison of the docked P4-P1 residues with VIR251. A solution was deemed to pass the filter if all four Cα atoms in P4-P1 were within 2 Å of the corresponding VIR251 Cα atoms. Protein-peptide complexes were visualised using PyMOL (open source, v. 2.3.0).[65]

System setup

The PL^{pro} structure (PDB 6WX4) was prepared with Reduce (MolProbity, Duke University),[66] with protonation states determined with H++ (Virginia Tech),[67] PROPKA3 (PDB2PQR),[68] and visual inspection. No flips were applied. All Asp/Glu residues were modelled as deprotonated; all Lys/Arg were protonated; all Cys residues were neutral and protonated, except the four Zn-coordinating residues (Cys189, 192, 224, 226), and Cys111 which is discussed below; His assignments are described in **Table S6.6**. For the active site Cys111 and His272 residues, both the neutral ("N"; His272 neutral, Nε-protonated) and zwitterionic ("Z") states were considered. Using PyMOL (open source, v. 2.3.0),[65] the protein N-terminal was capped with an acetyl group (ACE-EVRTI) and the C-terminal was capped with N-methyl (TTIKP-NME). The resultant charge of the protein, including Zn^{2+}, was +1. Crystallographic waters were retained. In the peptides, the

N-terminals were uncapped and the C-terminals NH_2-capped, in accord with experimental conditions. For the comparative modelling of the nsp peptides, the residues in the original peptide were modified to match the peptide sequences of interest using the mutagenesis tool of PyMOL (open source, v. 2.3.0).[65] For each residue, the least sterically clashing backbone-dependent rotamer was adopted.[69]

Molecular dynamics (MD) simulations

MD simulations were conducted with GROMACS (versions 2019.2 and 2020.4),[37] using the AMBER99SB-ILDN forcefield for amino acid residues,[38] non-bonded parameters for Zn,[70] and the TIP3P water model.[39] The system was placed in a rhombic dodecahedral box (1 nm buffer), solvated, neutralised with sodium/chloride ions, and minimised until maximum force fell below 1000 kJ mol^{-1} nm^{-1}. By generating random velocities at 298.15 K, three replicas per system were initiated, then equilibrated with non-hydrogen atom restraints for 200 ps (1 fs step) under NVT conditions at 298.15 K, equilibrated for 200 ps (1 fs step) under NPT conditions at 298.15 K and 1.0 bar, before a 100 ns or 200 ns (2 fs step) production MD. Temperature was maintained at 298.15 K with a velocity-rescaling thermostat with a stochastic term (time constant 0.1 ps; separate coupling of protein and non-protein),[71] and pressure at 1.0 bar with a Parrinello-Rahman barostat (time constant 2 ps).[72, 73] Long-range electrostatics was treated with smooth Particle-mesh Ewald (1 nm cut-off).[74, 75] Van der Waals interactions were treated with a 1 nm cut off. Analysis of production MD simulations was conducted using GROMACS tools (v 2019.2).[37] To obtain representative structures of peptide-bound PL^{pro}, clustering was performed with a 3 Å RMSD cut-off on the peptide backbone (when the focus is on peptide fold) or peptide non-hydrogen atoms (when the focus is on PL^{pro}-peptide interactions, including sidechain interactions) using the gromos algorithm.[76] A hydrogen bond (HB) was assigned if the donor-acceptor distance was below 3.5 Å and the hydrogen-donor-acceptor angle was below 30°. Peptide secondary structure assignments were conducted using DSSP (version 2.0.4).[77, 78]

Quantum mechanics/molecular mechanics (QM/MM) calculations

To investigate the preferred state of the catalytic triad, proton transfers between Cys111 and His272, and between His272 and Asp286, were modelled with QM/MM umbrella sampling (US). For each of the *apo* or peptide **2**-bound PL^{pro} in the N- or Z-state, one representative structure was extracted from each of the three MD repeats (200 ns for *apo* PL^{pro}; initial 100 ns for peptide **2**-bound PL^{pro}; frames analysed every ns) as the starting configuration, selecting the frame that had the lowest RMSD compared to the average PL^{pro}(-peptide) structure in the simulation while satisfying all of the following distance criteria: (i) (for N-state) *d*(C111_HG - H272_ND1) or (for Z-state) *d*(C111_SG - H272_HD1) ≤ 2.5 Å; (ii) *d*(D286_ODX - H272_HE2) ≤ 2.5 Å (X = 1 or 2); and in the peptide-bound cases, (iii) *d*(C111_SG - peptide P1_C) ≤ 3.5 Å. A 6 Å solvation shell was retained in the structure. Using Amber18,[41] each system was re-solvated with a 10 Å buffer region, neutralised, minimised, and equilibrated under NVT (10 ps) and NPT (20 ps) conditions to 298.15 K and 1.0 bar, using the MM forcefield restraining all protein and peptide atoms.

The QM region includes the sidechains of PL^{pro} Cys111, His272 and Asp286, and in the case of peptide **2**-bound PL^{pro}, the substrate P2 (atoms C and O), P1 (all atoms), and P1′ (atoms except C and O), with a total of 25 or 44 atoms including link atoms for *apo* or peptide **2**-bound PL^{pro}, respectively (**Figure S6.7**). The Cys111-His272 proton transfer was sampled (k = 100 kcal mol^{-1} $Å^{-2}$) with the reaction coordinate (RC) = *d*(C111_SG - H) - *d*(H272_ND1 - H) between -1 (N-state) and 1 (Z-state) Å in 0.1 Å intervals. In each of the 21 windows, 12.5 ps (1 fs step) of DFTB3/MM-MD was performed under NPT conditions,[79] with the first 2.5 ps discarded as equilibration. Similarly, the His272-Asp286 proton transfer was sampled with RC = *d*(H272_NE2 - H) - *d*(D286_ODX - H) (ODX being the less solvent-exposed carboxylate oxygen) from -1 (doubly protonated His272 and deprotonated Asp286) to 1 (neutral His272 and Asp286) Å in 0.1 Å intervals. In each window a 500 fs (1 fs step) DFT/MM-MD was performed at the PBE0-

D3BJ/6-31G(d) level of theory,[80-83] using Amber18/Gaussian16 (A.03).[41, 42] Free energy profiles were obtained by the weighted histogram analysis method (WHAM).[84]

Protein production

The PL^{pro} domain of the SARS-CoV-2 nsp3 (region Glu746-Thr1063) was produced using *E. coli* Lemo21(DE3) cells and purified as reported previously.[2]

Oligopeptide synthesis

Oligopeptides **4**-**11** were prepared from the C- to N-termini by solid phase peptide synthesis (SPPS) using a Liberty Blue peptide synthesiser (CEM Microwave Technology Ltd.), analogous to the reported synthesis of the oligopeptides **1**-**3**.[2] Rink Amide MBHA resin (AGTC Bioproducts Ltd.) was employed to obtain the oligopeptides as C-terminal amides following cleavage from the resin and purification by semi-preparative HPLC (Shimadzu UK Ltd.).

Solid phase extraction mass spectrometry (SPE-MS)

SPE-MS assays were performed using a RapidFire RF 365 high-throughput sampling robot (Agilent) attached to an iFunnel Agilent 6550 accurate mass quadrupole time-of-flight (Q-TOF) mass spectrometer operated in the positive ionization mode equipped with a C4 SPE cartridge, as reported.[2] Assay conditions: isolated recombinant SARS-CoV-2 PL^{pro} (0.2 μM), oligopeptides **4**-**11** (2.0 μM), and, if appropriate, the N-terminal acetylated C-terminal hydrolysis product peptide of the wt nsp2/3 peptide (**2**) (0.2 μM)[2] in buffer (50 mM Tris, pH 8.0) at 20 ºC or 37 ºC.

Data Availability

Detailed system setup and analysis of docking, MD, and QM/MM-US calculations are available in the Supporting Information. All PL^{pro}-peptide models and MD simulation data (including input and trajectory files) are openly available on GitHub at: https://github.com/duartegroup/PLpro_MD.

Author Contributions

HTHC, LB, CJS and FD conceptualised the study. HTHC carried out computational calculations and analyses. LB carried out oligopeptide synthesis, and SPE-MS experiments and analyses. PL and CS-D carried out PL^{pro} production with supervision from MAW. HTHC and LB wrote the original draft. All authors participated in reviewing of the manuscript. CJS and FD supervised the study.

References

1. Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D.*, et al.* Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **2020,** *6*, eabd4596.
2. Brewitz, L.; Kamps, J.; Lukacik, P.; Strain-Damerell, C.; Zhao, Y.*, et al.* Mass spectrometric assays reveal discrepancies in inhibition profiles for the SARS-CoV-2 papain-like protease. *ChemMedChem* **2022,** *17*, e202200016.
3. Moustaqil, M.; Ollivier, E.; Chiu, H. P.; Van Tol, S.; Rudolffi-Soto, P.*, et al.* SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg. Microbes Infect.* **2021,** *10*, 178.
4. Armstrong, L. A.; Lange, S. M.; Dee Cesare, V.; Matthews, S. P.; Nirujogi, R. S.*, et al.* Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLOS ONE* **2021,** *16*, e0253364.
5. Lv, Z.; Cano, K. E.; Jia, L.; Drag, M.; Huang, T. T.*, et al.* Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development. *Front. Chem.* **2022,** *9*, 819165.
6. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.*, et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374*, 1586.
7. US Food & Drug Administration Coronavirus (COVID-19) Update: FDA Authorizes First Oral Antiviral for Treatment of COVID-19. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19 (accessed 2021-12-22).
8. Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.*, et al.* Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J. Med. Chem.* **2022,** *65*, 6499.
9. Cannalire, R.; Cerchia, C.; Beccari, A. R.; Di Leva, F. S.; Summa, V. Targeting SARS-CoV-2 Proteases and Polymerase for COVID-19 Treatment: State of the Art and Future Opportunities. *J. Med. Chem.* **2022,** *65*, 2716.
10. Nepali, K.; Sharma, R.; Sharma, S.; Thakur, A.; Liou, J.-P. Beyond the vaccines: a glance at the small molecule and peptide-based anti-COVID19 arsenal. *J. Biomed. Sci.* **2022,** *29*, 65.
11. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.*, et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368*, 409.
12. MacDonald, E. A.; Frey, G.; Namchuk, M. N.; Harrison, S. C.; Hinshaw, S. M.*, et al.* Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **2021,** *7*, 2591.
13. Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.*, et al.* Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect. Dis.* **2020,** *6*, 2099.
14. Osipiuk, J.; Azizi, S.-A.; Dvorkin, S.; Endres, M.; Jedrzejczak, R.*, et al.* Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **2021,** *12*, 743.
15. Tan, H.; Hu, Y.; Jadhav, P.; Tan, B.; Wang, J. Progress and Challenges in Targeting the SARS-CoV-2 Papain-like Protease. *J. Med. Chem.* **2022,** *65*, 7561.
16. Meyer, B.; Chiaravalli, J.; Gellenoncourt, S.; Brownridge, P.; Bryne, D. P.*, et al.* Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat. Commun.* **2021,** *12*, 5553.
17. Reynolds, N. D.; Aceves, N. M.; Liu, J. L.; Compton, J. R.; Leary, D. H.*, et al.* The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect. Dis.* **2021,** *7*, 1483.
18. Mohamud, Y.; Xue, Y. C.; Liu, H.; Ng, C. S.; Bahreyni, A.*, et al.* The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem. Biophys. Res. Commun.* **2021,** *540*, 75.
19. Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.*, et al.* Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* **2020,** *39*, e106275.
20. Ratia, K.; Kilianski, A.; Baez-Santos, Y. M.; Baker, S. C.; Mesecar, A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. *PLOS Pathog.* **2014,** *10*, e1004113.
21. Ma, C.; Sacco, M. D.; Xia, Z.; Lambrinidis, G.; Townsend, J. A.*, et al.* Discovery of SARS-CoV-2 Papain-like Protease Inhibitors through a Combination of High-Throughput Screening and a FlipGFP-Based Reporter Assay. *ACS Cent. Sci.* **2021,** *7*, 1245.

22. Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.*, et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020,** *587*, 657.
23. Zhao, Y.; Du, X.; Duan, Y.; Pan, X.; Sun, Y.*, et al.* High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. *Protein Cell* **2021,** *12*, 877.
24. Fu, Z.; Huang, B.; Tang, J.; Liu, S.; Liu, M.*, et al.* The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **2021,** *12*, 488.
25. Shen, Z.; Ratia, K.; Cooper, L.; Kong, D.; Lee, H.*, et al.* Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity. *J. Med. Chem.* **2022,** *65*, 2940.
26. Xu, Y.; Chen, K.; Pan, J.; Lei, Y.; Zhang, D.*, et al.* Repurposing clinically approved drugs for COVID-19 treatment targeting SARS-CoV-2 papain-like protease. *Int. J. Biol. Macromol.* **2021,** *188*, 137.
27. Gil-Moles, M.; Basu, U.; Büssing, R.; Hoffmeister, H.; Türck, S.*, et al.* Gold Metallodrugs to Target Coronavirus Proteins: Inhibitory Effects on the Spike-ACE2 Interaction and on PLpro Protease Activity by Auranofin and Gold Organometallics. *Chem. Eur. J.* **2020,** *26*, 15140.
28. Weglarz-Tomczak, E.; Tomczak, J. M.; Talma, M.; Burda-Grabowska, M.; Giurg, M.*, et al.* Identification of ebselen and its analogues as potent covalent inhibitors of papain-like protease from SARS-CoV-2. *Sci. Rep.* **2021,** *11*, 3640.
29. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
30. Gerber, P. P.; Duncan, L. M.; Greenwood, E. J.; Marelli, S.; Naamati, A.*, et al.* A protease-activatable luminescent biosensor and reporter cell line for authentic SARS-CoV-2 infection. *PLOS Pathog.* **2022,** *18*, e1010265.
31. Kuo, C. J.; Chao, T. L.; Kao, H. C.; Tsai, Y. M.; Liu, Y. K.*, et al.* Kinetic Characterization and Inhibitor Screening for the Proteases Leading to Identification of Drugs against SARS-CoV-2. *Antimicrob. Agents Chemother.* **2021,** *65*, e02577.
32. Han, Y.-S.; Chang, G.-G.; Juo, C.-G.; Lee, H.-J.; Yeh, S.-H.*, et al.* Papain-Like Protease 2 (PLP2) from Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV): Expression, Purification, Characterization, and Inhibition. *Biochemistry* **2005,** *44*, 10349.
33. Shan, H.; Liu, J.; Shen, J.; Dai, J.; Xu, G.*, et al.* Development of potent and selective inhibitors targeting the papain-like protease of SARS-CoV-2. *Cell Chem. Biol.* **2021,** *28*, 855.
34. Zhao, Y.; Zhu, Y.; Liu, X.; Jin, Z.; Duan, Y.*, et al.* Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* **2022,** *119*, e2117142119.
35. Shaqra, A. M.; Zvornicanin, S. N.; Huang, Q. Y. J.; Lockbaum, G. J.; Knapp, M.*, et al.* Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nat. Commun.* **2022,** *13*, 3556.
36. Zhang, Y.; Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019,** *35*, 5121.
37. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
38. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
39. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
40. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
41. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T.E. Cheatham, I.*, et al. AMBER 2018*, University of California, San Francisco, 2018.
42. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.*, et al. Gaussian 16 Rev. A.03*, Wallingford, CT, 2016.
43. Paasche, A.; Zipper, A.; Schafer, S.; Ziebuhr, J.; Schirmeister, T.*, et al.* Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry* **2014,** *53*, 5930.
44. Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020,** *10*, 12544.

45. Mladenovic, M.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. On the Origin of the Stabilization of the Zwitterionic Resting State of Cysteine Proteases: A Theoretical Study. *J. Am. Chem. Soc.* **2008,** *130*, 8696.
46. Henderson, J. A.; Verma, N.; Harris, R. C.; Liu, R.; Shen, J. Assessment of proton-coupled conformational dynamics of SARS and MERS coronavirus papain-like proteases: Implication for designing broad-spectrum antiviral inhibitors. *J. Chem. Phys.* **2020,** *153*, 115101.
47. Hognon, C.; Marazzi, M.; García-Iriepa, C. Atomistic-Level Description of the Covalent Inhibition of SARS-CoV-2 Papain-like Protease. *Int. J. Mol. Sci.* **2022,** *23*, 5855.
48. Aplin, R. T.; Robinson, C. V.; Schofield, C. J.; Westwood, N. J. Does the observation of noncovalent complexes between biomolecules by electrospray ionisation mass spectrometry necessarily reflect specific solution interactions? *J. Chem. Soc., Chem. Commun.* **1994**, 2415.
49. Santos, L. H.; Kronenberger, T.; Almeida, R. G.; Silva, E. B.; Rocha, R. E. O.*, et al.* Structure-Based Identification of Naphthoquinones and Derivatives as Novel Inhibitors of Main Protease Mpro and Papain-like Protease PLpro of SARS-CoV-2. *J. Chem. Inf. Model.* **2022,** *62*, 6553.
50. Báez-Santos, Y. M.; Mielech, A. M.; Deng, X.; Baker, S.; Mesecar, A. D. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J. Virol.* **2014,** *88*, 12511.
51. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.*, et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265.
52. Drag, M.; Mikolajczyk, J.; Bekes, M.; Reyes-Turcu, F. E.; Ellman, J. A.*, et al.* Positional-scanning fluorigenic substrate libraries reveal unexpected specificity determinants of DUBs (deubiquitinating enzymes). *Biochem. J.* **2008,** *415*, 367.
53. Wydorski, P. M.; Osipiuk, J.; Lanham, B. T.; Tesar, C.; Endres, M.*, et al.* Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. *Nat. Commun.* **2023,** *14*, 2366.
54. Karanasios, E.; Stapleton, E.; Manifava, M.; Kaizuka, T.; Mizushima, N.*, et al.* Dynamic association of the ULK1 complex with omegasomes during autophagy induction. *J. Cell Sci.* **2013,** *126*, 5224.
55. Gassen, N. C.; Papies, J.; Bajaj, T.; Emanuel, J.; Dethloff, F.*, et al.* SARS-CoV-2-mediated dysregulation of metabolism and autophagy uncovers host-targeting antivirals. *Nat. Commun.* **2021,** *12*, 3818.
56. Lim, C. T.; Tan, K. W.; Wu, M.; Ulferts, R.; Armstrong, L. A.*, et al.* Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp3 papain-like protease. *Biochem. J.* **2021,** *478*, 2517.
57. Gorbalenya, A. E.; Snijder, E. J. Viral cysteine proteinases. *Perspect. Drug Discov. Des.* **1996,** *6*, 64.
58. Bakshi, S.; Holzer, B.; Bridgen, A.; McMullan, G.; Quinn, D. G.*, et al.* Dugbe virus ovarian tumour domain interferes with ubiquitin/ISG15-regulated innate immune cell signalling. *J. Gen. Virol.* **2013,** *94*, 298.
59. Reed, K. E.; Grakoui, A.; Rice, C. M. Hepatitis C virus-encoded NS2-3 protease: cleavage-site mutagenesis and requirements for bimolecular cleavage. *J. Virol.* **1995,** *69*, 4127.
60. Thibeault, D.; Maurice, R.; Pilote, L.; Lamarre, D.; Pause, A. In Vitro Characterization of a Purified NS2/3 Protease Variant of Hepatitis C Virus. *J. Biol. Chem.* **2001,** *276*, 46678.
61. Schregel, V.; Jacobi, S.; Penin, F.; Tautz, N. Hepatitis C virus NS2 is a protease stimulated by cofactor domains in NS3. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 5342.
62. Isken, O.; Langerwisch, U.; Jirasko, V.; Rehders, D.; Redecke, L.*, et al.* A conserved NS3 surface patch orchestrates NS2 protease stimulation, NS5A hyperphosphorylation and HCV genome replication. *PLOS Pathog.* **2015,** *11*, e1004736.
63. Boukadida, C.; Fritz, M.; Blumen, B.; Fogeron, M.-L.; Penin, F.*, et al.* NS2 proteases from hepatitis C virus and related hepaciviruses share composite active sites and previously unrecognized intrinsic proteolytic activities. *PLOS Pathog.* **2018,** *14*, e1006863.
64. Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Comput. Biol.* **2015,** *11*, e1004586.
65. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
66. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.*, et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018,** *27*, 293.

67. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012,** *40*, W537.
68. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525.
69. Dunbrack Jr., R. L.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997,** *6*, 1661.
70. Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991,** *113*, 8262.
71. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126*, 014101.
72. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52*, 7182.
73. Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50*, 1055.
74. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089.
75. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H., *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.
76. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F., *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
77. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E., *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015,** *43*, D364.
78. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577.
79. Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013,** *9*, 338.
80. Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999,** *110*, 6158.
81. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456.
82. Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* **1973,** *28*, 213.
83. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S., *et al.* Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982,** *77*, 3654.
84. Grossfield, A. WHAM: an implementation of the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham/ (accessed 2020-04-17).

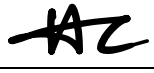
6.3 Statement of authorship

Statement of Authorship for joint/multi-authored papers for PGR thesis
To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2′ proline of the viral polyprotein
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Preprint deposited on bioRxiv; submitted to RSC Chemical Biology. Chan, H. T. H.; Brewitz, L.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A.; Schofield, C. J.; Duarte, F. Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2' proline of the viral polyprotein. *bioRxiv* **2023**, 2023.07.11.548309.

Student Confirmation

Student Name:	Ho Ting Henry CHAN		
Contribution to the Paper	Conceptualised the study; carried out computational calculations and analyses; wrote the original draft with input from other authors; reviewed the manuscript (as detailed in "Author contributions" section).		
Signature		Date	1 August 2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Fernanda Duarte		
Supervisor comments I confirm the description above regarding authorship and contributions is accurate.		
Signature 	Date	2 August 2023

This completed form should be included in the thesis, at the end of the relevant chapter.

7. PL^{pro} Binding of Ubiquitin-Like Substrates

7.1 Overview

In addition to catalysing the hydrolysis of linear peptide sequences in viral polyproteins, PL^{pro} recognises and cleaves the isopeptide bond in N^{ε}-lysine-branched peptides, a feature found in human and viral proteins modified by ubiquitin or ubiquitin-like modifiers, such as ISG15 (**Chapter 1**). The deubiquitinating and deISG15ylating roles of SARS-CoV-2 PL^{pro} have pro-viral roles including the evasion of human immune responses. Given the ubiquitous nature of these post-translational modifications and their potential importance to signalling processes, an efficient activity assay identifying ubiquitin-like substrates of PL^{pro} is of biological interest. In this Chapter, the potential of SPE-MS experiments and computational modelling towards this aim is demonstrated in collaborative studies. Based on models of polyprotein derived linear peptides (**Chapter 6**), PL^{pro} complexes with N^{ε}-lysine-branched peptides were investigated computationally, with substrate recognition again being manifested by loop dynamics.

7.2 Manuscript V

This manuscript (**Manuscript V**) has been submitted to Journal of Virology and is under review at the time of writing.

V. Brewitz, L.*; Chan, H. T. H.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A.; Duarte, F; Schofield, C. J.* Mass spectrometric assays monitoring the deubiquitinase activity of the SARS-CoV-2 papain-like protease inform on the basis of substrate selectivity and have utility for substrate identification. *Submitted to J. Virol.* **2023**.

I carried out all computational work, wrote the associated manuscript section, and contributed to discussions and review of manuscript. A Statement of Authorship is appended following the manuscript. The computational Supplementary Information is in the **Appendix for Chapter 7**.

Mass spectrometric assays monitoring the deubiquitinase activity of the SARS-CoV-2 papain-like protease inform on the basis of substrate selectivity and have utility for substrate identification

Abstract

The SARS-CoV-2 papain-like protease (PL^{pro}) and main protease (M^{pro}) are nucleophilic cysteine enzymes that catalyse the hydrolysis of the viral polyproteins pp1a/1ab. By contrast with M^{pro}, PL^{pro} is also a deubiquitinase (DUB) that accepts post-translationally modified human proteins as substrates. Here we report studies on the DUB activity of PL^{pro} using synthetic N^{ε}-lysine-branched oligopeptides as substrates that mimic post-translational protein modifications by ubiquitin (Ub) or Ub-like modifiers (UBLs), such as interferon stimulated gene 15 (ISG15). Mass spectrometry (MS)-based assays confirm the DUB activity of isolated recombinant PL^{pro}. They reveal that the sequence of both the peptide fragment derived from the post-translationally modified protein and that derived from the UBL affects PL^{pro} catalysis; however, the nature of substrate binding in the S sites appears to be more important for catalytic efficiency than binding in the S′ sites. Importantly, the results reflect the reported cellular substrate selectivity of PL^{pro}, *i.e.*, human proteins conjugated to ISG15 are better substrates than those conjugated to Ub or other UBLs. The combined experimental and modelling results imply that PL^{pro} catalysis is affected not only by the identity of the substrate residues binding in the S and S′ sites, but also by the substrate fold and the conformational dynamics of the blocking loop 2 of the PL^{pro}-substrate complex. N^{ε}-Lysine-branched oligopeptides thus have potential to help the identification of PL^{pro} substrates. More generally, the results imply that MS-based assays with N^{ε}-lysine-branched oligopeptides have potential to monitor catalysis by human DUBs and hence to inform on their substrate preferences.

Keywords

SARS-CoV-2 papain-like protease (PL^{pro}), nucleophilic cysteine protease, deubiquitinase (DUB), ubiquitin, interferon stimulated gene 15 (ISG15), mass spectrometry, ubiquitin-like modifier, virus-host interactions, interferon regulatory factor 3 (IRF3), eIF4E-homologous protein (4EHP)

Introduction

The SARS-CoV-2 genome encodes two nucleophilic cysteine proteases that catalyse hydrolysis of the viral polyproteins pp1a/1ab into functional non-structural proteins (nsps), *i.e.* the papain-like protease (PL^{pro}, a domain of nsp3) and the main protease (M^{pro}, nsp5).[1, 2] PL^{pro} catalyses the hydrolysis of peptide bonds in pp1a/1ab C-terminal to three LXGG motifs (X represents a non-conserved residue) to release nsps1-3, while M^{pro} catalyses the hydrolysis of peptide bonds in pp1a/1ab C-terminal to eleven glutamine residues in (L/F/V)Q(S/A/N) motifs to release nsps 4-16.[1] Catalysis of the two SARS-CoV-2 proteases is essential for viral replication and their inhibition is of therapeutic relevance,[3-7] *i.e.* the small-molecules nirmatrelvir and ensitrelvir, which are clinically used to treat COVID-19, inhibit M^{pro}.[8, 9]

The proteolytic functions of both SARS-CoV-2 M^{pro} and, in particular, PL^{pro} extend beyond catalysing the hydrolysis of pp1a/1ab to catalysing the hydrolysis of (iso)peptide amide bonds in human proteins, in accord with the reported substrate promiscuity of other coronavirus proteases.[10-14] Interestingly, only a few of the human proteins that contain (L/F/V)Q motifs in their coding sequence are validated substrates of isolated recombinant SARS-CoV-2 M^{pro}, *e.g.* human nuclear factor (NF)-κB essential modulator (NEMO), the M^{pro}-catalysed hydrolysis of which has been reported to induce the death of brain endothelial cells.[15, 16]

PL^{pro} has been proposed to catalyse the hydrolysis of peptide bonds C-terminal to LXGG motifs in the coding sequence of several human proteins, *e.g.* the serine/threonine-protein kinase ULK1 and protein S, on the basis of *in vitro* and/or cellular studies.[17-19] Although proteomic studies with samples from COVID-19 patients indicate that PL^{pro} may accept human proteins as substrates *in vivo*,[20] both the extent and the clinical significance of the PL^{pro}-catalysed degradation of human host proteins during SARS-CoV-2 infections are incompletely understood.

LXGG motifs are conserved in human proteins which are post-translationally modified with ubiquitin (Ub) or ubiquitin-like modifiers (UBL), *e.g.* interferon stimulated gene 15 (ISG15),[21-24]

neural precursor cell expressed, developmentally down-regulated 8 (NEDD8),[25, 26] and ubiquitin-related modifier-1 (URM1).[27, 28] Ub and UBLs are linked to one or multiple N^{ε}-amino groups of a protein lysine residue via an isopeptide amide bond formed with their C-terminal residue, which forms a LXGG motif in the case of Ub, ISG15, NEDD8, and URM1.[29, 30] The post-translational (poly)ubiquitinylation of human proteins has multiple regulatory functions, including signalling proteins for degradation via the proteasome.[31-33] Protein (poly)ubiquitinylation is often reversible; formation of the isopeptide amide bond is catalysed by Ub/UBL ligases while its hydrolysis is catalysed by deubiquitinases (DUBs); both Ub/UBL ligases and DUBs may show high levels of specificity, both with respect to the substrate protein and/or the Ub/UBL.[32-37]

SARS-CoV-2 PL^{pro} not only catalyses the hydrolysis of the pp1a/1ab polyproteins, but also possesses DUB activity and catalyses the hydrolysis of isopeptide amide bonds in human proteins which are post-translationally modified with Ub/UBLs,[38, 39] similar to the reported DUB activities of PL^{pro} from other coronaviruses.[10-12, 40] Cellular studies reveal that SARS-CoV-2 PL^{pro} preferentially catalyses the hydrolysis of isopeptide amide bonds C-terminal to ISG15, whereas SARS-CoV PL^{pro} prefers diubiquitin (Ub_2) substrates over ISG15;[38, 39] both Ub_2 and ISG15 bind to the active site of PL^{pro} and, additionally, to allosteric sites.[41-43] The apparent selectivity of SARS-CoV-2 PL^{pro} for catalysing deISG15ylations is precedented by the substrate selectivity of the human DUB ubiquitin specific peptidase 18 (USP18) which also employs a nucleophilic cysteine to selectively catalyse protein deISG15ylations over protein deubiquitinylations.[44] Notably, the DUB activity of SARS-CoV-2 PL^{pro} is reported to attenuate the host innate immune response via catalysing the deISG15ylation of ISG15ylated interferon regulatory factor 3 (IRF3), thus potentially enhancing virulence.[38, 45]

Reported assays which monitor the DUB activity of PL^{pro} typically employ Ub/UBL derivatives conjugated to a fluorescent group via a C-terminal amide (*e.g.*, Ub-AMC or ISG15-AMC; AMC: 7-amino-4-methylcoumarin), the hydrolysis of which can be spectroscopically monitored.[38, 39, 46, 47] Matrix-assisted laser desorption/ionisation (MALDI) mass spectrometry (MS)-based assays,

originally developed to monitor the activity of human DUBs,[48] have also been employed to monitor the DUB activity of PL^{pro}; however, the MALDI-MS assay uses Ub_2 as substrate which is apparently not the preferred PL^{pro} substrate in cells.[49] The use of substrates which do not bind to the PL^{pro} S′ sites, the relatively high costs associated with use of the Ub-AMC, ISG15-AMC, and Ub_2 substrates, and/or the use of artificial C-terminal fluorophores limit potential applications of these assays.

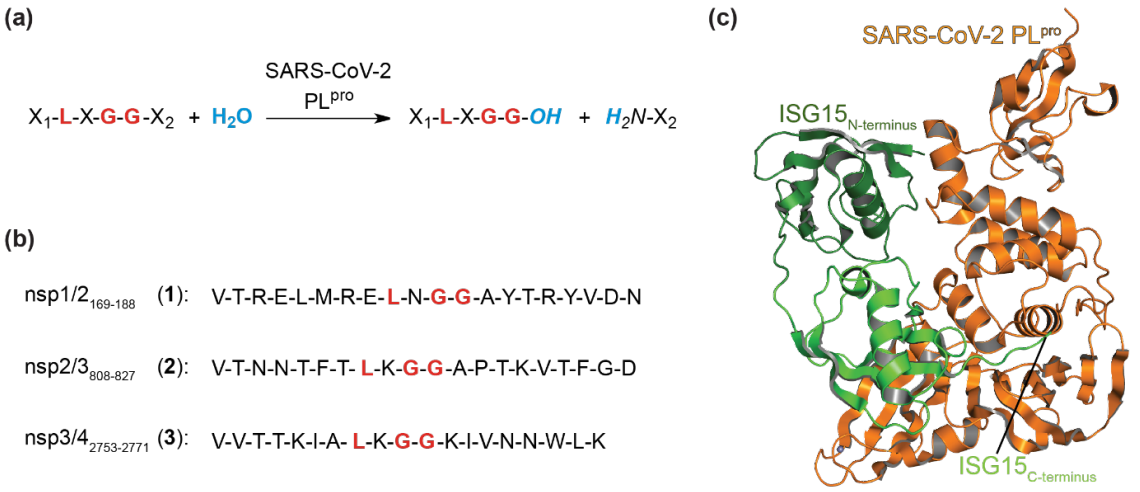


Figure 7.1: PL^{pro} catalysis. (**a**) PL^{pro} catalyses substrate hydrolysis C-terminal to LXGG motifs; (**b**) sequences of the reported SARS-CoV-2 pp1a/1ab-derived linear oligopeptides nsp1/2$_{169-188}$ (**1**), nsp2/3$_{808-827}$ (**2**), and nsp3/4$_{2753-2771}$ (**3**), previously used in SPE-MS assays; PL^{pro} catalysed the hydrolysis of **2** substantially more efficiently than that of **1** and **3**.[50] Residues of the LXGG motif are in red; (**c**) SARS-CoV-2 PL^{pro} (C111S variant, orange) in complex with ISG15 (PDB 7RBS),[43] which is composed of a C- and N-terminal Ub-like domain (light and dark green, respectively), that indicates binding to both the PL^{pro} active site via the C-terminus and the C-terminal Ub-like domain of ISG15 and binding to allosteric sites of PL^{pro} via the N-terminal Ub-like domain (dark green).

We have reported solid-phase extraction coupled to mass spectrometry (SPE-MS)-based assays which directly monitor PL^{pro}-catalysed hydrolysis of oligopeptide fragments derived from the sequence of the Wuhan-Hu-1 strain[51] of SARS-CoV-2 pp1a/1ab (**Figure 7.1b**). The assay results reveal that PL^{pro} preferentially catalyses hydrolysis of an oligopeptide which was based on the LXGG motif separating nsp2 and nsp3, *i.e.* nsp2/3$_{808-827}$ (**2**; **Figure 7.1**).[50] Here we report the results of SPE-MS assays employing N^{ε}-lysine branched synthetic oligopeptides as substrates, that bind to both the S and S′ sites of PL^{pro}, to monitor the DUB activity of isolated recombinant

SARS-CoV-2 PL^{pro}. The results inform on the substrate requirements for efficient PL^{pro} DUB catalysis and reveal that not only the presence of an LXGG motif but also the entire coding sequence of the substrate which binds in proximity of the active site, in particular that of the UBL, determines the catalytic efficiency of PL^{pro}.

Results

Assaying the DUB activity of SARS-CoV-2 PL^{pro} using mass spectrometry. SARS-CoV-2 PL^{pro} is reported to attenuate the human innate immune system by catalysing the deISG15ylation of post-translationally ISG15ylated IRF3;[38] however, to our knowledge, there are no reports on validation of the DUB activity of isolated recombinant PL^{pro} with oligopeptide substrates. We thus synthesised SPE-MS-compatible N^{ε}-lysine-branched oligopeptides based on the three reported ISG15ylation sites of IRF3, *i.e.* Lys193, Lys360, and Lys366,[52] using solid phase peptide synthesis (SPPS). The IRF3-derived peptide fragment was synthesised on a Rink amide resin from C- to N-termini employing Fmoc strategy, *N*-Fmoc lysine with a 4-methyltrityl (Mtt) protected N^{ε}-amine at the site of the lysine branching, and an N-terminal amino acid with a Boc-protected N^{α}-amine. Branching was introduced via selective deprotection of the lysine N^{ε}-Mtt group using $1\%_{v/v}$ trifluoroacetic acid (TFA) and $2\%_{v/v}$ triisopropylsilane (TIPS) in dichloromethane, followed by SPPS to build the ISG15 fragment from C- to N-termini starting with the N^{ε}-amino group of the IRF3 lysine at the C-terminus. The purified branched oligopeptides were obtained following global deprotection and cleavage from the resin with TFA and HPLC purification.

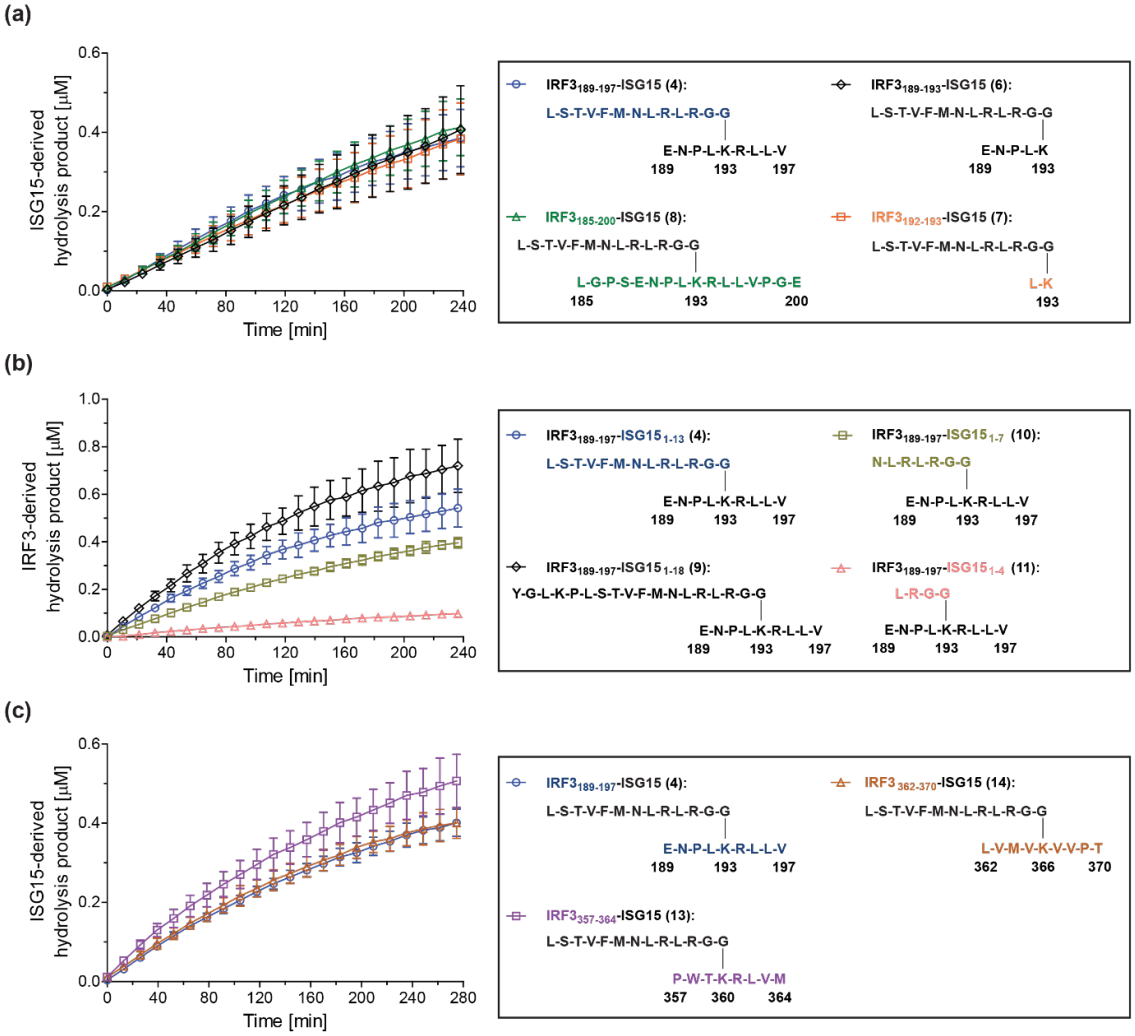


Figure 7.2: PL^{pro} catalyses the hydrolysis of isopeptide amide bonds in N^{ε}-lysine-branched oligopeptides which mimic ISG15ylated IRF3. (**a**) Levels of PL^{pro}-catalysed hydrolysis of the Lys193-branched oligopeptides IRF3$_{189\text{-}197}$-ISG15 (**4**; blue circles), IRF3$_{189\text{-}193}$-ISG15 (**6**; black diamonds), IRF3$_{192\text{-}193}$-ISG15 (**7**; orange boxes), and IRF3$_{185\text{-}200}$-ISG15 (**8**; green triangles); (**b**) levels of PL^{pro}-catalysed hydrolysis of the Lys193-branched IRF3$_{189\text{-}197}$-derived oligopeptides bearing ISG15$_{1\text{-}18}$ (**9**; black diamonds), ISG15$_{1\text{-}13}$ (**4**; blue circles), ISG15$_{1\text{-}7}$ (**10**; olive boxes), and ISG15$_{1\text{-}4}$ (**11**; salmon triangles); (**c**) levels of PL^{pro}-catalysed hydrolysis of the Lys360-branched oligopeptide IRF3$_{357\text{-}364}$-ISG15 (**13**; lavender boxes) and the Lys366-branched oligopeptide IRF3$_{362\text{-}370}$-ISG15 (**14**; brown triangles) compared to that of the Lys193-branched oligopeptide IRF3$_{189\text{-}197}$-ISG15 (**4**; blue circles). Conditions: SARS-CoV-2 PL^{pro} (0.2 μM), substrate peptide (2.0 μM), inert *N*-acetylated inert standard peptide (0.2 μM; Ac-LSTVFMNLRLRGG-NH$_2$ (**5**) in (**a**) and (**c**), Ac-ENPLKRLLV-NH$_2$ (**12**) in (**b**)) in buffer (50 mM Tris, pH 8.0, ambient temperature). Measurement times were normalised to the first sample injection analysed after the addition of PL^{pro} to the Substrate Mixture (t = 0 s), by which time low levels of substrate hydrolysis were manifest. The inert *N*-acetylated hydrolysis products **5** or **12** were used as an internal standard to quantify hydrolysis; SPE-MS assay results are a mean of independent triplicates (N = 3; mean ± standard deviation, SD).

The Lys193-branched IRF3$_{189\text{-}197}$-ISG15 oligopeptide **4** was incubated with the isolated recombinant PL^{pro} domain of SARS-CoV-2 nsp3 (0.2 μM; enzyme/substrate ratio 1/10)

employing our reported conditions for the PL^{pro}-catalysed hydrolysis of the linear pp1a/1ab-derived oligopeptide nsp2/$3_{808-827}$ (**2**) (50 mM Tris, pH 8.0, 37 °C).[50] Analysis of the reaction mixture after 14 h incubation by SPE-MS indicated apparent quantitative isopeptide amide bond hydrolysis, as supported by the observed masses of the resultant N- and C-terminal product fragments. Thus, the results confirmed the DUB activity of PL^{pro}, which had been assigned on the basis of cellular studies and on studies with ISG15 containing a C-terminal fluorophore,[38, 39, 46] with an N^{ε}-lysine-branched oligopeptide substrate.

Time course data were subsequently recorded under the same conditions using an automated SPE-MS setup, however, at ambient temperature rather than at 37 °C. Incomplete levels of PL^{pro}-catalysed isopeptide amide bond hydrolysis were observed within ~160 min incubations under these conditions (~35%; **Figure 7.2a**), which may, at least in part, reflect the inability of **4** to efficiently bind to PL^{pro} at allosteric sites as reported for full-length folded (poly)Ub/UBLs.[41-43] The reaction conditions were subsequently varied, with respect to, *e.g.*, buffer composition, pH, and salt additives; the tested variations did not, however, increase the rates of isopeptide amide bond hydrolysis substantially. Note that the corresponding N-terminally *N*-acetylated ISG15-derived hydrolysis product peptide, *i.e.*, Ac-LSTVFMNLRLRGG-NH_2 (**5**), was added as an inert internal standard to the reaction mixture to enable quantification of PL^{pro}-catalysed product formation; the presence of **5** in the reaction mixture did not affect PL^{pro} catalysis, at least substantially.

Derivatives of the Lys193-branched $IRF3_{189-197}$-ISG15 oligopeptide **4** were synthesised to investigate whether varying the length of the IRF3- or the ISG15-derived fragments affects PL^{pro} catalysis (**Figure 7.2a, b**). Interestingly, SPE-MS analysis of the PL^{pro}-catalysed hydrolysis of the oligopeptides **4**, **6**, **7** and **8** revealed that varying the length of the IRF3-derived fragment did not affect catalysis substantially, albeit the hydrolysis of the Lys193-branched $IRF3_{185-200}$-ISG15 oligopeptide **8**, with the relatively largest IRF3-derived fragment, appeared to be slightly favoured (**Figure 7.2a**). By contrast, varying the length of the ISG15-derived fragment had a substantial

effect on PL^{pro} catalysis; the oligopeptide **9** with the relatively largest ISG15-derived fragment bearing 18 C-terminal amino acids was hydrolysed substantially more efficiently than oligopeptide **11** with the relatively smallest ISG15-derived fragment bearing only the essential C-terminal LRGG tetrad (**Figure 7.2b**). Levels of PL^{pro}-catalysed substrate hydrolysis appear to increase with increasing length of the ISG15-derived fragment peptide; the rank order of PL^{pro} substrate preference was determined by quantifying product formation relative to Ac-ENPLKRLLV-NH_2 (**12**), which was used as an internal standard in the reactions. The combined oligopeptide results indicate that the binding of the N^{ε}-lysine-branched oligopeptides to the S′ sites is less important for catalysis than binding to the S sites. The situation may, however, be different with full-length folded protein substrates. Note that either the corresponding N-terminally *N*-acetylated ISG15-derived hydrolysis product peptide, *i.e.* Ac-LSTVFMNLRLRGG-NH_2 (**5**), or the corresponding N-terminally *N*-acetylated $IRF3_{189-197}$-derived hydrolysis product peptide, *i.e.* Ac-ENPLKRLLV-NH_2 (**12**), was added as an internal standard to the reaction mixtures to enable quantification of PL^{pro}-catalysed product formation; the presence of **5** or **12** in the reaction mixture did not affect PL^{pro} catalysis, at least substantially.

The Lys360-branched $IRF3_{357-364}$-ISG15 peptide **13** and the Lys366-branched $IRF3_{362-370}$-ISG15 peptide **14** were synthesised to investigate the potential of isolated recombinant PL^{pro} to catalyse regioselective deISG15ylations of post-translationally modified IRF3. Oligopeptides **4**, **13**, and **14** were incubated with PL^{pro} in the presence of the internal standard **5**, and isopeptide amide bond hydrolysis was monitored using SPE-MS (**Figure 7.2c**). The results reveal that the sequences of the three tested IRF3-derived peptide fragments, which likely bind to the S′ sites of PL^{pro}, do not affect the rate of peptide hydrolysis substantially, in accord with prior results which showed that the length of the IRF3-derived peptide fragment affects PL^{pro} catalysis less substantially than the length of the ISG15-derived peptide fragment (**Figure 7.2a, b**). The observation that PL^{pro} may catalyse the deISG15ylation of oligopeptide **13** slightly more efficiently than that of **4** and **14** might be a result of the comparatively lower purity of **13** compared to that of all other tested

peptides. Note that PL^{pro} may catalyse the regioselective deISG15ylation of full-length folded IRF3 in a cellular context.

The primary sequence of the PL^{pro} substrates affects catalysis. To investigate whether the substrate preference of SARS-CoV-2 PL^{pro} depends on the primary sequence of the UBLs, we synthesised the oligopeptide derivatives of Lys193-branched $IRF3_{189-197}$-ISG15 **4**, in which the ISG15-fragment was substituted for the corresponding first 13 C-terminal amino acids of Ub ($IRF3_{189-197}$-Ub, **15**) or the UBLs NEDD8 ($IRF3_{189-197}$-NEDD8, **16**), URM1 ($IRF3_{189-197}$-URM1, **17**), and small ubiquitin-related modifier 1 (SUMO1; $IRF3_{189-197}$-SUMO1, **18**), the latter of which does not contain a C-terminal LXGG motif and should hence not be a substrate of PL^{pro}. Lys193-branched $IRF3_{189-197}$ derivatives were attractive synthesis targets because the ubiquitinylation at Lys193 had been reported;[53] the corresponding NEDD8, URM1, and SUMO1 derivatives of **4** were also synthesised despite their unclear biological relevance. The length of the UBL fragment was kept constant to help enable comparison, as PL^{pro} catalyses the hydrolysis of derivatives of **4**, in which the length of the ISG15 fragment was varied, with different efficiencies (**Figure 7.2b**).

The oligopeptides **4**, **15**, **16**, **17**, and **18** were incubated with isolated recombinant SARS-CoV-2 PL^{pro}; hydrolysis of their isopeptide amide bond was monitored in the presence of the corresponding inert N-terminally *N*-acetylated $IRF3_{189-197}$-derived hydrolysis product peptide Ac-ENPLKRLLV-NH_2 (**12**) using SPE-MS (**Figure 7.3a**). The results revealed that PL^{pro} catalyses the hydrolysis of the $IRF3_{189-197}$-ISG15 (**4**) isopeptide amide bond more efficiently than those of $IRF3_{189-197}$-Ub (**15**) and, in particular, $IRF3_{189-197}$-NEDD8 (**16**) (**Figure 7.3a**). However, **4**, **15**, and **16** were substantially more efficient PL^{pro} substrates than $IRF3_{189-197}$-URM1 (**17**). As anticipated, $IRF3_{189-197}$-SUMO1 (**18**) was not a substrate for isolated recombinant PL^{pro} under the tested conditions, an observation which likely reflects the lack of an LXGG motif in **18**, which instead has a QTGG motif that cannot interact efficiently with the S1-S4 sites of PL^{pro} (*vide infra*).

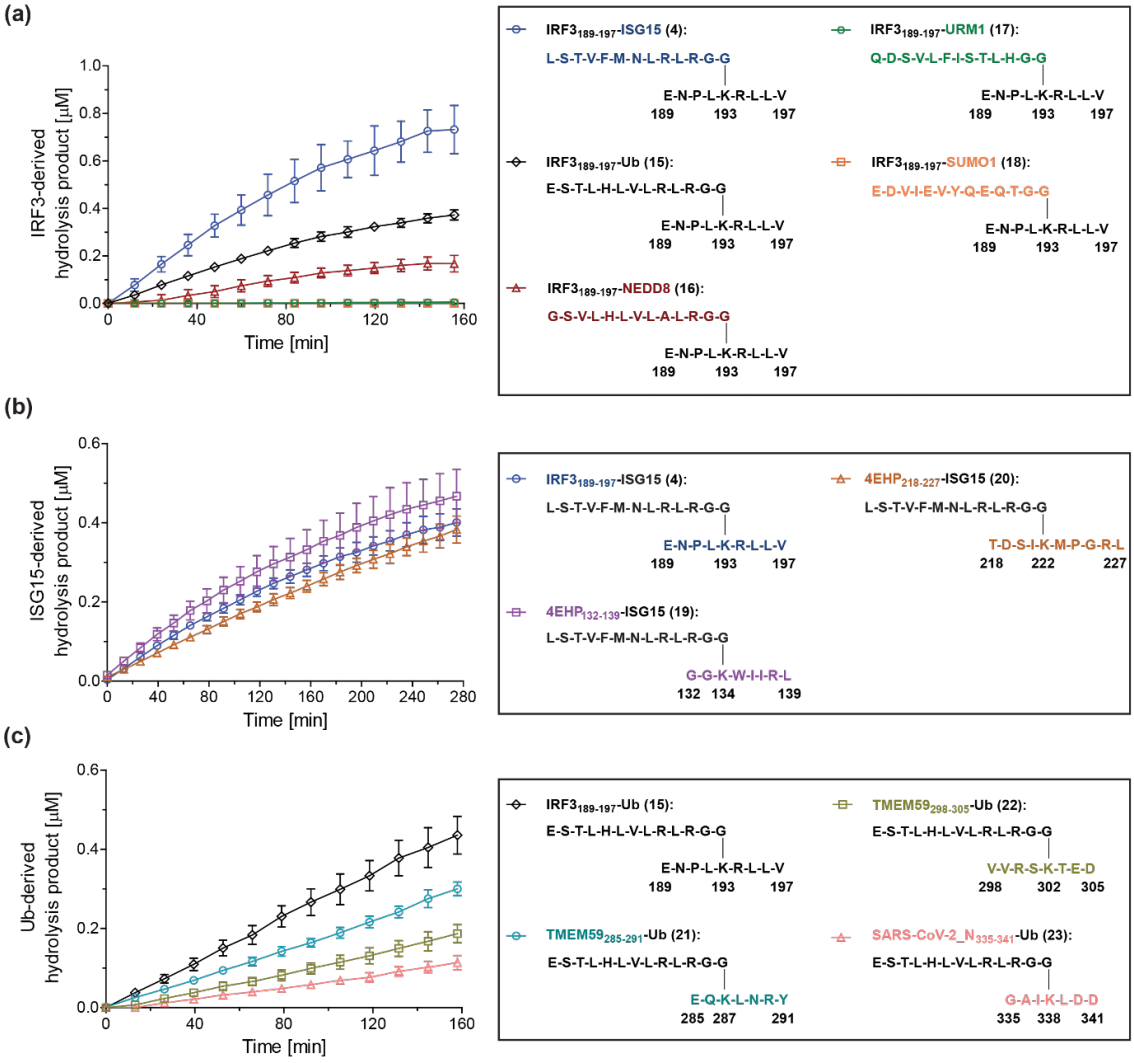


Figure 7.3: PL^{pro} catalyses the hydrolysis of isopeptide amide bonds in N^{ε}-lysine-branched oligopeptides that mimic post-translational protein modification by Ub/UBLs. (**a**) Levels of PL^{pro}-catalysed hydrolysis of oligopeptides derived from Lys193-branched IRF3$_{189-197}$ with ISG15 (**4**; blue circles), Ub (**15**; black diamonds), NEDD8 (**16**; red triangles), URM1 (**17**; green hexagons), and SUMO1 (**18**; orange boxes). Note that SUMO1 does not have an LXGG motif at its C-terminus; (**b**) levels of PL^{pro}-catalysed hydrolysis of the Lys134-branched oligopeptide 4EHP$_{132-139}$-ISG15 (**19**; lavender boxes) and the Lys222-branched oligopeptide 4EHP$_{218-227}$-ISG15 (**20**; brown triangles) compared to that of the Lys193-branched oligopeptide IRF3$_{189-197}$-ISG15 (**4**; blue circles); (**c**) levels of PL^{pro}-catalysed hydrolysis of oligopeptides derived from the lysine-branched IRF3$_{189-197}$-Ub (**15**; black diamonds), TMEM59$_{285-291}$-Ub (**21**; teal circles), TMEM59$_{298-305}$-Ub (**22**; olive boxes), and SARS-CoV-2_N$_{335-341}$-Ub (**23**; salmon triangles) peptides. Conditions: SARS-CoV-2 PL^{pro} (0.2 μM), substrate peptide (2.0 μM), inert *N*-acetylated standard peptide (0.2 μM; Ac-ENPLKRLLV-NH$_2$ (**12**) in (**a**), Ac-LSTVFMNLRLRGG-NH$_2$ (**5**) in (**b**), and Ac-ESTLHLVLRLRGG-NH$_2$ (**24**) in (**c**)) in buffer (50 mM Tris, pH 8.0, ambient temperature). Measurement times were normalised to the first sample injection analysed after the addition of PL^{pro} to the Substrate Mixture (t = 0 s), by which time low levels of substrate hydrolysis were manifest. The inert *N*-acetylated hydrolysis product **5**, **12** or **24** was used as an internal standard to quantify hydrolysis; SPE-MS assay results are a mean of independent triplicates (N = 3; mean ± SD).

$IRF3_{189\text{-}197}$-ISG15 (**4**), $IRF3_{189\text{-}197}$-Ub (**15**), and $IRF3_{189\text{-}197}$-NEDD8 (**16**) all have an LRGG motif, whereas $IRF3_{189\text{-}197}$-URM1 (**17**) has an LHGG motif which may rationalise why it is a substantially poorer substrate than **4**, **15**, and **16** (**Figure 7.3a**). Nonetheless, the assay results clearly show that the different hydrolysis rates of **4**, **15**, and **16** are not dependent on the presence of a LRGG motif or an $IRF3_{189\text{-}197}$ fragment, suggesting that the primary/coding sequence of the UBL N-terminal to the LXGG motif is important in PL^{pro} catalysis.

Interestingly, the results with **4**, **15**, and **16** reflect reported SARS-CoV-2 PL^{pro} substrate preferences observed in cellular studies which also showed that PL^{pro}-catalysed protein deISG15ylations are more efficient than the corresponding deubiquitinylations, which are both more efficient than deNEDD8ylations.[38, 39] Thus, the combined evidence suggests that not only the substrate fold affects PL^{pro} catalysis, *inter alia* via binding to allosteric sites, but also, in particular, the sequence identity of the amino acid fragment N-terminal to the LXGG motif of the substrate.

In general, the ISG15ylation sites of human proteins are poorly characterised, likely in part, because the standard workflow to localise protein ubiquitinylation sites, *i.e.*, protein denaturation followed by tryptic digestion and analysis of the resultant peptide fragments by MS/MS, affords the same lysine-branched peptide fragments for both ubiquitinylated and ISG15ylated proteins. Thus, a different workflow is required, *e.g.*, involving digestion using non-standard proteases with different substrate selectivities than trypsin.[34, 54-57] Apart from ISG15ylation sites of IRF3, the ISG15ylation sites of the eIF4E-homologous protein (4EHP) have been characterised, *i.e.* at Lys134 and Lys222.[58] 4EHP binds to the mRNA 5′-cap structure and suppresses translation by competing with the eukaryotic translation initiation factor 4E (eIF4E) for binding to the cap structure. The ISG15ylation of 4EHP is proposed to control translation during immune responses;[58] note that the translation initiation factor eIF4G is a reported substrate of rhinovirus $2A^{pro}$ and of the foot-and-mouth disease virus (FMDV) leader protease (L^{pro}), suggesting that

modulation of protein translation is a general strategy of *Riboviria* to evade the host immune system.[59]

The Lys134-branched 4EHP$_{132-139}$-ISG15 peptide **19** and the Lys222-branched 4EHP$_{218-227}$-ISG15 peptide **20** were synthesised to investigate the potential of isolated recombinant SARS-CoV-2 PL^{pro} to catalyse selective deISG15ylations of post-translationally modified 4EHP. Oligopeptides **4**, **19**, and **20** were incubated with PL^{pro} in the presence of the inert internal standard **5**, and isopeptide amide bond hydrolysis was monitored using SPE-MS (**Figure 7.3b**). Interestingly, the results reveal that the rates of PL^{pro}-catalysed peptide hydrolyses for **4**, **19**, and **20** are similar, within experimental error (**Figure 7.3b**). This observation is in accord with the previous results showing that PL^{pro}-catalysed the deISG15ylation of IRF3-derived fragment peptides regardless of the site of ISG15ylation (**Figure 7.2c**) and thus supports the proposal that the substrate binding in the S sites of PL^{pro} is more important for efficient catalysis than substrate binding in the S′ sites of PL^{pro}, in accord with reported studies on the substrate efficiency of oligopeptides based on SARS-CoV-2 nsps.[60] Note that the observed variation in the absolute efficiency of PL^{pro} catalysis appeared to depend on the batch of PL^{pro}; hence, we determined the rank order of substrate preference by quantifying product formation relative to internal standards, the rank order did not depend on the batch of PL^{pro}. The results suggest that ISG15ylated 4EHP may be a substrate of PL^{pro} in cells.

***N*$^{\varepsilon}$-Lysine-branched oligopeptides may help enable the identification of PL^{pro} substrates.** To further dissect the effect of the UBL fragment of the tested *N*$^{\varepsilon}$-lysine-branched oligopeptides, which binds to the PL^{pro} S sites, on the rate of isopeptide amide bond hydrolysis from that of the fragment which binds to the PL^{pro} S′ sites, and to investigate the potential of synthetic *N*$^{\varepsilon}$-lysine-branched oligopeptides for the identification of SARS-CoV-2 PL^{pro} substrates, we synthesised a set of *N*$^{\varepsilon}$-lysine-branched oligopeptides which mimic protein ubiquitinylation, *i.e.* via employing the 13 C-terminal residues of Ub. In general, the sites of protein (poly)ubiquitinylations have been characterised in greater detail than the sites of protein ISG15ylations. Potential SARS-CoV-2

PL^{pro} substrates were chosen on the basis of reported proteomic studies which showed, *e.g.*, that Lys338 of the SARS-CoV-2 nucleocapsid (N) protein can be ubiquitinylated.[61] Reduced levels of ubiquitinylation of the autophagy-related human transmembrane protein TMEM59 (at *e.g.* Lys287 and Lys302) have been associated with SARS-CoV-2 infections, which may suggest that TMEM59 could be a substrate of SARS-CoV-2 PL^{pro}.[61] We thus synthesised the corresponding Lys287-branched $TMEM59_{285\text{-}291}$-Ub peptide **21**, the Lys302-branched $TMEM59_{298\text{-}305}$-Ub peptide **22**, and the Lys338-branched SARS-CoV-2_$N_{335\text{-}341}$-Ub peptide **23**.

The oligopeptides **15**, **21**, **22**, and **23** were incubated with isolated recombinant SARS-CoV-2 PL^{pro} and hydrolysis of their isopeptide amide bond was monitored using SPE-MS (**Figure 7.3c**). Note that the corresponding N-terminally *N*-acetylated Ub-derived hydrolysis product peptide, *i.e.*, Ac-ESTLHLVLRLRGG-NH_2 (**24**), was added as an inert internal standard to the reaction mixtures to enable quantification of PL^{pro}-catalysed product formation; the presence of **24** in the reaction mixture did not affect PL^{pro} catalysis substantially. The results reveal that SARS-CoV-2 PL^{pro} catalyses the hydrolysis of the $IRF3_{189\text{-}197}$-Ub (**15**) isopeptide amide bond more efficiently than those of **21**, **22**, and **23** (**Figure 7.3c**). The PL^{pro}-catalysed isopeptide hydrolysis in $IRF3_{189\text{-}197}$-Ub (**15**) was ~2-fold more efficient as in SARS-CoV-2_$N_{335\text{-}341}$-Ub (**23**), which was the least efficient substrate of the four substrates tested. This observation indicates that PL^{pro} catalysis can be affected by substrate binding to the S′ sites and, by implication, also by the substrate fold in proximity of the S′ sites, by contrast to previous results with peptide fragments mimicking the ISG15ylation of IRF3 and 4EHP (**Figure 7.2)**.

The combined results suggest that PL^{pro} may be promiscuous (including compared to M^{pro}) with respect to its *in vivo* protein substrates, and that it likely has the capability to catalyse a broad range of deISG15ylation and deubiquitinylation reactions of both host and viral proteins. The results also highlight the potential of N^{ε}-lysine-branched oligopeptides to help enable identification of PL^{pro} substrates, including post-translationally modified viral proteins.

Single amino acid substitutions in PL^{pro} substrates can affect catalysis. The observation that IRF3$_{189-197}$-ISG15 (**4**), IRF3$_{189-197}$-Ub (**15**), and IRF3$_{189-197}$-NEDD8 (**16**) all have an LRGG motif, whereas IRF3$_{189-197}$-URM1 (**17**) has an LHGG motif and is, relative to **4**, **15**, and **16**, a substantially poorer PL^{pro} substrate raises the question as to whether PL^{pro} catalyses the hydrolysis of the corresponding IRF3$_{189-197}$-URM1-R (**25**) variant of **17**, in which the LHGG histidine residue has been substituted for an arginine residue, more efficiently than that of **17** (**Figure 7.3a**). **25** was thus synthesised and incubated with isolated recombinant PL^{pro}, and its hydrolysis was monitored using SPE-MS (**Figure 7.4a**).

Interestingly, although the PL^{pro}-catalysed hydrolysis of **25** was ~4-fold more efficient than that of **17** (after 4 h), it remained ~7-fold less efficient compared to the PL^{pro}-catalysed hydrolysis of IRF3$_{189-197}$-ISG15 (**4**) (after 4 h), indicating that peptides with an LRGG motif are, in principle, better substrates than those with an LHGG motif (**Figure 7.4a**). However, considering that **25** and **4** are based on the identical IRF3$_{189-197}$ fragment C-terminal to their LRGG motif which likely binds to the S′ sites of PL^{pro} in an identical manner, the results support the proposal that the primary sequence of the substrate N-terminal to the LXGG motif can substantially affect catalysis. This observation may reflect the importance of substrate binding to S sites other than S1-S4, including at allosteric positions, and/or the presence of secondary structural elements in the substrate that favour PL^{pro} catalysis. Note that substrate binding to allosteric sites by full-length folded substrate proteins has been proposed to be important in PL^{pro} catalysis on the basis of crystallographic studies.[41-43]

The PL^{pro}-catalysed hydrolysis of TMEM59$_{298-305}$-Ub (**22**) and SARS-CoV-2_N$_{335-341}$-Ub (**23**), which bear an acidic ED or DD sequence, respectively, in the sequence C-terminal to their LXGG motif, is less efficient compared to that of IRF3$_{189-197}$-Ub (**15**) (**Figure 7.3c**), which bears a hydrophobic LV sequence at the corresponding position. This observation raises the possibility that acidic residues proximate to the substrate P1′ lysine residue involved in the isopeptide amide bond, are detrimental for PL^{pro} catalysis. To test this proposal, we synthesised the corresponding

SARS-CoV-2_$N_{335\text{-}341}$-Ub_$L_{340}V_{341}$ variant (**26**), in which the DD sequence was substituted for an LV sequence as present in **15**, and monitored its PL^{pro}-catalysed hydrolysis using SPE-MS (**Figure 7.4b**). The results reveal that PL^{pro} catalyses the hydrolysis of **26** with similar efficiency as that of **15** and ~2.5-fold more efficiently than that of **23** (after 4 h), supporting the proposal that hydrophobic residues may be preferred in proximity of the substrate lysine residue involved in the isopeptide amide bond, potentially due to improved binding to the PL^{pro} S′ sites.

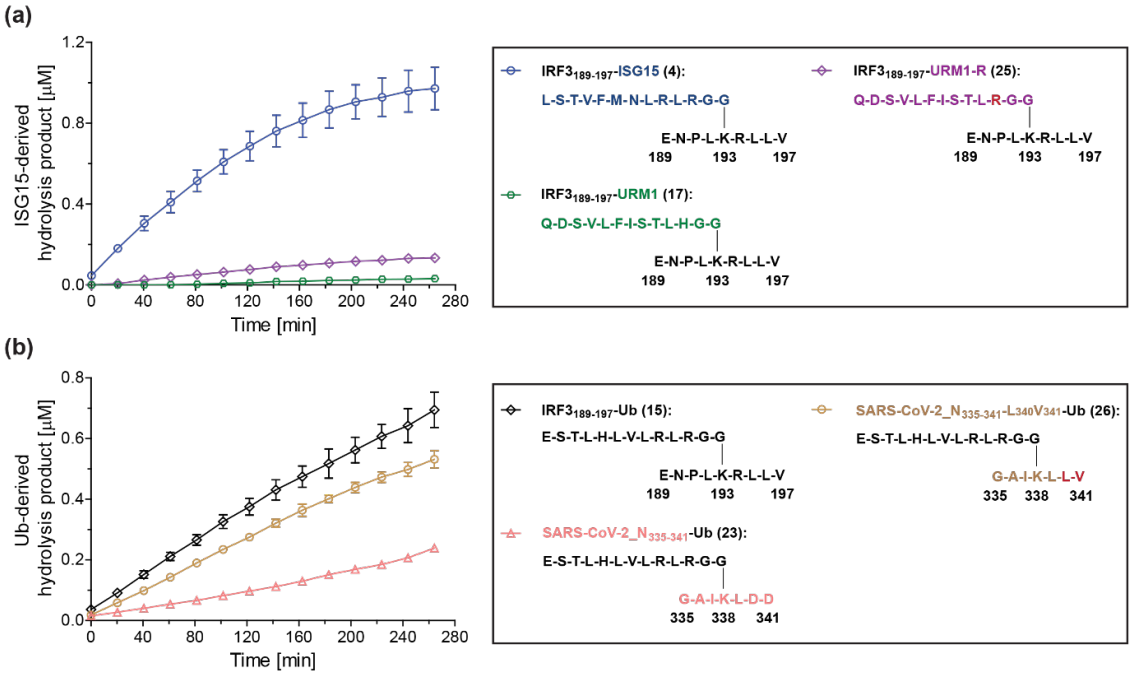


Figure 7.4: The primary sequence of *N*ε-lysine-branched oligopeptides affects the efficiency of PL^{pro} catalysis. (**a**) PL^{pro}-catalysed hydrolysis of the H-to-R variant of the Lys193-branched IRF3189-197-URM1 peptide **17** (green hexagons), *i.e.* **25** (lavender diamonds), compared to that of IRF3189-197-ISG15 (**4**; blue circles); (**b**) PL^{pro}-catalysed hydrolysis of SARS-CoV-2_$N_{335\text{-}341}$-Ub_$L_{340}V_{341}$ (**26**; brown circles) compared to that of SARS-CoV-2_$N_{335\text{-}341}$-Ub (**23**; salmon triangles) and IRF3189-197-Ub (**15**; black diamonds). Substituted residues are in red. Conditions: SARS-CoV-2 PL^{pro} (0.2 µM), substrate peptide (2.0 µM), inert *N*-acetylated standard peptide (0.2 µM; Ac-ENPLKRLLV-NH_2 (**12**) in (**a**), Ac-ESTLHLVLRLRGG-NH_2 (**24**) in (**b**)) in buffer (50 mM Tris, pH 8.0, ambient temperature). Measurement times were normalised to the first sample injection analysed after the addition of PL^{pro} to the Substrate Mixture (t = 0 s), by which time low levels of substrate hydrolysis were manifest. The inert *N*-acetylated hydrolysis products were used as internal standard to quantify hydrolysis; SPE-MS assay results are a mean of independent triplicates (N = 3; mean ± SD).

The combined results clearly indicate that the primary sequence of oligopeptides affects SARS-CoV-2 PL^{pro} catalysis and that apparently minor changes in their sequence, which may be distal

to the LXGG motif, can have a pronounced effect on catalysis. It thus appears that substrate binding to both the S and S′ sites of PL^{pro} has potential to modulate catalysis.

Substrate competition studies. To investigate whether N^{ε}-lysine-branched oligopeptides or the reported pp1a/1ab-derived linear oligopeptide nsp2/3$_{808\text{-}827}$ (**2**) are more efficient substrates of SARS-CoV-2 PL^{pro}, we attempted to determine k_{cat}/K_m-values. These efforts were, however, unsuccessful due to technical limitations of the SPE-MS assay, *i.e.*, high peptide concentrations saturated the sensor of the mass spectrometer impeding the quantification of PL^{pro} catalysis. Thus, direct substrate competition studies were performed as an alternative, using the linear pp1a/1ab-derived oligopeptide nsp2/3$_{808\text{-}827}$ (**2**) and the two lysine-branched oligopeptide PL^{pro} substrates IRF3$_{189\text{-}197}$-ISG15 (**4**) or IRF3$_{189\text{-}197}$-Ub (**15**). Note that nsp2/3$_{808\text{-}827}$ was employed as a pp1a/1ab-derived linear substrate because it was a more efficient substrate compared to linear peptides based on the nsp1/2 or nsp3/4 cleavage site,[50] in accord with studies that employed a LCMS assay to investigate the substrate preference of SARS-CoV-2 PL^{pro}.[62]

Equimolar amounts of nsp2/3$_{808\text{-}827}$ (**2**) and either IRF3$_{189\text{-}197}$-ISG15 (**4**) or IRF3$_{189\text{-}197}$-Ub (**15**) were incubated with isolated recombinant SARS-CoV-2 PL^{pro} in the same reaction vessel, together with the corresponding inert *N*-acetylated hydrolysis product peptides of **2** and **4** or **15** (as internal standards). PL^{pro}-catalysed peptide hydrolysis was monitored using SPE-MS, which was suitable for these experiments provided that substrates and products have different masses.[63] The combined results reveal that PL^{pro}-catalysed hydrolysis of **2** and **4**/**15** does not affect each other under the tested conditions, *i.e.* the PL^{pro}-catalysed hydrolysis of **2** proceeds as efficiently in the presence of **4** or **15** as in its absence, and vice versa (**Figure 7.5**). This observation may indicate that, under the tested conditions, the initial binding of the peptides to PL^{pro} is not rate-limiting regardless of whether peptides are branched or linear, *i.e.*, the on/off-rates for peptide binding are faster than the peptide hydrolysis rates, potentially because the peptides lack appropriate folding.

The results also reveal that the PL^{pro}-catalysed hydrolysis of nsp2/3$_{808-827}$ (**2**) is less efficient than that of **4** or **15**, *i.e.* ~10% of **2** was observed to be hydrolysed after 5 h whereas ~35-40% of **4** and **15** were observed to be hydrolysed after 5 h which may indicate that the k_{cat}/K_m-values of the N^{ε}-lysine-branched oligopeptides are higher than that of **2**. Notably, it appears that the absolute levels of PL^{pro} catalysis depended on the batch of PL^{pro} and substrate sequence used.

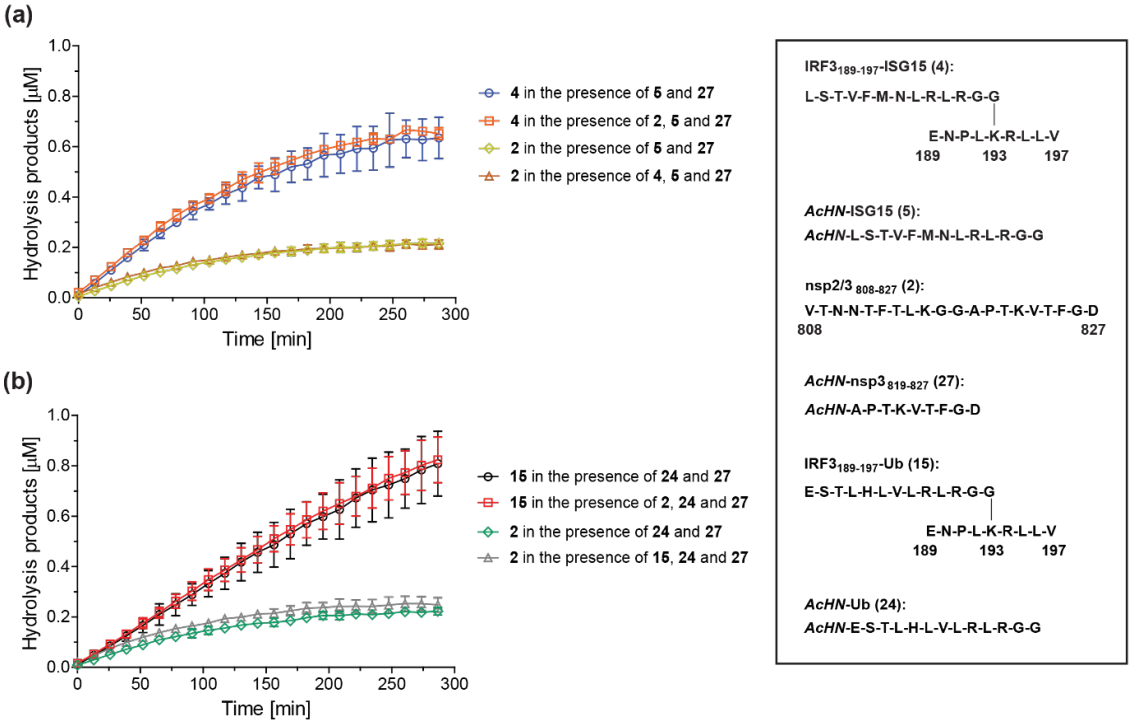


Figure 7.5: N^{ε}-Lysine-branched oligopeptides are more efficient substrates of isolated recombinant PL^{pro} than linear oligopeptides. (**a**) The SARS-CoV-2 PL^{pro}-catalysed hydrolysis of IRF3$_{189-197}$-ISG15 (**4**) is not affected by the presence of nsp2/3$_{808-827}$ (**2**) and vice versa, **4** appears to be a more efficient PL^{pro} substrate than **2**; (**b**) the SARS-CoV-2 PL^{pro}-catalysed hydrolysis of IRF3$_{189-197}$-Ub (**15**) is not affected by the presence of nsp2/3$_{808-827}$ (**2**) and vice versa, **15** appears to be a more efficient PL^{pro} substrate than **2**. Conditions: SARS-CoV-2 PL^{pro} (0.2 µM), substrate peptide(s) (2.0 µM), inert *N*-acetylated standard peptides (0.2 µM, as indicated) in buffer (50 mM Tris, pH 8.0, ambient temperature). Measurement times were normalised to the first sample injection analysed after the addition of PL^{pro} to the Substrate Mixture (t = 0 s), by which time low levels of substrate hydrolysis were manifest. The inert *N*-acetylated hydrolysis products **27** [50] and **5** or **24** were used as internal standard to quantify hydrolysis; SPE-MS assay results are a mean of independent triplicates (N = 3; mean ± SD).

Conformational analysis of substrate binding. To investigate the binding mode of N^{ε}-lysine-branched oligopeptides to PL^{pro}, we computationally modelled nine non-covalently bound PL^{pro}-oligopeptide complexes, including the IRF3$_{189-197}$ oligopeptide N^{ε}-branched with ISG15 (**4**), Ub

(**15**), NEDD8 (**16**), URM1 (**17**), or SUMO1 (**18**), as well as the $TMEM59_{285-291}$ (**21**), $TMEM59_{298-305}$ (**22**), and SARS-CoV-2_$N_{335-341}$ (**23**) oligopeptides, all N^{ε}-branched with Ub. The models were based on a reported PL^{pro} structure (PDB 6WX4),[64] using two of the reported AutoDock CrankPep (ADCP)[65]-docked poses of the $nsp2/3_{808-827}$ peptide **2** [60] as templates for modelling the N- and C-terminal fragments (**Figures S7.1-3** in **Appendix for Chapter 7**). For **17**, we modelled both the neutral (^{N}) and positive charged ($^{+}$) states of the P3 histidine imidazole group, as PL^{pro} has a reported preference for positively charged residues at this position.[64] The resultant PL^{pro}-oligopeptide complexes were subjected to 3 × 200 ns molecular dynamics (MD) simulations, and the stability of the complexes was assessed by calculating the root mean square deviations (RMSDs) and root mean square fluctuations (RMSFs) (**Figures S7.4-13**).

Analysis of the backbone RMSFs of the PL^{pro}-bound N^{ε}-lysine-branched oligopeptides reveals that the P6-P1 residues of the UBL-derived peptide fragment are stable, with values <3 Å (**Figure S7.10**); an exception was $IRF3_{189-197}$-URM1 (**17**) with a neutral P3 histidine imidazole, an observation suggesting that the protonation state of this group may affect complex stability. RMSF analysis of the oligopeptide sidechains reveals that the conformation of the P4 residue is rigid relative to its adjacent P5 and P3 residues (**Figure S7.11**), likely reflecting the conserved nature of the LXGG motif leucine residue which binds to the PL^{pro} S4 site. Notably, the conformation of the P4 glutamine residue of the $IRF3_{189-197}$-SUMO1 oligopeptide (**18**), which is not a substrate of isolated recombinant PL^{pro} and which lacks the LXGG motif (**Figure 7.3a**), is also rigid, suggesting that complex stability itself may not necessarily indicate productive substrate recognition.

Conserved hydrogen bonding interactions appear to stabilise binding of the oligopeptide backbone to the PL^{pro} S4-S1 sites (**Figures 7.6, S7.14-15**), reminiscent of the interactions observed in the modelled PL^{pro}-$nsp2/3_{808-827}$ peptide **2** complex.[60] Hydrogen bonding interactions involving residues of the UBL-derived peptide fragment N-terminal to P4-P1 may also contribute to binding, *e.g.* with Thr75 and Gln174; however, they appear to be less conserved than those

with the P4-P1 LXGG motif (**Table S7.1**; **Figures 7.6, S7.14-15**), consistent with previous observations on the modelled interactions of pp1a/1ab-derived linear oligopeptides with PL^{pro}.[60] Notably, the IRF3$_{189-197}$-SUMO1 oligopeptide (**18**) can apparently engage in similar interactions with PL^{pro} as the other modelled *N*$^{\varepsilon}$-lysine-branched oligopeptides; however, it appears that its P4 glutamine sidechain binds to the S4 site less efficiently than a leucine sidechain (**Figures S7.16-20**).

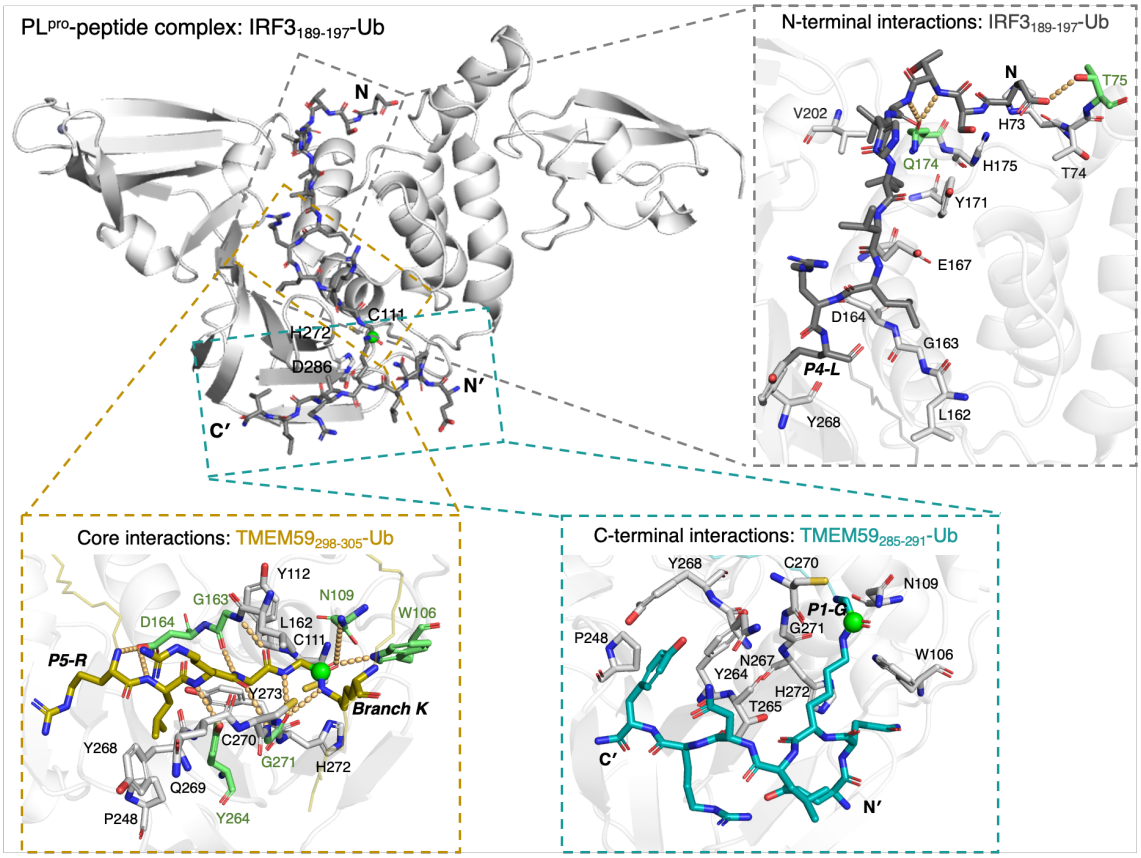


Figure 7.6: Modelled interactions of PL^{pro} with lysine-branched oligopeptides. Views of clustering-derived representative PL^{pro}-oligopeptide complex structures exemplifying the overall binding mode of the oligopeptides and their interactions with PL^{pro} N-terminal and C-terminal to the S4-S1 sites. Peptides are coloured as in **Figure 7.3**, with the P1 scissile amide carbons as green spheres; the N-terminus of the UBL-derived fragment and the N- and C-termini of the S′ binding fragment are labelled N, N′, and C′, respectively. PL^{pro} residues interacting via hydrogen bonding with the oligopeptide are in lime (the calculated occurrence is ≥25%; orange dashes; not necessarily present in the frame). PL^{pro} residues within 4 Å of the peptide residues, that were calculated to contribute ≥0.5 kcal mol^{-1} binding energy, are in white.

In all the modelled PL^{pro}-substrate complexes, the residues C-terminal to the substrate LXGG motif manifest high conformational flexibility, with backbone RMSFs exceeding 4 Å (**Figure S7.12**). The V-shaped RMSF plots indicate that the peptide backbone is flexible at both its N- and C-termini (**Figure S7.12**). Nonetheless, transient hydrogen bonding and dispersion interactions were observed between the substrates and *e.g.* Asn109 of PL^{pro} (**Table S7.1**), as well as with residues around Tyr268 which form a flexible β-hairpin loop (Gly266-Gly271), *i.e.* the blocking loop 2 (BL2), that controls substrate access to the active site.[66-69] In the *apo* form of PL^{pro}, the BL2 loop is present in an open conformation; substrate binding likely triggers an induced fit mechanism resulting in closure of the BL2 loop, in a manner helping to orient the C-terminus of the UBL productively towards the active site.[69] The conformational changes in the BL2 loop associated with substrate binding to PL^{pro} are conserved in, at least, some human DUBs, suggesting that both viral and human DUBs employ similar mechanisms to regulate substrate selectivity.[70, 71] Note that SARS-CoV and SARS-CoV-2 PL^{pro} inhibitors can also bind proximate to BL2 and alter or stabilise its conformation,[38, 66, 69, 72-75] as precedented by the inhibitor-induced conformational changes in BL2 of human DUBs, such as USP7 and USP14.[76]

Our previous modelling studies have shown that the BL2 conformation is stable in its closed form in the PL^{pro}-nsp2/$3_{808-827}$ (**2**) substrate complex, but alternates between open and closed forms with *apo* PL^{pro}, or when the active site is occupied by a peptide that is not efficiently hydrolysed.[60] We analysed BL2 dynamics in all the nine modelled PL^{pro}-substrate complexes by backbone RMSF of the most flexible residue in BL2 (*i.e.* Tyr268) (**Figures 7.7, S7.21-S22**) and the Tyr268-Pro248 Cα-Cα distance (**Figures S7.23-24**); in the latter case, a distance of >12 Å was considered to be indicative of an open BL2 conformation. The results reveal that BL2 remains stably closed in the PL^{pro}-$IRF3_{189-197}$-ISG15 (**4**) and PL^{pro}-$IRF3_{189-197}$-Ub (**15**) complexes (**Figure 7.7**), which are the most efficient substrates among the modelled peptides (**Figure 7.3**). By contrast, BL2 is flexible in the modelled PL^{pro} complex with the relatively inefficient substrate $IRF3_{189-197}$-URM1 (**17**), regardless of the charge state of the P3 histidine imidazole group. Similarly, in the modelled PL^{pro}-

IRF3$_{189\text{-}197}$-SUMO1 (**18**) complex, BL2 frequently opens and closes (**Figure S7.24**); note that **18** is not a PL^{pro} substrate (**Figure 7.3**). The latter observation is likely a result of the P4 leucine to glutamine substitution in **18**; consistent with this proposal, previous reports have shown that the binding of hydrophobic sidechains to the PL^{pro} S4 site induces BL2 closure.[64, 69, 77, 78]

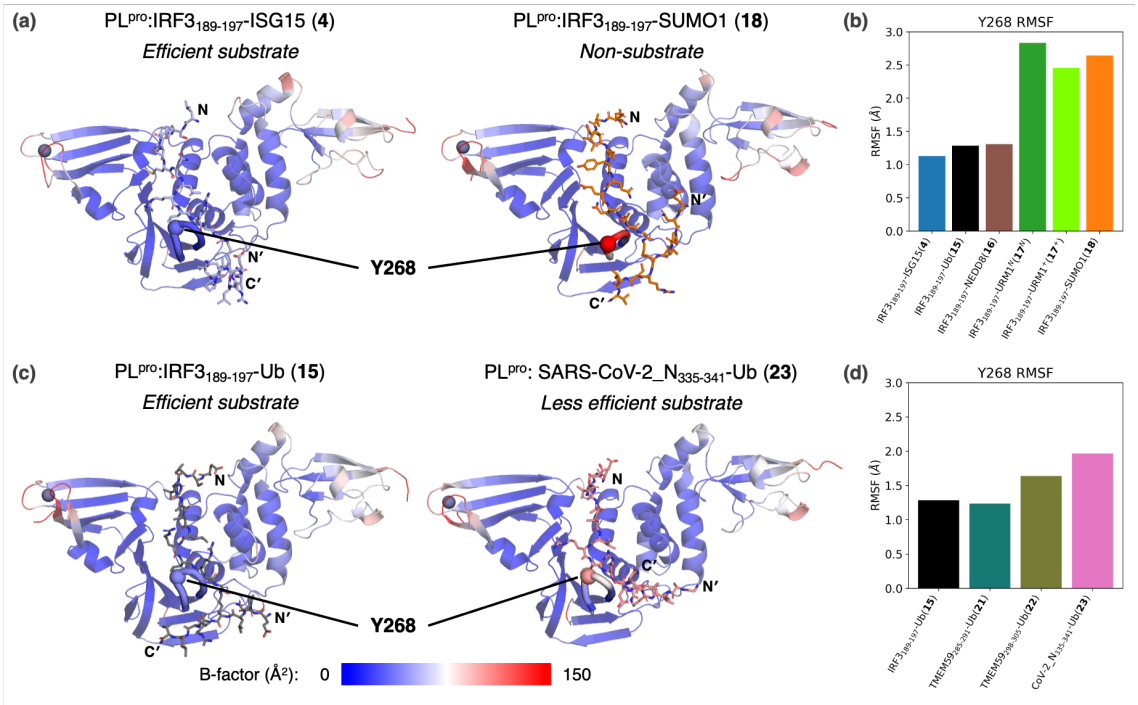


Figure 7.7: BL2 flexibility in PL^{pro}-oligopeptide complexes is indicative of poor hydrolysis efficiency. Views of clustering-derived representative structures of (**a**) the PL^{pro}-IRF3$_{189\text{-}197}$-ISG15 (**4**) and PL^{pro}-IRF3$_{189\text{-}197}$-SUMO1 (**18**) complexes and (**c**) the PL^{pro}-IRF3$_{189\text{-}197}$-Ub (**15**) and PL^{pro}-SARS-CoV-2_N$_{335\text{-}341}$-Ub (**23**) complexes. PL^{pro} residues are coloured based on MD-derived backbone B-factors; red represents high B-factors, which is indicative of relatively high residue flexibility. The bar plots show comparisons of Tyr268 backbone RMSF for PL^{pro} in complex with (**b**) oligopeptides **4**, **15**, **16**, **17**, and **18**, and (**d**) oligopeptides **15**, **21**, **22**, and **23**. Residues forming BL2 (Gly266-Gly271) are shown as a thickened loop, with Tyr268 Cα being shown as a sphere. The N-terminus of the UBL-derived fragment and the N- and C-termini of the S′ binding fragment are labelled N, N′, and C′, respectively.

In the modelled PL^{pro} complexes with the N^{ε}-lysine-branched oligopeptides that mimic post-translational protein ubiquitinylation (*i.e.* **15**, **21**, **22**, and **23**), BL2 remains rigidly closed for the relatively more efficient substrates (*i.e.* **15** and **21**), as observed for the pp1a/1ab-derived nsp2/3$_{808\text{-}827}$ peptide **2**,[60] but not for the less efficient substrates (*i.e.* **22** and **23**) (**Figure 7.7, S7.23-24**). These observations are in accord with the results on the stability of the complexes of PL^{pro}

and N^{ε}-lysine-branched oligopeptides mimicking post-translational protein ISG15ylation. In particular, it appears that hydrophobic residues C-terminal to the P1′ lysine residue can affect BL2 flexibility (**Figures S7.9, 19-20, 25-27**), in accord with the experimentally observed preference of PL^{pro} for hydrophobic residues C-terminal to the substrate P1′ lysine residue involved in the isopeptide amide bond (**Figure 7.4**).

The combined modelling and experimental results reveal the importance of the conformational dynamics of the BL2 loop; it not only enables productive substrate binding in the active site, but also governs substrate selectivity. Importantly, the BL2 flexibility and opening tendency observed in the MD simulations correlate with the experimentally observed substrate selectivity, *i.e.*, conformational BL2 flexibility in the PL^{pro}-substrate complex appears to result in inefficient catalysis (**Figure 7.7**). Although substrate peptide dissociation was not observed over the course of the MD simulations, the observations suggest that a sustained closed state of BL2 may be necessary for initiating productive catalysis, potentially because BL2 is proximal to His272 which is part of the catalytic triad.[67] Thus, the observation that the PL^{pro}-catalysed hydrolysis of the linear pp1a/1ab-derived nsp2/$3_{808\text{-}827}$ oligopeptide (**2**) and the N^{ε}-lysine-branched oligopeptide substrates do not appear to affect each other when incubated in the same reaction vessel (**Figure 7.5**), implies that inefficient substrate binding to PL^{pro} results in substrate dissociation being faster than substrate hydrolysis.

Discussion

SARS-CoV-2 PL^{pro} catalyses the hydrolysis of peptide amide bonds in pp1a/1ab that are C-terminal to three LXGG motifs to release functional nsps1-3.[1, 2] By contrast with M^{pro}, PL^{pro} is also a deubiquitinase (DUB) catalysing the hydrolysis of isopeptide amide bonds of human proteins that are post-translationally modified with Ub/UBLs, in particular ISG15.[38, 39] Thus, the DUB substrate preference of PL^{pro} is apparently similar to that of the human ubiquitin specific

peptidase 18 (USP18), a DUB which also employs a nucleophilic cysteine and which selectively catalyses protein deISG15ylations.[34, 44] It has been proposed that the DUB activity of PL^{pro} modulates virulence via catalysing the deISG15ylation of ISG15ylated IRF3.[38] However, to our knowledge, the DUB activity of isolated recombinant PL^{pro} has not yet been validated with oligopeptide substrates *in vitro*.

We developed SPE-MS assays that directly monitor the hydrolysis of the isopeptide amide bond of synthetic N^{ε}-lysine-branched oligopeptides, which mimic post-translational protein modification by Ub/UBLs, catalysed by isolated recombinant SARS-CoV-2 PL^{pro}, extending the scope of our reported PL^{pro} SPE-MS assays with linear oligopeptides as substrates to N^{ε}-lysine-branched oligopeptide substrates.[50] Due to variations in the catalytic efficiency of different PL^{pro} batches, we thus determined the rank order of substrate preference by quantifying product formation relative to internal standards.

Importantly, our results with N^{ε}-lysine-branched oligopeptides support cellular studies concerning the DUB activities of PL^{pro}, including its reported preference for catalysing the hydrolysis of isopeptide amide bonds C-terminal to the LRGG motif of ISG15 rather than Ub/Ub_2 and, in particular, NEDD8.[38, 39] The rank order observed for PL^{pro}-catalysed hydrolysis of Lys193-branched IRF3-derived oligopeptides mimicking post-translational modification by UBLs was: $IRF3_{189\text{-}197}$-ISG15 (**4**) > $IRF3_{189\text{-}197}$-Ub (**15**) > $IRF3_{189\text{-}197}$-NEDD8 (**16**) > $IRF3_{189\text{-}197}$-URM1 (**17**) > $IRF3_{189\text{-}197}$-SUMO1 (**18**), the latter of which was not a substrate of PL^{pro} (**Figure 7.3a**). The results thus imply that the substrate preference of PL^{pro} not only depends on the fold of the substrate, interactions of the substrate at the PL^{pro} S1-S4 sites,[41, 64] and on allosteric interactions with PL^{pro},[41-43] but also on the sequence identity of the substrate which binds to the active site beyond S1-S4. This proposal is further supported by the pronounced effect of amino acid substitutions in the N^{ε}-lysine-branched oligopeptide substrates on PL^{pro} catalysis (**Figure 7.4**), as well as by reported work on the hydrolysis of linear oligopeptides catalysed by SARS-CoV-2 PL^{pro} and PL^{pro}s of other coronaviruses.[79, 80]

The combined results show that the primary sequence of both the peptide fragment derived from the C-terminus of the UBL and that of the post-translationally modified human protein affects the efficiency of PL^{pro} catalysis, the latter potentially by binding to the S′ sites of PL^{pro}, possibly in a substrate context-dependent manner. Hence, SPE-MS assays have potential to predict the efficacy with which PL^{pro} catalyses the hydrolysis of post-translationally modified (human) proteins, although factors other than the primary sequence also contribute to catalysis efficiency (see above). It will be useful to perform studies with the N^{ε}-lysine-branched oligopeptides and isolated recombinant SARS-CoV PL^{pro} and MERS-CoV PL^{pro}, as well as reported variants of SARS-CoV-2 PL^{pro}, to inform on different substrate preferences and on how these differences manifest in altered virulence. Similarly, the substrate preferences of viral proteases other than coronavirus PL^{pro}s, that have been reported to catalyse protein deISG15ylation and/or deubiquitinylation,[81-83] should be examined.

The results indicate that PL^{pro} may catalyse the deISG15ylation of IRF3 regardless of the ISG15ylation site (**Figure 7.2c**); studies with SARS-CoV-2 infected human cells are required to investigate the significance of this observation in a cellular context. Note that the ISG15ylation of IRF3 at the N^{ε}-amino group of Lys193, Lys366, and, by implication, Lys360 may compete with other post-translational modifications, *e.g.* ubiquitinylation at Lys193,[53] and/or acetylation[84] and methylation[85] at Lys366, in a manner potentially affecting PL^{pro} catalysis. SPE-MS assay results with N^{ε}-lysine-branched oligopeptides mimicking the ISG15ylation of 4EHP, the ISG15ylation of which has been proposed to control translation during immune response,[58] suggest that ISG15ylated 4EHP may be a substrate of SARS-CoV-2 PL^{pro} in cells (**Figure 7.3b**). The relevance of the PL^{pro}-catalysed deISG15ylation of ISG15ylated 4EHP for virulence should thus also be probed in cellular studies.

It is likely that PL^{pro} not only catalyses the deISG15ylation of IRF3 and 4EHP (**Figure 7.3**), but also of other human proteins.[86] Further substrate selectivity studies are, however, currently hampered by the lack of information on the exact ISG15ylation sites of human proteins.[34, 54-57]

Note that evidence suggests that residues other than lysine residues may be ISG15ylated, *e.g.* cysteine residues;[87, 88] thus, the ability of SARS-CoV-2 PL^{pro} to catalyse the deISG15ylation of other residues than lysine should be subject of future work. Considering that protein ubiquitinylation is apparently more widespread than ISG15ylation, it is possible that human ubiquitinylated proteins may be identified which are more efficient substrates than ISG15ylated IRF3 *in vitro* and in cells.

Viral proteins can also be ISG15ylated by host proteins, in a manner which may impede their function, *e.g.*, during replication.[89-91] PL^{pro} may catalyse the deISG15ylation of ISG15ylated viral proteins to counteract the host innate immune response. Although, to our knowledge, little information is currently available on SARS-CoV-2 protein ISG15ylation, multiple ubiquitinylation sites have been identified in SARS-CoV-2 proteins in proteomic MS studies.[61] Our work provides MS evidence that isolated recombinant PL^{pro} catalyses the hydrolysis of the isopeptide amide bond of an N^{ε}-lysine-branched oligopeptide based on a reported ubiquitinylation site of the SARS-CoV-2 N protein, *i.e.* SARS-CoV-2_$N_{335-341}$-Ub **23** (**Figure 7.3c**). It is thus possible that PL^{pro} catalyses the deubiquitinylation of at least some of the (poly)ubiquitinylated lysine residues of viral proteins to *e.g.* regulate protein function or to prevent protein degradation via the host proteasome, as (poly)ubiquitinylation can be a signal for proteasomal degradation and the removal of the (poly)ubiquitin may hence stabilise the viral protein; conversely, PL^{pro} has potential to also catalyse the deISG15ylation of ISG15ylated viral proteins.

In general, it appears that viral proteases may have evolved to catalyse deISG15ylation reactions as a strategy to evade the host innate immune response, at least during early stages of infection.[91-93] Interestingly, viral proteases catalyse deISG15ylation reactions at different sites*, e.g.* the leader protease of foot-and-mouth-disease virus (FMDV) is reported to catalyse the hydrolysis of ISG15 N-terminal to diglycine of the LRGG motif,[81] opposed to SARS-CoV-2 PL^{pro} which catalyses the hydrolysis of ISG15 C-terminal to diglycine of the LRGG motif. Viral proteases have also been reported to catalyse the direct hydrolysis of the main chain of interferon regulatory factors,[94-96]

including that of IRF3 by SARS-CoV-2 PL^{pro},[17] further highlighting the importance of modulating interferon regulatory factors such as IRF3 during viral infections.

At least in most cases, the identity of the substrate residues binding to the S sites affects PL^{pro} catalysis to a relatively greater extent than the identity of the substrate residues binding to the S′ sites (**Figure 7.2**). This observation may reflect the different substrate selectivities of PL^{pro} and M^{pro}. M^{pro} apparently requires the presence of specific residues (*i.e.*, S/A/N) at the S1′ site for efficient catalysis, though its substrate requirements for S2′-S4′ appear to be less stringent.[97] Defining the full substrate scope of M^{pro} is, however, the subject of ongoing investigations.[98, 99] Notably, NEMO, the M^{pro}-catalysed hydrolysis of which has been reported to induce the death of brain endothelial cells,[15, 16] has a valine residue at the P1′ position, indicating that the presence of (L/F/V)Q(S/A/N) motifs in human proteins may not be sufficient to predict the efficiency of M^{pro} catalysis.

Eleven sites in pp1a/1ab fulfil the apparently somewhat narrower substrate requirements for M^{pro}, whereas only three sites in pp1a/1ab fulfil the substrate requirements for the apparently more promiscuous PL^{pro}.[1, 2] Human proteases that accept (L/F/V)Q(S/A/N) motifs as substrates are currently unknown, suggesting that M^{pro} may have evolved to exclusively catalyse the release of those nsps from pp1a/1ab that are directly involved in replication and transcription (*e.g.* nsp12, which has a RNA-dependent RNA polymerase domain, and the nsp13 helicase), under strict spatiotemporal control. By contrast, in principle, human DUBs could catalyse the hydrolysis of pp1a/1ab at three sites, releasing *inter alia* nsp1 and nsp2 which have been proposed to help the virus evade the host immune system,[100] and thus compete with PL^{pro} for viral substrates. The apparent lower substrate specificity of PL^{pro} may reflect its functions in enabling the virus to counteract the host immune system, with respect to catalysing both the release of nsp1 and nsp2 and the deISG15ylation of human and, potentially, viral proteins, a process which might not require strict spatiotemporal control and thus tolerates reduced substrate specificity.

The ability of PL^{pro}, but not M^{pro} (as far as is known), to catalyse hydrolysis of isopeptide as well as peptide amide bonds is striking. The comparison of substrate and non-substrate binding modes for the two proteases is thus of interest, because the precise nature of the dynamic interactions that regulate the different substrate selectivities of PL^{pro} and M^{pro} is not defined. Reported PL^{pro} structures reveal that substrate binding can induce conformational changes of flexible regions proximate to the active site, including the PL^{pro} blocking loop 2 (BL2) which is a key factor in regulating the accessibility of potential substrates to the active site.[69] The role of the PL^{pro} BL2 in regulating substrate recognition is precedented by BL2 in, at least some, human DUBs, including *e.g.* USP7 and USP14;[70, 71] however, in human DUBs, post-translational modification of BL2 residues may further alter catalysis.[101] It appears that the PL^{pro} substrate selectivity is not only a result of the affinity of the substrate to bind to S sites and, to a lesser extent, S′ sites, but also of the ability of a potential substrate to stabilise the closed conformation of BL2 so enabling a catalytically productive enzyme-substrate conformation which may be achieved via interactions with PL^{pro} that may be relatively remote from the active site, including by specific interactions of hydrophobic substrate residues with the S′ sites (**Figure 7.6**).

Movement of active site bordering loops to bind, enclose, and/or orientate substrates (and reject non-substrates) are common in enzyme catalysis, *e.g.* substrate binding can also induce conformational changes in flexible regions of M^{pro} that directly contribute to shape the S sites, in particular of an α-helix (Thr45-Leu50), which contributes to form the S2 site, and of a loop (Asp187-Gln192), which contributes to form the S4 site.[97, 102-105] The presence of a BL2-type loop in PL^{pro}, but not M^{pro}, is interesting. Whether or not the presence of this loop affects the ability of PL^{pro}, but not M^{pro}, to accept different types of substrates (peptide and isopeptide amide bonds), whilst manifesting an apparently narrower selectivity with respect to pp1a/1ab hydrolysis, is unclear. In this regard, it will be of interest to investigate how the substrate selectivities of PL^{pro} homologues change during the course of viral evolution, both with respect to pp1a/1ab and proteins in different host organisms, *e.g.* SARS-CoV-2 PL^{pro} prefers ISG15 as a substrate, whereas

SARS-CoV PL^{pro} prefers diubiquitin.[38, 39] Notably, the efficiency and selectivity of MERS-CoV PL^{pro} catalysis has been proposed to be distinct from that of SARS-CoV and SARS-CoV-2 PL^{pro}, in part as a result of amino acid variations in the BL2 loop.[46, 73]

In summary, SPE-MS assays have been used to investigate post-translational protein modifications and catalysis of nucleophilic cysteine enzymes, including isolated recombinant SARS-CoV-2 PL^{pro} and M^{pro}.[9, 47, 50, 106-111] SPE-MS assays employing synthetic N^{ε}-lysine-branched oligopeptides as substrates of isolated recombinant PL^{pro} have enabled studies on the substrate scope and substrate selectivity of PL^{pro} and thus helped to characterise the DUB activity of PL^{pro}. The combination of analogous MS-based assays with synthetic N^{ε}-lysine-branched oligopeptides as substrates of isolated recombinant human DUBs, together with proteomic studies, will enable studies on the substrate preferences of human DUBs, such as USP18 which also catalyses protein deISG15ylation, and which is a current medicinal chemistry target.[34, 44]

Materials and Methods

Production and purification of isolated recombinant SARS-CoV-2 PL^{pro}. The PL^{pro} domain of the SARS-CoV-2 nsp3 (E746-T1063) was produced using *E. coli* Lemo21(DE3) cells and purified as reported previously.[50]

Peptide synthesis. Linear oligopeptides were prepared by solid phase peptide synthesis (SPPS) using a Liberty Blue peptide synthesiser (CEM Microwave Technology Ltd.), as reported for the synthesis of the oligopeptides **1**-**3**.[50]

The N^{ε}-lysine-branched oligopeptides mimicking post-translationally modified proteins were synthesised by microwave-assisted SPPS using the Fmoc-protection strategy from the C- to N-terminus on Rink Amide MBHA resin (AGTC Bioproducts Ltd.; loading: 0.6-0.8 mmol/g) similar to reported procedures.[50] Initially, a linear oligopeptide was synthesised; *N*-Fmoc lysine with a 4-methyltrityl (Mtt) protected N^{ε}-amine was used at the site of the lysine branching, and an amino acid with a Boc-protected N^{α}-amine was used at the N-terminus. Branching was introduced via selective deprotection of the lysine N^{ε}-Mtt group of the resin-bound peptides using 1%$_{v/v}$ trifluoroacetic acid (TFA) and 2%$_{v/v}$ triisopropylsilane (TIPS) in dichloromethane at ambient temperature while shaking (300 rpm) (5 cycles; after each cycle, the deprotection mixture was removed and the resin was washed with dichloromethane). Following Mtt deprotection, the UBL-derived peptide fragment was synthesised by SPPS from the C- to N-terminus starting with the free N^{ε}-amino group of the resin-bound peptides at the C-terminus.

After completion of the synthesis, the resin-bound peptides were washed with dichloromethane and subsequently cleaved from the resin and simultaneously deprotected using a mixture of trifluoroacetic acid, triisopropylsilane, 1,3-dimethoxybenzene, and water (92.5/2.5/2.5/2.5%$_{v/v}$, respectively). Solids were separated; the remaining clear solution was diluted with diethyl ether (45 mL/0.1 mmol resin). After incubation for 30 min at 0 °C, the mixture was centrifuged for 10 min using a Beckman Coulter Allegra X-30R centrifuge equipped with a SX4400 rotor (4500

rpm) and the supernatant discarded. The solid residue was dissolved in a water/acetonitrile mixture, frozen using liquid N_2, and then lyophilised. The dried crude product was dissolved in a water/acetonitrile mixture, filtered, and purified using a semi-preparative HPLC machine (Shimadzu UK Ltd.) equipped with a reverse phase column (Gemini 00G-4454-U0-AX; phase: NX-C18). A linear gradient (typically 2−47%$_{v/v}$ over 38 min) of acetonitrile in milli-Q grade water (each containing 0.1%$_{v/v}$ trifluoroacetic acid) was used as eluent. Fractions were analysed by SPE-MS and those containing the pure peptide were combined and lyophilised.

SPE-MS assays. PL^{pro} assays for turnover and competition experiments were performed in 96-well polypropylene assay plates (Greiner), either with a 1.0 or 0.5 mL final reaction volume, using isolated recombinant SARS-CoV-2 PL^{pro} (0.2 μM), substrate peptide(s) (2.0 μM), inert *N*-acetylated standard peptides (0.2 μM, as indicated in the individual experiments) in buffer (50 mM Tris, pH 8.0) at ambient temperature. PL^{pro} catalysis was directly monitored using SPE-MS. The RapidFire RF 365 high-throughput sampling robot used was programmed to aspirate samples from the reaction mixture at the time intervals indicated in the individual experiments.

MS-analyses were performed using a RapidFire RF 365 high-throughput sampling robot (Agilent) attached to an iFunnel Agilent 6550 accurate mass quadrupole time-of-flight (Q-TOF) mass spectrometer operated in the positive ionization mode. Assay samples were aspirated under vacuum for 0.6 s and loaded onto a C4 solid phase extraction (SPE) cartridge. After loading, the C4 SPE cartridge was washed with 0.1%$_{v/v}$ aqueous formic acid to remove non-volatile buffer salts (5.5 s, 1.5 mL/min). The peptide was eluted from the SPE cartridge with 0.1%$_{v/v}$ aqueous formic acid in 85/15$_{v/v}$ acetonitrile/water into the mass spectrometer (5.5 s, 1.25 mL/min) and the SPE cartridge re-equilibrated with 0.1%$_{v/v}$ aqueous formic acid (0.5 s, 1.25 mL/min). The mass spectrometer parameters were: capillary voltage: 4000 V; nozzle voltage: 1000 V; fragmentor voltage: 365 V; gas temperature: 280 °C; gas flow: 13 L/min; sheath gas temperature: 350 °C; sheath gas flow: 12 L/min.

For data analysis and to quantify product formation, the charge states of both the C-terminal and N-terminal product peptides and the corresponding *N*-acetylated C-terminal and N-terminal product peptides (*i.e.* **5**, **12**, and **24**), which were used as internal standards, were used to extract ion chromatogram data (*m*/*z* +1 for **12** and the corresponding product peptide; *m*/*z* +2 for **5** and **24**, as well as for the corresponding product peptides); peak areas were integrated using the RapidFire Integrator software (Agilent). Data were exported into Microsoft Excel and used to calculate the product peptide concentrations using the equation: peptide concentration = 0.2 μM × (integral C- or N-terminal product peptide) / (integral *N*-acetylated C- or N-terminal product peptide).

Preparation of PL^{pro}-oligopeptide models. A reported SARS-CoV-2 PL^{pro} structure (PDB 6WX4)[64] was prepared for modelling as described (**Table S7.2**).[60] Based on reported quantum mechanics/molecular mechanics-umbrella sampling (QM/MM-US) calculations on proton transfer processes in the PL^{pro} catalytic triad,[60] Cys111 was modelled in its deprotonated form, His272 as doubly protonated, and Asp286 as deprotonated.

The N^{ε}-lysine-branched oligopeptides complexed with PL^{pro} were constructed using reported AutoDock CrankPep (ADCP)[65]-docked conformations of the linear nsp2/3-derived peptide **2**.[60] The residues N-terminal to the scissile amide in the top ranked pose (d2_01, reported nomenclature),[60] which was successful in placing P4-P1 LKGG in the respective S4-S1 subsites, was used to build the N-terminal ubiquitin (Ub) or Ub-like modifier (UBL) derived fragment of the N^{ε}-lysine-branched oligopeptides. The C-terminal fragment of the N^{ε}-lysine-branched oligopeptides was built based on the 48th-ranked nsp2/3 pose (d2_48), which did not pass through S4-S1, but in which the P4′ lysine sidechain N^{ε} amine was proximate to Cys111 (**Figure S7.1**). The lysine sidechain branching was modelled by linking the C^{ε} atom of the d2_48 P4′ lysine sidechain to the carboxamide N atom of the 2_01 P1′ alanine. To prepare each of the eight branched peptides in $IRF3_{189\text{-}197}$-(ISG15/Ub/NEDD8/URM1/SUMO1) and ($TMEM59_{285\text{-}}$

$_{291}$/TMEM59$_{298\text{-}305}$/SARS-CoV-2_N$_{335\text{-}341}$)-Ub (**Figure 7.3**), the two nsp2/3 peptide poses were modified to the lengths of the N- and C-terminal fragments. The residues were modified to the target sequences using the mutagenesis tool of PyMOL (open source, v. 2.3.0),[112] selecting the least sterically clashing backbone-dependent rotamer in each case.[113] The N-termini of both fragments were uncapped, while the C-terminus of the IRF3$_{189\text{-}197}$/TMEM59$_{285\text{-}291}$/TMEM59$_{298\text{-}305}$/SARS-CoV-2_N$_{335\text{-}341}$ fragment was NH_2-capped. Peptide histidine residues were modelled in their neutral state, with N^{ε} protonated (HIE in AMBER nomenclature),[114] except for IRF3$_{189\text{-}197}$-URM1 where both the neutral HIE and the positively charged, doubly protonated (HIP) states of the P3 histidine residue were considered, denoted as URM1^{N} and URM1^{+} respectively.

Parametrisation and molecular dynamics (MD). MD simulations of the PL^{pro}-oligopeptide complexes were performed in GROMACS (v. 2019.2, 2020.4),[115] using the AMBERFF99SB-ILDN/TIP3P forcefield.[114, 116] Non-bonded Zn parameters were employed.[117] For the non-standard *N*$^{\varepsilon}$-branched lysine residue, *N*-butyl acetamide (**Figure S7.28**) was used as a model for parametrization with AMBER atom types and RESP charges (Gaussian16 and antechamber in Amber18).[118-122] The new residue, designated as "LYC", was incorporated into GROMACS (**Figure S7.29**), with its N^{ε} atom bonded to the P1 Gly C atom. Amide improper dihedral parameters centred at the GLY_C and LYC_NZ atoms were included to maintain amide group planarity.

Each of the nine PL^{pro}-oligopeptide complexes (eight unique sequences, with URM1^{N} and URM1^{+} for two possible charge states of the P3 histidine group in **17**) was centred in a rhombic dodecahedral box with at least 1.0 nm separation from box edges, solvated, neutralised with sodium/chloride ions (110,788-110,847 atoms in total), and minimised until the maximum force was below 1000 kJ mol^{-1} nm^{-1}. From the minimised system, three replicas were initiated using random velocities at 298.15 K, subjected to 200 ps (1 fs step) restrained NVT equilibration at 298.15 K, followed by 200 ps (1 fs step) NPT equilibration at 298.15 K and 1.0 bar. The

equilibrated complexes were subjected to 200 ns production MD (2 fs step). A velocity-rescaling thermostat with a stochastic term (time constant 0.1 ps; protein and non-protein coupled separately)[123] and a Parrinello-Rahman barostat (time constant 2 ps) were used.[124, 125] Long-range electrostatic interactions were calculated by smooth Particle-mesh Ewald (1 nm cut-off).[126, 127] Van der Waals interactions were cut off at 1 nm.

The 3 × 200 ns MD trajectories were fitted based on the PL^{pro} backbone and analysed using GROMACS tools (v 2019.2).[115] A hydrogen bond was defined on observation of a donor-acceptor distance <3.5 Å and hydrogen-donor-acceptor angle <30°. To obtain representative structures, clustering was performed with a 3 Å RMSD cut-off of the peptide backbone, using the gromos algorithm.[128] Per residue decomposition of PL^{pro}-oligopeptide binding energies calculated by molecular mechanics/generalised Born surface area (MM/GBSA)[129-132] was performed on frames in 5 ns intervals, using MMPBSA.py (AMBER18) with an ionic strength of 0.15 M, mbondi2 radii, and igb=5.[122, 133, 134]

References

1. V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2021,** *19*, 155.
2. Thiel, V.; Ivanov, K. A.; Putics, Á.; Hertzig, T.; Schelle, B.*, et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **2003,** *84*, 2305.
3. Li, G.; Hilgenfeld, R.; Whitley, R.; De Clercq, E. Therapeutic strategies for COVID-19: progress and lessons learned. *Nat. Rev. Drug Discov.* **2023,** *22*, 449.
4. Cannalire, R.; Cerchia, C.; Beccari, A. R.; Di Leva, F. S.; Summa, V. Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: state of the art and future opportunities. *J. Med. Chem.* **2022,** *65*, 2716.
5. Ghosh, A. K.; Mishevich, J. L.; Mesecar, A.; Mitsuya, H. Recent drug development and medicinal chemistry approaches for the treatment of SARS-CoV-2 infection and COVID-19. *ChemMedChem* **2022,** *17*, e202200440.
6. Lv, Z.; Cano, K. E.; Jia, L.; Drag, M.; Huang, T. T.*, et al.* Targeting SARS-CoV-2 proteases for COVID-19 antiviral development. *Front. Chem.* **2022,** *9*, 819165.
7. Ton, A.-T.; Pandey, M.; Smith, J. R.; Ban, F.; Fernandez, M.*, et al.* Targeting SARS-CoV-2 papain-like protease in the postvaccine era. *Trends Pharmacol. Sci.* **2022,** *43*, 906.
8. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.*, et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374*, 1586.
9. Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.*, et al.* Discovery of S-217622, a noncovalent oral SARS-CoV-2 3CL protease inhibitor clinical candidate for treating COVID-19. *J. Med. Chem.* **2022,** *65*, 6499.
10. Mielech, A. M.; Kilianski, A.; Baez-Santos, Y. M.; Mesecar, A. D.; Baker, S. C. MERS-CoV papain-like protease has deISGylating and deubiquitinating activities. *Virology* **2014,** *450-451*, 64.
11. Yang, X.; Chen, X.; Bian, G.; Tu, J.; Xing, Y.*, et al.* Proteolytic processing, deubiquitinase and interferon antagonist activities of middle east respiratory syndrome coronavirus papain-like protease. *J. Gen. Virol.* **2014,** *95*, 614.
12. Lindner, H. A.; Lytvyn, V.; Qi, H.; Lachance, P.; Ziomek, E.*, et al.* Selectivity in ISG15 and ubiquitin recognition by the SARS coronavirus papain-like protease. *Arch. Biochem. Biophys.* **2007,** *466*, 8.
13. Ratia, K.; Kilianski, A.; Baez-Santos, Y. M.; Baker, S. C.; Mesecar, A. Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease. *PLoS Pathog.* **2014,** *10*, e1004113.
14. Chen, Z.; Wang, Y.; Ratia, K.; Mesecar, A. D.; Wilkinson, K. D.*, et al.* Proteolytic processing and deubiquitinating activity of papain-like proteases of human coronavirus NL63. *J. Virol.* **2007,** *81*, 6007.
15. Wenzel, J.; Lampe, J.; Müller-Fielitz, H.; Schuster, R.; Zille, M.*, et al.* The SARS-CoV-2 main protease Mpro causes microvascular brain pathology by cleaving NEMO in brain endothelial cells. *Nat. Neurosci.* **2021,** *24*, 1522.
16. Hameedi, M. A.; T. Prates, E.; Garvin, M. R.; Mathews, I. I.; Amos, B. K.*, et al.* Structural and functional characterization of NEMO cleavage by SARS-CoV-2 3CLpro. *Nat. Commun.* **2022,** *13*, 5285.
17. Moustaqil, M.; Ollivier, E.; Chiu, H.-P.; Van Tol, S.; Rudolffi-Soto, P.*, et al.* SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg. Microbes Infect.* **2021,** *10*, 178.
18. Reynolds, N. D.; Aceves, N. M.; Liu, J. L.; Compton, J. R.; Leary, D. H.*, et al.* The SARS-CoV-2 SSHHPS recognized by the papain-like protease. *ACS Infect. Dis.* **2021,** *7*, 1483.
19. Mohamud, Y.; Xue, Y. C.; Liu, H.; Ng, C. S.; Bahreyni, A.*, et al.* The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem. Biophys. Res. Commun.* **2021,** *540*, 75.
20. Meyer, B.; Chiaravalli, J.; Gellenoncourt, S.; Brownridge, P.; Bryne, D. P.*, et al.* Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat. Commun.* **2021,** *12*, 5553.
21. Haas, A. L.; Ahrens, P.; Bright, P. M.; Ankel, H. Interferon induces a 15-kilodalton protein exhibiting marked homology to ubiquitin. *J. Biol. Chem.* **1987,** *262*, 11315.
22. Loeb, K. R.; Haas, A. L. The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J. Biol. Chem.* **1992,** *267*, 7806.
23. Narasimhan, J.; Wang, M.; Fu, Z.; Klein, J. M.; Haas, A. L.*, et al.* Crystal structure of the interferon-induced ubiquitin-like protein ISG15 *J. Biol. Chem.* **2005,** *280*, 27356.

24. Kang, J. A.; Kim, Y. J.; Jeon, Y. J. The diverse repertoire of ISG15: more intricate than initially thought. *Exp. Mol. Med.* **2022,** *54*, 1779.
25. Kumar, S.; Tomooka, Y.; Noda, M. Identification of a set of genes with developmentally down-regulated expression in the mouse brain. *Biochem. Biophys. Res. Commun.* **1992,** *185*, 1155.
26. Kamitani, T.; Kito, K.; Nguyen, H. P.; Yeh, E. T. H. Characterization of NEDD8, a developmentally down-regulated ubiquitin-like protein. *J. Biol. Chem.* **1997,** *272*, 28557.
27. Furukawa, K.; Mizushima, N.; Noda, T.; Ohsumi, Y. A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J. Biol. Chem.* **2000,** *275*, 7462.
28. Van der Veen, A. G.; Schorpp, K.; Schlieker, C.; Buti, L.; Damon, J. R.*, et al.* Role of the ubiquitin-like protein Urm1 as a noncanonical lysine-directed protein modifier. *Proc. Natl. Acad. Sci. USA* **2011,** *108*, 1763.
29. van der Veen, A. G.; Ploegh, H. L. Ubiquitin-like proteins. *Ann. Rev. Biochem.* **2012,** *81*, 323.
30. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **2009,** *458*, 422.
31. Dikic, I.; Schulman, B. A. An expanded lexicon for the ubiquitin code. *Nat. Rev. Mol. Cell Biol.* **2023,** *24*, 273.
32. Glickman, M. H.; Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* **2002,** *82*, 373.
33. Lecker, S. H.; Goldberg, A. L.; Mitch, W. E. Protein degradation by the ubiquitin–proteasome pathway in normal and disease states. *J. Am. Soc. Nephrol.* **2006,** *17*, 1807.
34. Jiménez Fernández, D.; Hess, S.; Knobeloch, K.-P. Strategies to target ISG15 and USP18 toward therapeutic applications. *Front. Chem.* **2020,** *7*, 923.
35. Cappadocia, L.; Lima, C. D. Ubiquitin-like protein conjugation: structures, chemistry, and mechanism. *Chem. Rev.* **2018,** *118*, 889.
36. Lange, S. M.; Armstrong, L. A.; Kulathu, Y. Deubiquitinases: from mechanisms to their inhibition by small molecules. *Mol. Cell* **2022,** *82*, 15.
37. Komander, D.; Clague, M. J.; Urbé, S. Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **2009,** *10*, 550.
38. Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.*, et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020,** *587*, 657.
39. Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.*, et al.* Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* **2020,** *39*, e106275.
40. Lei, J.; Kusov, Y.; Hilgenfeld, R. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antiviral Res.* **2018,** *149*, 58.
41. Patchett, S.; Lv, Z.; Rut, W.; Békés, M.; Drag, M.*, et al.* A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. *Cell Rep.* **2021,** *36*, 109754.
42. van Vliet, V. J. E.; Huynh, N.; Palà, J.; Patel, A.; Singer, A.*, et al.* Ubiquitin variants potently inhibit SARS-CoV-2 PLpro and viral replication via a novel site distal to the protease active site. *PLOS Pathog.* **2022,** *18*, e1011065.
43. Wydorski, P. M.; Osipiuk, J.; Lanham, B. T.; Tesar, C.; Endres, M.*, et al.* Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. *Nat. Commun.* **2023,** *14*, 2366.
44. Malakhov, M. P.; Malakhova, O. A.; Kim, K. I.; Ritchie, K. J.; Zhang, D.-E. UBP43 (USP18) specifically removes ISG15 from conjugated proteins. *J. Biol. Chem.* **2002,** *277*, 9976.
45. Gold, I. M.; Reis, N.; Glaser, F.; Glickman, M. H. Coronaviral PLpro proteases and the immunomodulatory roles of conjugated versus free Interferon Stimulated Gene product-15 (ISG15). *Semin. Cell Dev. Biol.* **2022,** *132*, 16.
46. Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.*, et al.* Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infect. Dis.* **2020,** *6*, 2099.
47. Redhead, M. A.; Owen, C. D.; Brewitz, L.; Collette, A. H.; Lukacik, P.*, et al.* Bispecific repurposed medicines targeting the viral and immunological arms of COVID-19. *Sci. Rep.* **2021,** *11*, 13208.
48. Ritorto, M. S.; Ewan, R.; Perez-Oliva, A. B.; Knebel, A.; Buhrlage, S. J.*, et al.* Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat. Commun.* **2014,** *5*, 4763.
49. Armstrong, L. A.; Lange, S. M.; Dee Cesare, V.; Matthews, S. P.; Nirujogi, R. S.*, et al.* Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLoS One* **2021,** *16*, e0253364.

50. Brewitz, L.; Kamps, J. J. A. G.; Lukacik, P.; Strain-Damerell, C.; Zhao, Y., *et al.* Mass Spectrometric Assays Reveal Discrepancies in Inhibition Profiles for the SARS-CoV-2 Papain-Like Protease. *ChemMedChem* **2022,** *17*, e202200016.
51. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W., *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265.
52. Shi, H.-X.; Yang, K.; Liu, X.; Liu, X.-Y.; Wei, B., *et al.* Positive regulation of interferon regulatory factor 3 activation by Herc5 via ISG15 modification. *Mol. Cell. Biol.* **2010,** *30*, 2424.
53. Chattopadhyay, S.; Kuzmanovic, T.; Zhang, Y.; Wetzel, J. L.; Sen, G. C. Ubiquitination of the transcription factor IRF-3 activates RIPA, the apoptotic pathway that protects mice from viral pathogenesis. *Immunity* **2016,** *44*, 1151.
54. Giannakopoulos, N. V.; Luo, J.-K.; Papov, V.; Zou, W.; Lenschow, D. J., *et al.* Proteomic identification of proteins conjugated to ISG15 in mouse and human cells. *Biochem. Biophys. Res. Commun.* **2005,** *336*, 496.
55. Li, C.; Nelson, T. G.; Vertegaal, A. C. O.; Thibault, P. Proteomic strategies for characterizing ubiquitin-like modifications. *Nat. Rev. Methods Primers* **2021,** *1*, 53.
56. Zhao, C.; Denison, C.; Huibregtse, J. M.; Gygi, S.; Krug, R. M. Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 10200.
57. Zhang, Y.; Thery, F.; Wu, N. C.; Luhmann, E. K.; Dussurget, O., *et al.* The in vivo ISGylome links ISG15 to metabolic pathways and autophagy upon Listeria monocytogenes infection. *Nat. Commun.* **2019,** *10*, 5383.
58. Okumura, F.; Zou, W.; Zhang, D.-E. ISG15 modification of the eIF4E cognate 4EHP enhances cap structure-binding activity of 4EHP. *Genes Dev.* **2007,** *21*, 255.
59. Glaser, W.; Skern, T. Extremely efficient cleavage of eIF4G by picornaviral proteinases L and 2A in vitro. *FEBS Lett.* **2000,** *480*, 151.
60. Chan, H. T. H.; Brewitz, L.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A., *et al.* Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2′ proline of the viral polyprotein. *bioRxiv* **2023**, https://doi.org/10.1101/2023.07.11.548309.
61. Stukalov, A.; Girault, V.; Grass, V.; Karayel, O.; Bergant, V., *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **2021,** *594*, 246.
62. Kuo, C.-J.; Chao, T.-L.; Kao, H.-C.; Tsai, Y.-M.; Liu, Y.-K., *et al.* Kinetic characterization and inhibitor screening for the proteases leading to identification of drugs against SARS-CoV-2. *Antimicrob. Agents Chemother.* **2021,** *65*, e02577.
63. Nakashima, Y.; Brewitz, L.; Tumber, A.; Salah, E.; Schofield, C. J. 2-Oxoglutarate derivatives can selectively enhance or inhibit the activity of human oxygenases. *Nat. Commun.* **2021,** *12*, 6478.
64. Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D., *et al.* Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **2020,** *6*, eabd4596.
65. Zhang, Y.; Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019,** *35*, 5121.
66. Báez-Santos, Y. M.; St. John, S. E.; Mesecar, A. D. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* **2015,** *115*, 21.
67. Henderson, J. A.; Verma, N.; Harris, R. C.; Liu, R.; Shen, J. Assessment of proton-coupled conformational dynamics of SARS and MERS coronavirus papain-like proteases: Implication for designing broad-spectrum antiviral inhibitors. *J. Chem. Phys.* **2020,** *153*, 115101.
68. Gao, X.; Qin, B.; Chen, P.; Zhu, K.; Hou, P., *et al.* Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm. Sin. B* **2021,** *11*, 237.
69. Fu, Z.; Huang, B.; Tang, J.; Liu, S.; Liu, M., *et al.* The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **2021,** *12*, 488.
70. Hu, M.; Li, P.; Li, M.; Li, W.; Yao, T., *et al.* Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* **2002,** *111*, 1041.
71. Hu, M.; Li, P.; Song, L.; Jeffrey, P. D.; Chernova, T. A., *et al.* Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J.* **2005,** *24*, 3747.
72. Ratia, K.; Pegan, S.; Takayama, J.; Sleeman, K.; Coughlin, M., *et al.* A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. USA* **2008,** *105*, 16119.
73. Lee, H.; Lei, H.; Santarsiero, B. D.; Gatuz, J. L.; Cao, S., *et al.* Inhibitor recognition specificity of MERS-CoV papain-like protease may differ from that of SARS-CoV. *ACS Chem. Biol.* **2015,** *10*, 1456.

74. Osipiuk, J.; Azizi, S.-A.; Dvorkin, S.; Endres, M.; Jedrzejczak, R.*, et al.* Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **2021,** *12*, 743.
75. Shen, Z.; Ratia, K.; Cooper, L.; Kong, D.; Lee, H.*, et al.* Design of SARS-CoV-2 PLpro inhibitors for COVID-19 antiviral therapy leveraging binding cooperativity. *J. Med. Chem.* **2022,** *65*, 2940.
76. Wertz, I. E.; Murray, J. M. Structurally-defined deubiquitinase inhibitors provide opportunities to investigate disease mechanisms. *Drug Discov. Today Technol.* **2019,** *31*, 109.
77. Zhao, Y.; Du, X.; Duan, Y.; Pan, X.; Sun, Y.*, et al.* High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. *Protein Cell* **2021,** *12*, 877.
78. Ma, C.; Sacco, M. D.; Xia, Z.; Lambrinidis, G.; Townsend, J. A.*, et al.* Discovery of SARS-CoV-2 Papain-like Protease Inhibitors through a Combination of High-Throughput Screening and a FlipGFP-Based Reporter Assay. *ACS Cent. Sci.* **2021,** *7*, 1245.
79. Drag, M.; Mikolajczyk, J.; Bekes, M.; Reyes-Turcu, F. E.; Ellman, J. A.*, et al.* Positional-scanning fluorigenic substrate libraries reveal unexpected specificity determinants of DUBs (deubiquitinating enzymes). *Biochem. J.* **2008,** *415*, 367.
80. Rut, W.; Zmudzinski, M.; Snipas, S. J.; Bekes, M.; Huang, T. T.*, et al.* Engineered unnatural ubiquitin for optimal detection of deubiquitinating enzymes. *Chem. Sci.* **2020,** *11*, 6058.
81. Swatek, K. N.; Aumayr, M.; Pruneda, J. N.; Visser, L. J.; Berryman, S.*, et al.* Irreversible inactivation of ISG15 by a viral leader protease enables alternative infection detection strategies. *Proc. Natl. Acad. Sci. USA* **2018,** *115*, 2371.
82. Akutsu, M.; Ye, Y.; Virdee, S.; Chin, J. W.; Komander, D. Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains. *Proc. Natl. Acad. Sci. USA* **2011,** *108*, 2228.
83. James, T. W.; Frias-Staheli, N.; Bacik, J.-P.; Levingston Macleod, J. M.; Khajehpour, M.*, et al.* Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease. *Proc. Natl. Acad. Sci. USA* **2011,** *108*, 2222.
84. Huai, W.; Liu, X.; Wang, C.; Zhang, Y.; Chen, X.*, et al.* KAT8 selectively inhibits antiviral immunity by acetylating IRF3. *J. Exp. Med.* **2019,** *216*, 772.
85. Wang, C.; Wang, Q.; Xu, X.; Xie, B.; Zhao, Y.*, et al.* The methyltransferase NSD3 promotes antiviral innate immunity via direct lysine methylation of IRF3. *J. Exp. Med.* **2017,** *214*, 3597.
86. Munnur, D.; Teo, Q.; Eggermont, D.; Lee, H. H. Y.; Thery, F.*, et al.* Altered ISGylation drives aberrant macrophage-dependent immune responses during SARS-CoV-2 infection. *Nat. Immunol.* **2021,** *22*, 1416.
87. Bade, V. N.; Nickels, J.; Keusekotten, K.; Praefcke, G. J. K. Covalent protein modification with ISG15 via a conserved cysteine in the hinge region. *PLoS One* **2012,** *7*, e38294.
88. Jeon, S. J.; Chung, K. C. Covalent conjugation of ubiquitin-like ISG15 to apoptosis-inducing factor exacerbates toxic stimuli-induced apoptotic cell death. *J. Biol. Chem.* **2022,** *298*, 102464.
89. Durfee, L. A.; Lyon, N.; Seo, K.; Huibregtse, J. M. The ISG15 conjugation system broadly targets newly synthesized proteins: implications for the antiviral function of ISG15. *Mol. Cell* **2010,** *38*, 722.
90. Zhao, C.; Hsiang, T.-Y.; Kuo, R.-L.; Krug, R. M. ISG15 conjugation system targets the viral NS1 protein in influenza A virus–infected cells. *Proc. Natl. Acad. Sci. USA* **2010,** *107*, 2253.
91. Perng, Y.-C.; Lenschow, D. J. ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* **2018,** *16*, 423.
92. Frias-Staheli, N.; Giannakopoulos, N. V.; Kikkert, M.; Taylor, S. L.; Bridgen, A.*, et al.* Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe* **2007,** *2*, 404.
93. Isaacson, M. K.; Ploegh, H. L. Ubiquitination, ubiquitin-like modifiers, and deubiquitination in viral infection. *Cell Host Microbe* **2009,** *5*, 559.
94. Hung, H.-C.; Wang, H.-C.; Shih, S.-R.; Teng, I. F.; Tseng, C.-P.*, et al.* Synergistic inhibition of enterovirus 71 replication by interferon and rupintrivir. *J. Infect. Dis.* **2011,** *203*, 1784.
95. Xiang, Z.; Liu, L.; Lei, X.; Zhou, Z.; He, B.*, et al.* 3C protease of enterovirus D68 inhibits cellular defense mediated by interferon regulatory factor 7. *J. Virol.* **2016,** *90*, 1613.
96. Lei, X.; Xiao, X.; Xue, Q.; Jin, Q.; He, B.*, et al.* Cleavage of interferon regulatory factor 7 by enterovirus 71 3C suppresses cellular responses. *J. Virol.* **2013,** *87*, 1690.
97. Zhao, Y.; Zhu, Y.; Liu, X.; Jin, Z.; Duan, Y.*, et al.* Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2022,** *119*, e2117142119.
98. Pablos, I.; Machado, Y.; de Jesus, H. C. R.; Mohamud, Y.; Kappelhoff, R.*, et al.* Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CLpro substrate degradome. *Cell Rep.* **2021,** *37*, 109892.

99. Koudelka, T.; Boger, J.; Henkel, A.; Schönherr, R.; Krantz, S.*, et al.* N-Terminomics for the identification of in vitro substrates and cleavage site specificity of the SARS-CoV-2 main protease. *Proteomics* **2021,** *21*, 2000246.
100. Yan, W.; Zheng, Y.; Zeng, X.; He, B.; Cheng, W. Structural biology of SARS-CoV-2: open the door for novel therapies. *Sig. Transduct. Target. Ther.* **2022,** *7*, 26.
101. Xu, D.; Shan, B.; Lee, B.-H.; Zhu, K.; Zhang, T.*, et al.* Phosphorylation and activation of ubiquitin-specific protease-14 by Akt regulates the ubiquitin-proteasome system. *eLife* **2015,** *4*, e10510.
102. Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.*, et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020,** *11*, 3202.
103. Lee, J.; Kenward, C.; Worrall, L. J.; Vuckovic, M.; Gentile, F.*, et al.* X-ray crystallographic characterization of the SARS-CoV-2 main protease polyprotein cleavage sites essential for viral processing and maturation. *Nat. Commun.* **2022,** *13*, 5196.
104. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
105. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical nonequilibrium molecular dynamics simulations Identify allosteric sites and positions associated with drug resistance in the SARS-CoV-2 main protease. *JACS Au* **2023,** *3*, 1767.
106. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
107. Brewitz, L.; Dumjahn, L.; Zhao, Y.; Owen, C. D.; Laidlaw, S. M.*, et al.* Alkyne derivatives of SARS-CoV-2 main protease inhibitors including nirmatrelvir inhibit by reacting covalently with the nucleophilic cysteine. *J. Med. Chem.* **2023,** *66*, 2663.
108. Malla, T. R.; Brewitz, L.; Muntean, D.-G.; Aslam, H.; Owen, C. D.*, et al.* Penicillin derivatives inhibit the SARS-CoV-2 main protease by reaction with its nucleophilic cysteine. *J. Med. Chem.* **2022,** *65*, 7682.
109. Thun-Hohenstein, S. T. D.; Suits, T. F.; Malla, T. R.; Tumber, A.; Brewitz, L.*, et al.* Structure-activity studies reveal scope for optimisation of ebselen-type inhibition of SARS-CoV-2 main protease. *ChemMedChem* **2022,** *17*, e202100582.
110. Miura, T.; Malla, T. R.; Owen, C. D.; Tumber, A.; Brewitz, L.*, et al.* In vitro selection of macrocyclic peptide inhibitors containing cyclic γ2,4-amino acids targeting the SARS-CoV-2 main protease. *Nat. Chem.* **2023,** *15*, 998.
111. Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.*, et al.* Rapid planning and analysis of high-throughput experiment arrays for reaction discovery. *Nat. Commun.* **2023,** *14*, 3924.
112. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
113. Dunbrack Jr, R. L.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997,** *6*, 1661.
114. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
115. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
116. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
117. Hoops, S. C.; Anderson, K. W.; Merz, K. M., Jr. Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991,** *113*, 8262.
118. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269.
119. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993,** *115*, 9620.
120. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.*, et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179.
121. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.*, et al. Gaussian 16 Rev. A.03*, Wallingford, CT, 2016.

122. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T.E. Cheatham, I.*, et al. AMBER 2018*, University of California, San Francisco, 2018.
123. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126*, 014101.
124. Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50*, 1055.
125. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52*, 7182.
126. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089.
127. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.*, et al.* A smooth particle mesh Ewald method. *j. Chem. Phys.* **1995,** *103*, 8577.
128. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.*, et al.* Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
129. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate−DNA helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401.
130. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.*, et al.* Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000,** *33*, 889.
131. Gohlke, H.; Kiel, C.; Case, D. A. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *J. Mol. Biol.* **2003,** *330*, 891.
132. Metz, A.; Pfleger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K.-H.*, et al.* Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein–protein interface. *J. Chem. Inf. Model.* **2012,** *52*, 120.
133. Miller, B. R., III; McGee, T. D., Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.*, et al.* MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314.
134. Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004,** *55*, 383.

7.3 Statement of authorship

Statement of Authorship for joint/multi-authored papers for PGR thesis
To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Mass spectrometric assays monitoring the deubiquitinase activity of the SARS-CoV-2 papain-like protease inform on the basis of substrate selectivity and have utility for substrate identification
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Brewitz, L.; Chan, H. T. H.; Lukacik, P.; Strain-Damerell, C.; Walsh, M. A.; Duarte, F; Schofield, C. J. Mass spectrometric assays monitoring the deubiquitinase activity of the SARS-CoV-2 papain-like protease inform on the basis of substrate selectivity and have utility for substrate identification. *Submitted to J. Virol.* **2023**.

Student Confirmation

Student Name:	Ho Ting Henry CHAN		
Contribution to the Paper	Carried out computational calculations and analyses; contributed to the writing of the original draft, discussions, and the review of the manuscript.		
Signature		Date	1 August 2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Fernanda Duarte			
Supervisor comments I confirm the description above regarding authorship and contributions is accurate.			
Signature		Date	2 August 2023

This completed form should be included in the thesis, at the end of the relevant chapter.

8. Conclusions

The COVID-19 pandemic presented the global scientific community with the unprecedented medicinal challenge of rapidly developing effective vaccines and drugs against the SARS-CoV-2 coronavirus. Due to their essential roles of polyprotein processing in the viral lifecycle, the two cysteine proteases of SARS-CoV-2, the main protease (M^{pro}) and the papain-like protease (PL^{pro}), quickly emerged as protein targets of interest (**Chapter 1**).

An international collaborative effort to understand the activity and inhibition mechanisms of M^{pro} is presented in **Chapter 3** (**Manuscript I**), where complexes of M^{pro} with each of its 11 native polyprotein derived substrates were modelled and investigated through various computational chemistry approaches (**Chapter 2**). This work was initiated at a time when experimental structural information on such complexes was unavailable.[1] The models and analyses obtained through different methods led to a remarkable consensus concerning the conserved nature of M^{pro}-substrate interactions across the P4-P4′ positions that are key to selective substrate recognition by M^{pro}, which acts as molecular "scissors" in catalysing hydrolytic peptide cleavage (**Figure 8.1**).

Since substrate binding is required prior to hydrolysis, and catalysis relies on enzymatic stabilisation of the reaction transition state(s), it was envisaged that the conserved M^{pro}-substrate interactions could be exploited in the design of inhibitors that bind and deactivate M^{pro}. The translational relationship between M^{pro}-substrate and M^{pro}-inhibitor interactions was demonstrated in the design of competitive peptide inhibitors based on substrate complex models,[1] along with analyses of binding modes of fragments,[1] fragment-based ligands,[1] and β-lactam inhibitors (**Chapter 3**).[2] Indeed, the binding mode of the peptidomimetic inhibitor nirmatrelvir, the active pharmaceutical ingredient in the first clinically authorised oral COVID-19 treatment, shares numerous common interaction features with the native M^{pro} substrates, notably in the glutamine-recognising S1 pocket, the hydrophobic S2 pocket, and the backbone binding S3 subsite.[3] These inhibitors, which are structurally diverse and yet utilise similar patterns of

interactions with M^{pro}, act as molecular "rocks" to bind tightly and hinder the action of M^{pro} towards the substrates ("papers") (**Figure 8.1**).

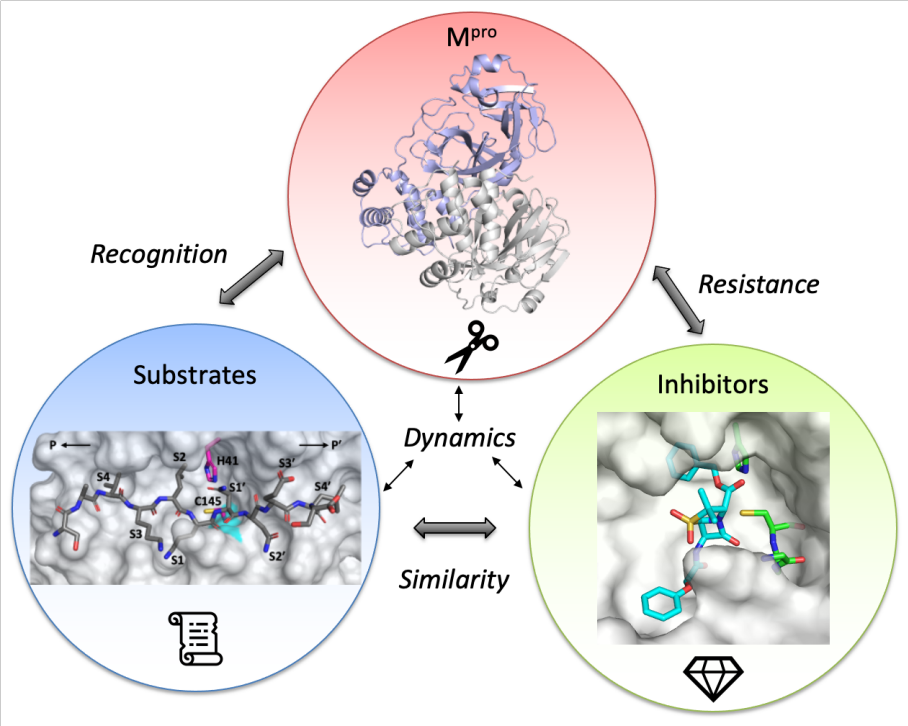


Figure 8.1: The "scissors-paper-rock" interrelationship between M^{pro}, its substrates, and its inhibitors, as discussed in **Manuscript I**.[1]

Central to the relationship between M^{pro}, its substrates, and its inhibitors is the role of protein dynamics, the investigations of which remain limited with currently available experimental methodology. Time-evolved MD simulations offer an approach to explore the protein response to substrate or inhibitor binding, and the associated communication networks within and between protein protomers through propagation of movement "signals" (**Chapter 4**). While these dynamical events might be rarely observed during equilibrium simulations in currently attainable timescales, the D-NEMD approach induces these responses by forcing the system out of equilibrium, enabling statistical analysis of observed responses.[4] The structural responses of M^{pro} to substrate removal analysed by the D-NEMD approach demonstrate communication networks across the M^{pro} dimer interface and between the active site and a validated allosteric inhibition site.[5] The latter means that inhibitor binding will modulate M^{pro} dynamics during the substrate

binding process and hence likely affect the rate and/or affinity of substrate binding (**Figure 8.2**).[6] Furthermore, the D-NEMD approach enables the observation of signals arising from the binding of the substrate P1 glutamine sidechain in the S1 pocket, which propagates to the active site and prompts a transition of the Cys145-His41 catalytic dyad into a catalytically active conformation (**Chapter 5**). This analysis provides new insight into the selectivity of M^{pro} in catalysing hydrolysis at its designated cleavage sites in the viral polyproteins.

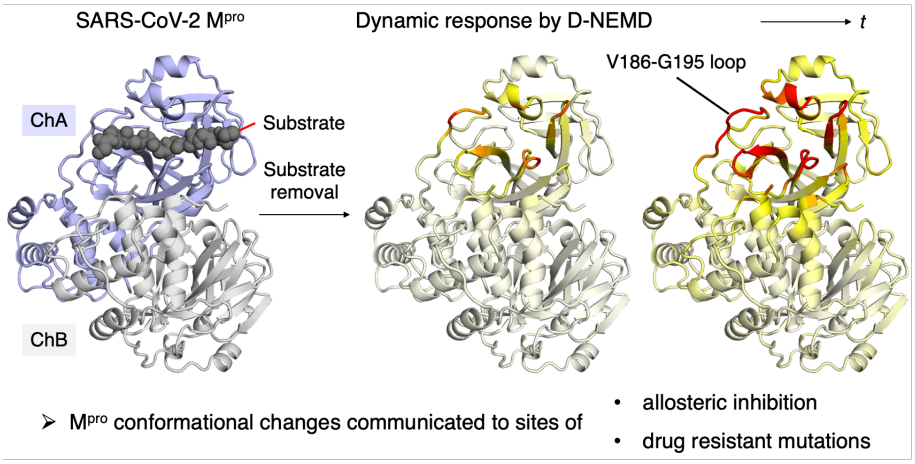


Figure 8.2: D-NEMD investigation of M^{pro} dynamic responses upon substrate removal. Adapted with permission from **Manuscript II**: *JACS Au* **2023**, *3*, 1767. Copyright 2023 American Chemical Society.[6]

Given the similarity in binding modes of substrates and inhibitors, protein dynamics also play fundamental roles in inhibitor binding, as exemplified by the different extents of structural responses associated with the binding of various hydrophobic sidechains in the S2 pocket (**Chapter 5**). Significant structural adaptations in M^{pro} occur during the binding of the P2-Phe in the slower substrate s02 and the P2-Trp in the peptide inhibitor p12, compared to the binding of P2-Leu, which is the predominant P2 residue in the native substrates. In addition, the dynamic contributions of remote M^{pro} residues, such as Leu50 and Gln192, to substrate binding are revealed by the D-NEMD investigations. Mutations of these residues could alter their dynamic behaviour and thus affect the binding affinities and/or reactivities of M^{pro} with its substrates and inhibitors, and in certain cases lead to resistance towards these inhibitors.

In particular, significant movements are observed in Gln192 and the entire Val186-Gly195 loop upon the complexation of the substrate s05 (**Chapter 4**), and more specifically, its P4-Val sidechain (**Chapter 5**). Structural adaptation of this loop also accommodates the P2-Phe and P2-Trp in s02 and p12, respectively. Overall, the induced fit of the Val186-Gly195 loop is key to accommodating the diverse residues, in size and shape, at the P4 and P2 positions in M^{pro} substrates.

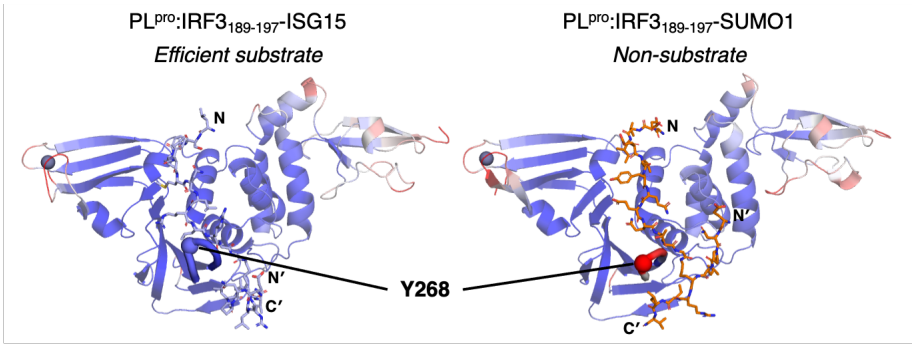


Figure 8.3: The differential dynamic behaviour of the PL^{pro} loop BL2 when PL^{pro} is complexed with an efficient substrate peptide and with a non-substrate. Adapted from **Figure 7.7** (**Manuscript V**).

Loop movements are similarly important in PL^{pro} catalysis. MD simulations of PL^{pro} complexed with polyprotein derived linear peptides (**Chapter 6**) or with N^{ε}-lysine-branched peptides representing Ub- or UBL-modified proteins (**Chapter 7**) emphasise the importance of loop flexibility for substrate recognition, which is, at least in part, regulated by the opening and closing of the Tyr268-containing loop BL2 (**Figure 8.3**). The stability of BL2 in remaining rigidly closed in the PL^{pro}-peptide complex correlates with the experimentally measured efficiencies of PL^{pro}-catalysed hydrolysis. However, hydrolytic efficiency depends on a complex interplay of factors that involve multiple positions or regions in the peptide sequence. Notably, substrate recognition requires a leucine residue, or possibly other AAs with an extended hydrophobic sidechain (*e.g.*, in the reverse binding modes explored in **Chapter 6**), in the S4 pocket. The neighbouring P3 position apparently tolerates any proteinogenic AAs, although a P3-His in the $IRF3_{189-197}$-URM1 peptide appears to be less favourable than Arg (**Chapter 7**). Together with the narrowness of the

S2-S1 channel, these provide PL^{pro} with its observed selectivity towards the P4-P1 LXGG sequence. Nevertheless, MD simulations of PL^{pro}-substrate models suggest that peptide residues beyond P4-P1, in both the N-terminal and C-terminal directions, exhibit interactions with PL^{pro} residues and thus affect catalytic efficiency. Allosteric N-terminal interactions have been proposed to result in preferential deISG15ylation by SARS-CoV-2 PL^{pro} over deubiquitinylation.[7, 8] The importance of C-terminal interactions is showcased by the role of the P2′-Pro in the polyprotein nsp2/3-derived linear peptide, which induces a bend in the overall conformation of the P′ residues and likely plays a role in promoting and/or stabilising the closed BL2 conformation prior to reaction (**Figure 8.4**). Similarly, the S′ binding, C-terminal fragment sequence of the *N*$^{\varepsilon}$-lysine-branched peptides affects BL2 flexibility and thus rate of hydrolysis. Given the complex relationship between catalytic efficiency and the peptide primary sequence, the measurement of BL2 flexibility by MD simulations may provide a computational assay to predict the efficiency PL^{pro}-catalysed hydrolysis of novel Ub- or UBL-modified viral/host proteins (**Chapter 7**).

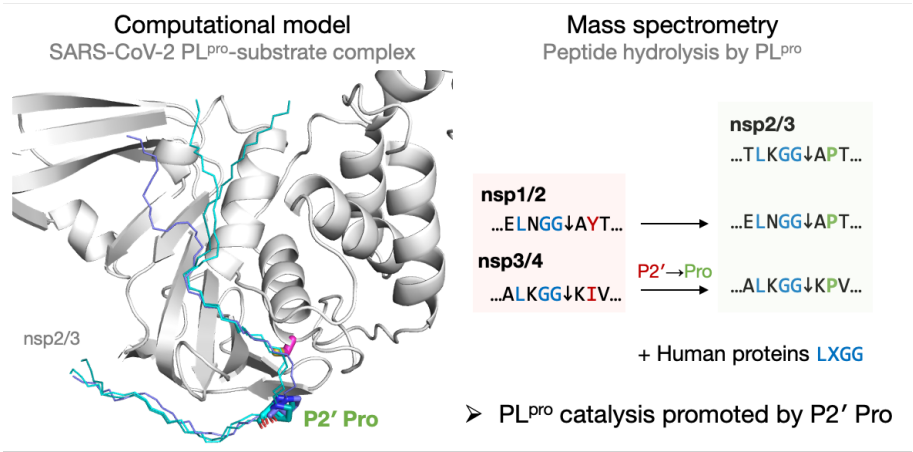


Figure 8.4: The MD-observed effect of the peptide P2′-Pro residue on the conformation and S′ binding interactions of the P′ fragment of PL^{pro} native polyprotein derived substrates, and its experimentally validated role in promoting PL^{pro}-catalysed peptide hydrolysis.

An interesting question remains as to how coronaviruses have evolved to utilise two distinct cysteine proteases for the same function of polyprotein processing. Unlike M^{pro}, PL^{pro} appears to be more promiscuous towards viral and host protein derived substrates, yet it is only responsible

for processing three cleavage sites in the viral polyproteins compared to M^{pro}, which cleaves at 11 positions. In addition, PL^{pro} primary sequence and substrate recognition patterns vary more across coronaviruses (**Chapter 1**).[7, 9] For instance, MERS-CoV PL^{pro} accepts a P4-Ile in its nsp3/4 cleavage site, and substitution of Tyr268 with Thr renders inhibitors of SARS-CoV and SARS-CoV-2 PL^{pro}s ineffective for MERS-CoV.[10]

One possible reason for the distinct natures and functional roles of M^{pro} and PL^{pro} arises from differences in the polyprotein secondary structures near the cleavage sites.[11] PL^{pro} recognises a P2-P1 GG motif, which has a much lower propensity to adopt a helical structure than the LQ motif present in most M^{pro} native substrates.[12] The deubiquitinating and deISG15ylating functions of PL^{pro} provide it with additional pro-viral roles, by interfering with host immune responses.[8, 13] This interaction between PL^{pro} and host proteins may result in greater selection pressure on PL^{pro} to adapt and evolve according to the nature of host cellular processes.[14] It is notable that nsps 1-3, released by PL^{pro}, tend to engage in interactions with host cellular processes, while the remaining nsps released by M^{pro} primarily contribute to internal roles in the viral lifecycle, such as viral replication (**Table 1.1**).[15-20] Maintaining control over these downstream processes might depend on a fine-tuned release of the corresponding nsp by M^{pro},[21] suggesting that sequence conservation near the M^{pro} cleavage sites in the polyproteins might be more advantageous for viral evolution. The localisation of PL^{pro} within nsp3, which also contains transmembrane segments, compared to M^{pro}, which is a standalone nsp, might also govern the distinct roles of these cysteine proteases.[22]

Given the threat to existing inhibitor/drug efficacy caused by mutations in M^{pro} or PL^{pro} during viral evolution, the development of a dual inhibitor that targets both proteases simultaneously is an attractive strategy to counteract mutation-led drug resistance.[11] Considering their respective substrate recognition patterns, a peptidic/peptidomimetic inhibitor based on an LQGG sequence might be worth investigating. The rationale is that the leucine residue is recognised by both the M^{pro} S2 and PL^{pro} S4 pockets, and the glutamine residue can target both the M^{pro} S1 and PL^{pro} S3

subsites. These possibilities could be explored in future computational studies, using a combination of methods discussed in this Thesis.

Utilising the computational models of PL^{pro} complexed with peptide substrates presented in this Thesis (**Chapters 6 and 7**), in a similar manner as performed in for M^{pro} (**Chapter 4**), the D-NEMD approach could be employed to study the structural adaptation of PL^{pro} involved in substrate complexation. A future investigation into the communication networks, particularly the dynamic role of BL2 and of the S2-S1 channel accommodating the substrate Gly-Gly sequence, would be of interest for the discovery and optimisation of inhibitors targeting PL^{pro}.[20]

While the primary focus of this Thesis was on the recognition of substrates prior to hydrolysis, the efficient release of product fragments is also important for efficient catalytic turnover, yet challenging to investigate experimentally. Further computational investigations into the processes involved in product release, including using the D-NEMD approach, should offer insights into the full catalytic mechanism in polyprotein processing.

While the effect of the COVID-19 global pandemic has lessened, the wealth of experimental and computational data on the SARS-CoV-2 cysteine proteases that have been obtained by scientists worldwide, and the advancement of novel and collaborative computational chemistry approaches for investigating protease activity and inhibition, would strengthen our preparedness for future coronavirus outbreaks via ongoing drug discovery and optimisation efforts. Furthermore, there are many other nucleophilic cysteine enzymes beyond coronavirus M^{pro} and PL^{pro} that are of significant medicinal importance. An example is the *Mycobacterium tuberculosis* L,D-transpeptidase Ldt_{Mt2}, an established drug target for tuberculosis.[23, 24] Given the extensive data presently available on the SARS-CoV-2 M^{pro} and PL^{pro}, these proteases can serve as model systems for investigating other cysteine enzymes, as well as other proteases and enzymes in general.

References

1. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
2. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
3. Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.*, et al.* An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374*, 1586.
4. Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur. Phys. J. B* **2021,** *94*, 144.
5. Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.*, et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021,** *372*, 642.
6. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3*, 1767.
7. Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.*, et al.* Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect. Dis.* **2020,** *6*, 2099.
8. Wydorski, P. M.; Osipiuk, J.; Lanham, B. T.; Tesar, C.; Endres, M.*, et al.* Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. *Nat. Commun.* **2023,** *14*, 2366.
9. Ratia, K.; Kilianski, A.; Baez-Santos, Y. M.; Baker, S. C.; Mesecar, A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. *PLOS Pathog.* **2014,** *10*, e1004113.
10. Ullrich, S.; Nitsche, C. SARS-CoV-2 Papain-Like Protease: Structure, Function and Inhibition. *ChemBioChem* **2022,** *23*, e202200327.
11. Yan, F.; Gao, F. An overview of potential inhibitors targeting non-structural proteins 3 (PLpro and Mac1) and 5 (3CLpro/Mpro) of SARS-CoV-2. *Comput. Struct. Biotechnol. J.* **2021,** *19*, 4868.
12. Pace, C. N.; Scholtz, J. M. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998,** *75*, 422.
13. Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.*, et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020,** *587*, 657.
14. Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.*, et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020,** *583*, 459.
15. Jin, Y.; Ouyang, M.; Yu, T.; Zhuang, J.; Wang, W.*, et al.* Genome-Wide Analysis of the Indispensable Role of Non-structural Proteins in the Replication of SARS-CoV-2. *Front. Microbiol.* **2022,** *13*, 907422.
16. Raj, R. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochem. Biophys. Rep.* **2021,** *25*, 100847.
17. Lubin, J. H.; Zardecki, C.; Dolan, E. M.; Lu, C.; Shen, Z.*, et al.* Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. *Proteins* **2022,** *90*, 1054.
18. Yan, W.; Zheng, Y.; Zeng, X.; He, B.; Cheng, W. Structural biology of SARS-CoV-2: open the door for novel therapies. *Signal Transduct. Target. Ther.* **2022,** *7*, 26.
19. Stukalov, A.; Girault, V.; Grass, V.; Karayel, O.; Bergant, V.*, et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **2021,** *594*, 246.
20. Cannalire, R.; Cerchia, C.; Beccari, A. R.; Di Leva, F. S.; Summa, V. Targeting SARS-CoV-2 Proteases and Polymerase for COVID-19 Treatment: State of the Art and Future Opportunities. *J. Med. Chem.* **2022,** *65*, 2716.
21. Meyer, B.; Chiaravalli, J.; Gellenoncourt, S.; Brownridge, P.; Bryne, D. P.*, et al.* Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat. Commun.* **2021,** *12*, 5553.
22. Báez-Santos, Y. M.; St. John, S. E.; Mesecar, A. D. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* **2015,** *115*, 21.

23. Steiner, E. M.; Schneider, G.; Schnell, R. Binding and processing of beta-lactam antibiotics by the transpeptidase LdtMt2 from Mycobacterium tuberculosis. *FEBS J.* **2017,** *284*, 725.
24. Li, W.-J.; Li, D.-F.; Hu, Y.-L.; Zhang, X.-E.; Bi, L.-J.*, et al.* Crystal structure of L,D-transpeptidase LdtMt2 in complex with meropenem reveals the mechanism of carbapenem against Mycobacterium tuberculosis. *Cell Res.* **2013,** *23*, 728.

Appendix for Chapter 3

A. Supplementary Information to Manuscript I

Section S3.1: Methods

S3.1a QM/MM studies on catalytic dyad protonation state

A crystal structure of the TSAVLQ↓SGFRK (↓ indicates the scissile amide bond) substrate peptide-bound inactivated H41A variant of SARS-CoV BJ01 M^{pro} (PDB 2Q6G; 2.50 Å resolution)[1] was used as a starting point for modelling SARS-CoV-2 M^{pro} (M^{pro}) in complex with the substrate peptide s01. The substituted residue Ala41 was converted back to His41 using the automodel routine in Modeller.[2] The crystallographically observed waters were conserved, and hydrogens were added in silico. Protonation states of titratable residues were assigned using PROPKA3;[3] the hydrogen bonding (HB) network was optimised using the protein preparation wizard in Maestro (Schrodinger).[4] Neutral and zwitterionic catalytic dyad models were prepared for the three possible protonation states of His163 (δ-, ε-, or both nitrogens protonated), resulting in six scenarios. The FF14SB force field[5] was used to describe the protein and substrate. A solvation shell of TIP3P water molecules[6] was created 5 Å around the protein using SOLVATE.[7] Further solvation was achieved by construction of a truncated octahedral cell of TIP3P water using LEaP (AmberTools19),[8] with a 10 Å distance from the initial solvation cell to the edge of the box. Na^+ and Cl^- ions were added randomly throughout the solvent with a concentration of 0.1 M NaCl.

Each system was minimised by restraining the protein and allowing relaxation of the solvent and ions. This process was followed by a minimisation restraining only the backbone atoms, a further minimisation restraining only Cαs, and a final minimisation allowing the full system to relax. Each minimisation included 1000 steps of steepest descent, followed by 10,000 steps of conjugate gradient minimisation. The systems were heated by increasing the temperature to 310 K over 100 ps. Langevin dynamics was used with a collision frequency of 5 ps^{-1}; backbone atoms were restrained with a force constant of 5 kcal mol^{-1} $Å^{-2}$. SHAKE was used to restrain bonds to hydrogen atoms.[9] A timestep of 2 fs was used. NPT equilibration was carried out for a total of 9 ns, slowly releasing backbone restraints, with a Monte Carlo barostat maintaining the pressure at 1.01325 bar. Three repeat simulations of 250 ns of production MD were carried out on each of the six systems using pmemd.cuda (AMBER18).[8]

The following QM/MM protocol was followed to carry out umbrella sampling simulations of the proton transfer (PT) between Cys145 and His41. Simulations were performed using sander.MPI (AMBER18).[8] The QM region consisted of the sidechains of Cys145 and His41. The reaction coordinate was defined as the difference between the S-H and N-H distances and was restrained with a force constant of 50 kcal mol^{-1} $Å^{-2}$. The reaction coordinate describing the PT was varied between -1.0 Å and 1.0 Å in steps of 0.1 Å, where a value of -1.0 Å denotes the neutral catalytic dyad and a value of 1.0 Å denotes the zwitterionic catalytic dyad. The reaction coordinate was followed in the forward and reverse directions, with starting snapshots selected from the MD trajectories of the neutral and zwitterionic catalytic dyad, respectively, to ascertain if hysteresis affects the free energy surface for PT. A total of 100 ps of sampling at the DFTB3/MM[10] level of theory was carried out in each reaction coordinate window, but the first 25 ps of sampling was treated as equilibration and discarded. A further backward run of 5 ps of sampling per window was carried out at the ωB97X-D/6-31G*/MM[11-13] level of theory for the structure with HID163 using the interface to Gaussian 16.[14]

S3.1b Interactive Molecular Dynamics in Virtual Reality (iMD-VR) and subsequent implicit-solvent MD

A crystal structure of apo dimeric SARS-CoV-2 M^{pro} (PDB 6YB7)[15] was used as the target for substrate and peptide inhibitor docking using iMD-VR. Protonation states of histidines and other titratable residues were the same as described in **main text Section: "Protonation state of the catalytic dyad"**. The M^{pro}, three natural substrates (s01, s02, s05), and two peptide inhibitors (p12 and p13) tested were parameterised

using the LEaP programme (AMBER19)[8] employing the AMBER99SB-ILDN force field[16] and the OBC2 implicit solvent water model (igb=5).[17] M^{pro} was minimised using OpenMM[18] prior to iMD-VR simulation.

For all iMD-VR simulations, a temperature of 300 K was used with a timestep of 0.5 fs. M^{pro}, all substrates, and both peptide inhibitors remained fully flexible. Whilst in VR, each substrate and peptide inhibitor was docked to M^{pro} following the guidance of 'trace atoms' representing where the s01 substrate should bind; this visual representation was taken from the positions of the s01 backbone atoms in the crystal structure of the H41A mutant of SARS-CoV M^{pro}.[1] These 'trace atoms' were used as a rough visual guide to aid the docking, and the main focus whilst in VR was on establishing key hydrogen bond contacts between the protease and the three substrates and the protease and the two peptide inhibitors.

Once each substrate was docked in VR, a structure where the oxyanion hole interactions were successfully reformed was extracted. These structures were minimised, equilibrated, and subjected to 3 x 200 ns replicates of production MD in implicit solvent to ensure the substrates remained bound. In the case of the docked peptide inhibitors, the docking was repeated 5 times, and structures where the oxyanion hole interactions were successfully reformed were extracted, resulting in 5 docked structures per peptide inhibitor. The 5 docked structures were minimised, equilibrated, and subjected to 500 ns of production MD in implicit solvent. The process of minimisation and equilibration was the same for all substrate and peptide inhibitor structures, and is as follows: First, the structures were iteratively energy minimised at 10 K using slowly decreasing degrees of positional restraint. Restraints of 5 kcal mol^{-1} $Å^{-2}$, 2.5 kcal mol^{-1} $Å^{-2}$, and 1.25 kcal mol^{-1} $Å^{-2}$ were applied to all backbone atoms for the first three rounds of minimisation respectively, and no restraints were applied for the final round. The system was heated by running 10 stages totalling 20 ps of MD, starting at 0 K and linearly increasing the temperature by 30 K at each stage until a temperature of 298 K was reached (each step had a backbone atom restraint of 5 kcal mol^{-1} $Å^{-2}$). 8 rounds of 500 ps of NPT MD with slowly decreasing backbone restraints were run. Restraints were initially 5 kcal mol^{-1} $Å^{-2}$ and halved after each step; once backbone restraints were below 1 kcal mol^{-1} $Å^{-2}$, only the restraints on Cα atoms were retained. The eighth and final stage had no restraints in the system at all.

For the docked substrate structures (s01, s02, and s05), following minimisation and equilibration, 3 x 200 ns replicates of production MD in OBC2 implicit solvent[17] was run for each docked substrate structure, with a protein backbone restraint of 5 kcal mol^{-1} $Å^{-2}$, resulting in 3 x 200 ns MD trajectories for each substrate. In the case of the peptide inhibitors, following minimisation and equilibration, 500 ns of production MD in OBC2 implicit solvent[17] was run for each iMD-VR docked peptide structure, with a protein backbone restraint of 5 kcal mol^{-1} $Å^{-2}$, resulting in 5 x 500 ns MD trajectories for each peptide inhibitor (due to 5 independent docked structures from iMD-VR).

S3.1c Contact interaction mapping

Contact maps. Snapshots from MD models as well as XChem derived crystal structures and covalent docking poses were analysed using Arpeggio.[19] The ligand-M^{pro} complex was processed as described by Jubb et al.[19] by cleaning the PDB file using PDBtools[20] and running Arpeggio on all ligand-M^{pro} contacts. For the MD snapshots of the substrate and designed peptides, a representative snapshot for each complex was chosen by selecting the highest populated cluster and the conformation within the cluster that has the lowest RMSD to all other snapshots in the cluster. From the docked covalent Moonshot submission compounds, the lowest energy pose of the highest populated cluster was chosen. The analysis of the XChem fragments and Moonshot designs was done using the published crystallised conformation.[21-23] The resulting Arpeggio contact map consists of a bit vector for each identified atom-atom contact and classifies them as "Clash", "Covalent", "VdW Clash", "VdW", "Proximal", "Hydrogen Bond", "Weak Hydrogen Bond", "Halogen Bond", "Ionic", "Metal Complex", "Aromatic", "Hydrophobic", "Carbonyl", "Polar" or "Weak Polar".[19]

Hydrophilicity maps. To calculate whether a given protein subsite corresponds to a hydrophilic or hydrophobic pocket, a hydrophilicity score was introduced. All identified atom-atom contacts that interact with a given residue in the substrate are classified as either hydrophobic (Hydrophobic, Aromatic, Halogen Bond) or hydrophilic (Hydrogen Bond, Weak Hydrogen Bond, Ionic, Carbonyl, Polar), excluding the "VdW" and "Weak Polar" interaction types since they were deemed too insignificant and usually redundant as individual atom-atom contacts. The sum of all hydrophobic atom-atom contacts was then subtracted from the sum of all hydrophilic atom-atom contacts to create a hydrophilicity score for each subsite.

Interaction fingerprints. A bit vector was created for every analysed protein-ligand complex, denoting the absence (0) or presence (1) of an interaction of a single ligand with every protein residue that was found to interact with any of the known actives (namely the substrates or the XChem fragments). In order to compare the interaction networks of the ligands, a Tanimoto distance can be calculated between the fingerprint bit vectors using the Jaccard distance[24] and the ligands clustered by their calculated Tanimoto distance.

To investigate potential fragment elaboration pathways, the atom-atom contacts present in each fragment cluster were used as a baseline to investigate fragment growth. To investigate if a designed small molecule ligand exhibits the same binding profile as one of the identified fragment clusters, a standardised cluster profile was created for each fragment cluster which records the presence of a residue level contact if it was classified by Arpeggio as one of the following major contacts: Aromatic, Hydrophobic, Halogen Bond, Polar, Hydrogen Bond, Ionic, Carbonyl. If more than 70% of all recorded residue level contacts of a particular cluster are occupied for an individual ligand, we classify the ligand as a member of that cluster.

S3.1d BigDFT calculations

Snapshots generated from MD (see **main text Section: "Models of SARS-CoV-2 M^{pro}-substrate peptide complexes"**) were studied by Quantum Mechanical (QM) modelling, as implemented in the BigDFT suite.[25] The approach employs Daubechies wavelets to express the electronic structure of the assemblies in the framework of Kohn-Sham (KS) formalism of Density Functional Theory (DFT).[26] With such an approach, the code provides QM results for full systems of large sizes, thanks to the systematic approach offered by wavelets. The electronic structure is expressed by both the density matrix and the KS hamiltonian operator in an underlying basis set of so-called support functions, which are a set of localised functions that are adapted to the chemical environment of the system. Such functions are then expressed in Daubechies wavelets and there are only a few per atom (between 1 and 4 according to the chemical species). The code delivers excellent performance on massively parallel supercomputers and provides the user with the possibility of treating the entire system with the same QM level of theory. A single calculation on one MD-clustered snapshot at the PBE-D3 DFT level requires about 2 h of walltime on 2048 CPU cores (16 nodes) of the IRENE-Rome supercomputer at the TGCC Supercomputing centre in Saclay (Paris). This computational setup, with the inclusion of frozen-core approximation enforced by norm conserving pseudopotentials, was employed for all the DFT calculations presented here. The information to set up the full QM calculation (input file, code version) is available in the GitHub project associated with this publication.

The electronic density matrices as well as the KS hamiltonians expressed in the BigDFT basis were analysed to provide quantum observables on the various portions of the systems. Such a method of analysis has been employed previously and shown to be able to i) evaluate reliable physico-chemical observables on the systems' moieties, thereby decomposing an observable into fragment-based pseudo-observables and ii) assess the pertinence of a given partitioning, by providing an indicator of the quality of the pseudo-observables.[26] In particular, we analysed the strength of the QM interaction on each of the systems' residues, calculated as the matrix elements of the KS hamiltonian reduced on the amino acids. Such analysis provides a linear-response approximation of the energetic contribution by the corresponding residue to the enzyme-peptide interaction and enables a characterisation of the chemical bonding between portions of the system.

For the XChem crystallographic positions, a scheme to equilibrate the position was used, as follows. The pdbfixer program[27] was employed to optimise the crystallographic positions. Water molecules were removed, as well as the hydrogen lost when a ligand formed a covalent bond. Only the M^{pro} monomer-ligand complex was considered for this preliminary dataset. The resulting positions were then optimised with the GFN-FF force field provided by the XTB program.[28]

S3.1e Experimental studies on M^{pro} activity and inhibition

Protein production and purification. Recombinant M^{pro} protein was produced and purified as reported.[29]

Peptide synthesis. Peptide synthesis was performed as reported.[29] s01, s01-LP2W , s01-QP1W, p12, p13, p13-WP2L, p15 and p16 were synthesised on a 0.1 - 0.25 mmol scale from C- to N-terminus on Rink amide-MBHA resin (100–200 mesh, 0.6–0.8 mmol g^{-1} loading, AGTC Bioproducts) using a microwave assisted LibertyBlue peptide synthesiser (CEM) and N-Fmoc protected α-amino acids (CS Bio, Novabiochem, Sigma-Aldrich, TCI, Alfa Aesar, Merck or AGTC Bioproducts). N,N′-

diisopropylcarbodiimide (TCI Europe) and Oxyma Pure (Merck) in DMF and 20% (v/v) piperidine in DMF (peptide synthesis grade, AGTC Bioproducts) were used for iterative cycles of coupling and deprotection respectively under the manufacturer's standard protocol. Following the terminal Fmoc-deprotection step, the resin was washed with CH_2Cl_2, dried in air, then treated with 5-10 mL of a deprotection solution (2.5:2.5:2.5:92.5 (v/v) 1,3-dimethoxybenzene, triisopropylsilane, MilliQ water and trifluoroacetic acid) for 3 h at ambient temperature. The resulting mixture was filtered and the filtrate was diluted with cooled Et_2O (3 x 45 mL) to precipitate the peptide. Et_2O was decanted, peptide dried on air and lyophilised overnight.

Peptides apart from P1 mutant of s01 were dissolved in DMSO and quantified by spiking the sample with 3 mg mL^{-1} of an internal standard 3-(Trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt. The eleven substrate peptides (s01-s11) were purchased from GLBioChem (Shanghai).

Substrate turnover analysis under denaturing conditions. 20 μM stock of all 11 native substrate peptide sequences were prepared in the assay buffer (20 mM HEPES, pH 7.5, 50 mM NaCl). E1-ClipTip™ Bluetooth™ Electronic multichannel pipette (ThermoFisher) was used to dispense 5 μL/well (x24) of each peptide in a single row of a 384 well plate. The first column was treated with a final concentration of 1% (v/v) aqueous formic acid to obtain a 0 min time point. M^{pro} was dispensed using Multidrop to obtain a final concentration of 0.15 μM M^{pro} with 2 μM peptides in all wells. Each column was sequentially quenched (every minute) with 1% (v/v) aqueous formic acid. Samples were analysed by solid-phase extraction (SPE) coupled to mass spectrometry (MS) using a RapidFire Mass Spectrometer. The operating parameters in the positive ion mode were: capillary voltage (4000 V), nozzle voltage (1000 V), fragmentor voltage (365 V), drying gas temperature (280 °C), gas flow (13 L min^{-1}), sheath gas temperature (350 °C) and sheath gas flow (12 L min^{-1}). Samples were loaded onto a SPE C4-cartridge, which was then washed with 0.1% (v/v) aqueous formic acid to remove non-volatile buffer salts (5.5 s, 1.5 mL min^{-1}) followed by elution with aqueous 85% (v/v) acetonitrile in 0.1% (v/v) formic acid (5.5 s, 1.25 mL min^{-1}). The cartridge was equilibrated with 0.1% (v/v) aqueous formic acid (0.5 s, 1.25 mL min^{-1}) prior to every sample injection. Data were exported in a plate list mode and processed in Excel to calculate percentage product turnover.

Substrate binding and turnover analysis under non-denaturing conditions. Non-denaturing mass spectra were obtained using a Waters Synapt HDMS Q-TOF mass spectrometer coupled with an automated chip-based nano-electrospray ion source (TriVersa Nanomate, Advion). A larger concentration of M^{pro} than the one used in the denaturing MS assays was used to provide sufficient sensitivity. 5 μM of M^{pro} was mixed with 13-fold molar excess of a substrate (s01-s11) in 200 mM of ammonium acetate (pH 6.9) at room temperature and electrosprayed (1.77 kV spray voltage, 0.55 psi spray backing gas pressure and 4.3 mbar inlet pressure). The sample and extractor cone voltages were maintained at 180 V and 1 V, respectively; no in-source dissociation of M^{pro} dimers was observed at these voltages. Mass spectra were recorded after 1, 3, 6, 9 and 12 min incubation. Measurements were taken in duplicate for each substrate. Data collection and analysis were carried out using Waters MassLynx software. Integrated peak areas of the substrate ions and cleavage product ions were compared at different time points: the sum of substrate and product ions intensities was set at 100% for each measurement, and the level of depletion of the substrate ions was used as a measure of the turnover efficiency.

Dose response curve analysis. Methods for SPE coupled RapidFire MS-based inhibition assay were as reported.[29] In brief, in a 384 polypropylene well plate, 100 μL of 2.5 mM stocks of the designed peptides were transferred. 11 point 3 fold serial dilutions of the peptides were performed in 60 μL using E1-ClipTip™ Bluetooth™ Electronic multichannel pipette (ThermoFisher) in DMSO with 5 mix cycles of 30 μL volume for mixing. 10 μL was drawn from each well and 5 μL was transferred to two wells of a new destination 384 well polypropylene plate. 5 μL of DMSO (positive turnover control) and 5 μL of 10% (v/v) aqueous formic acid (negative turnover control) were added to 16 wells each on every destination plate. 25 μL/well of x2 stock of enzyme in assay buffer (20 mM HEPES, pH 7.5, 50 mM NaCl) was dispensed using a Multidrop Combi machine; incubation for 15 minutes was followed by dispensation of x2 stock of substrate in each well to obtain 0.15 μM M^{pro} and 2 μM s01 concentration. The reaction was allowed to progress for 10 min (~ 50% turnover in DMSO control), then quenched with 5 μL of 10% (v/v) aqueous formic acid. The plates were centrifuged for ~15 s after addition of each reagent at 2500 rpm (Star lab) to ensure all dispensed solutions were pooled at the bottom of the plate. The plates were analysed by SPE coupled MS under the conditions specified above in "*Substrate turnover analysis under denaturing conditions*". RapidFire integrator software was used to extract and integrate abundance peaks of the +1 charge states of the substrate (1191.68 Da) and N-terminal cleaved product (617.34 Da). Data was exported in a plate list mode and processed in Excel to calculate percentage product turnover, normalisation of percentage activity followed by deduction of percentage inhibition. Normalised percentage inhibition data

were exported to GraphPad Prism 8 and non-linear regression analysis was performed to obtain IC_{50} values. Top and bottom constraints of 100% and 0% were applied respectively for the analysis of reported IC_{50} values curves. Z' of the assay was always ≥0.8.

Dose response curve analysis with varying substrate concentrations. The designed peptides were dispensed using an Echo 550 acoustic liquid handling robot. Samples were prepared as described above ("*Dose response curve analysis*") with final substrate concentrations of 2 µM, 10 µM, 20 µM and 40 µM TSAVLQ/SGFRK-NH_2 (s01) with 10, 10, 15 and 20 minutes of incubation with substrates, respectively.

Designed peptide turnover analysis under denaturing conditions. 100 µM stocks of p12, p13, p15, p16, p13-WP2L, s01-LP2W and s01-QP1W were prepared. 0.15 µM of enzyme was dispensed and incubated with 2 µM peptide (see above**:** "*Substrate turnover analysis under denaturing conditions*"); the reaction was allowed to proceed overnight at 37°C, 300 rpm in a thermomixer. Samples were analysed by SPE coupled MS. After integration using RapidFire Integrator, the data was analysed in Excel and presented using GraphPad Prism 8.

Table S3.1.1: Observed mass (Da) and (*m/z*) charge states of the peptides that were extracted using RapidFire Integrator for peak integration.

Peptides	Sequence	Substrate (Da) (*m/z* charge state)	Product (Da) (*m/z* charge state)
s01	TSAVLQ↓SGFRK	1191.68 (+1)	617.34 (+1)
s02	SGVTFQ↓SAVKR	1177.65 (+1)	637.30 (+1)
s03	KVATVQ↓SKMSD	1191.62 (+1)	
s04	NRATLQ↓AIASE	1171.55 (+1)	
s05	SAVKLQ↓NNELS	1200.57 (+1)	644.37 (+1)
s06	ATVRLQ↓AGNAT	1099.53 (+1)	686.36 (+1)
s07	REPMLQ↓SADAQ	1243.51 (+1)	772.39 (+1)
s08	PHTVLQ↓AVGAC	2185.98 (+2)	
s09	NVATLQ↓AENVT	1157.60 (+1)	644.35 (+1)
s10	TFTRLQ↓SLENV	1305.62 (+1)	764.37 (+1)
s11	FYPKLQ↓SSQAW	1352.59 (+1)	794.37 (+1)
p12	KYTFWQYSQFY	1558.75 (+1)	
p13	KYLTWQNSQIN	1392.70 (+1)	
p15	LTINWQKYFNT	1427.62 (+1)	
p16	WFTLKQYWQTN	1514.70 (+1)	
p13-WP2L	KYLTLQNSQIN	1319.71 (+1)	
s01-LP2W	TSAVWQ↓SGFRK	1264.65 (+1)	690.33 (+1)
s01-QP1W	TSAVLWSGFRK	1249.68 (+1)	

LCMS analysis for designed peptides. LCMS experiments were performed using an Agilent Infinity Series II System attached to QTOF 6650 using an Agilent Zorbax C-18 Extend column. Solvent A: LCMS grade water with 0.1% formic acid, and solvent B: 100% acetonitrile in 0.1% (v/v) formic acid was used at 0.2 mL min^{-1} flow rate to elute the peptides over a gradient of 22-55% of solvent B over 8 minutes. The operating parameters for the LCMS were the same as above ("*Substrate turnover analysis under denaturing conditions*"). In a 96 well plate, samples consisting of 0.15 μM M^{pro} were prepared. p12, p13, p15, p16 and s01 were transferred from source wells to destination wells with M^{pro} using the multi injector programme and samples injected immediately after mixing. 30 min, 3 h, 6 h, 1 day and 2 days time points were obtained for peptides. Samples were covered with a polypropylene cover to limit evaporation.

Designed peptide binding and turnover analysis under non-denaturing conditions. The binding of designed peptides p12, p13, p15 and p16 to M^{pro} dimers and their effects on substrate turnover were investigated using non-denaturing mass spectrometry (see above: "*Substrate binding and turnover analysis under non-denaturing conditions*"). 5 μM of M^{pro} was mixed with designed peptides at different levels of peptide excess in 200 mM of ammonium acetate (pH 6.9) at room temperature. Non-denaturing mass spectra were recorded for different protein-peptide molar concentration ratios (up to 16-fold excess of peptide relative to the protein). At the final step, the native s01 substrate was added to the protein-peptide mixture at 4-fold excess over the protein, and its turnover recorded after 3- and 6-min incubation.

S3.1f Protein-ligand docking

Dataset. A large-scale crystal-based fragment screen against M^{pro} has been conducted using the Diamond synchronton.[23] >500 fragments were screened leading to the discovery of 92 active fragments, 44 of which are covalently bound to Cys145. The structures are available on Fragalysis.[30] The Poster.AI Moonshot project crowdsourced the design of M^{pro} inhibitors based on the original fragment screen. All submissions are made available on the Moonshot project GitHub.[22] For our covalent docking workflow, the dataset as of the 12th of July, 2020 was used, which features 10001 submissions. A subset was created by selecting only submissions with a matching covalent warhead that cite one covalent fragment as their inspiration, correcting for duplicate and incorrect structures, which gave a final dataset of 540 compounds.

Docking workflow. The goal of the workflow was to match each compound design to the corresponding covalent origin fragment and to include the binding pose information of the fragment into docking. The inspiration fragment was cited by the designer of the compound; a list of all designs and inspirations can be found on the Moonshot project GitHub.[22] Each design was matched with the corresponding fragment, the maximum common substructure (MCS) between them was identified, and the conformation of the design was aligned to the fragment before docking. The alignment was performed using a custom alignment script similar to the constrained alignment method in RDKit[31] to force the corresponding atom positions of the MCS into the same conformation, followed by a constrained energy minimisation, keeping the conformation of the MCS constant. Docking was performed using AutoDock4 (AD4), which considers ring conformations to be rigid when sampling ligand conformations before docking.[32] As a result, all rings present in the MCS are already aligned to the crystallographically observed binding pose. Each design was docked to the corresponding M^{pro} crystal structure of the origin fragment, after generation of the homodimer and charge optimization using Protonate3D in MOE.[33]

The FlexRes method in AD4 was used for covalent docking.[32] The covalent adduct of the COVID Moonshot design *after* reaction with the active site Cys145 was selected as the flexible residue and a water molecule included as the "dummy" ligand. Docking and grid parameter files were generated for each Cys145-inhibitor adduct individually with the rest of the corresponding co-crystallised dimeric M^{pro} structure treated as the rigid receptor molecule. Docking with AD4 was performed using the Lamarckian Genetic Algorithm (LGA) and the following AD4 hyperparameters: population size 300; maximum number of energy evaluations 250 000; maximum number of generations 27 000; number of dockings 100.

The scoring function used by AD4 [32] includes pairwise evaluation of intermolecular interactions of the ligand and the protein, intramolecular interactions between residues of the protein and covalent adduct, and an estimation of the conformational entropy lost upon binding. For the evaluation of the covalent docking procedure, only the *intramolecular* terms are relevant, since they correspond to the changes in the energy of the flexible residues plus covalent adduct. Since AD4 automatically clusters docking results by the total estimated free energy of binding, covalent docking results must be re-clustered using the "Final Total Internal Energy" instead (as reported in the DLG docking log file). For clustering the docked poses of the

covalent adducts, the same hierarchical clustering procedure as used in the native AD4 method was employed. A new cluster was seeded with the lowest energy pose, and all remaining poses within a threshold (< 2 Å RMSD) are added to that cluster. The procedure was repeated for the next lowest energy pose, until all docked poses have been assigned to a cluster. RMSD values between poses were calculated using the Open Drug Discovery Toolkit (ODDT),[34] to account for intramolecular symmetry, such as equivalent methyls in tertiary butyl groups.

Docked poses were compared to the original origin fragment crystal structure using SuCOS.[35] SuCOS produces normalised scores to a value between 0 and 1, where 1 indicates perfect overlap and identical molecules. Both the shape and pharmacophoric feature overlaps are weighted equally in the SuCOS score. Based on work by Leung *et al.*,[35] a SuCOS score of 0.55 between two molecules was found to be equivalent to a pose-pose RMSD of 2 Å.

Covalent docking for PF-07321332 was performed identically to that for the other covalent Moonshot designs with the exception that no pre-alignment of the ligand was performed prior to docking. Instead, a random conformation of the ligand was used to seed the docking process. The azanide nitrogen was assigned a negative charge prior to docking.

Section S3.2: Supplementary Results – Substrate Binding and Recognition

S3.2a QM/MM studies of proton transfer in the catalytic dyad

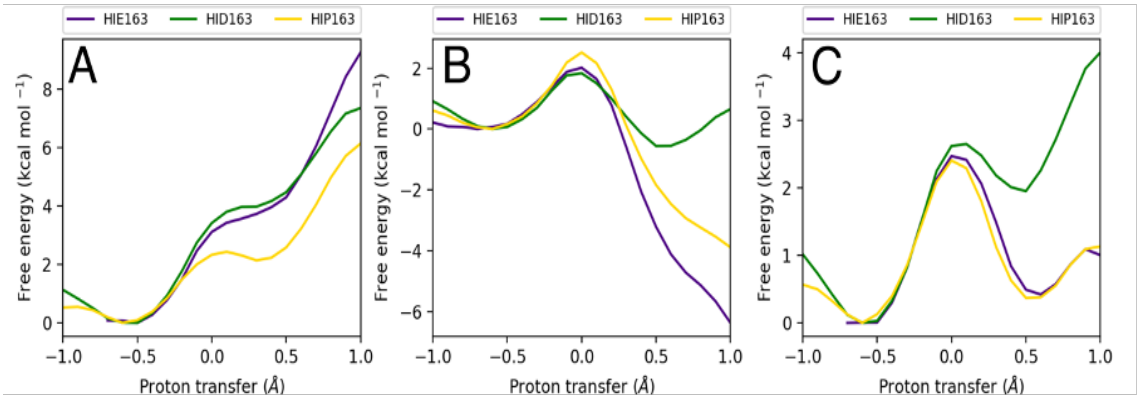


Figure S3.2.1: Free energy profiles of the proton transfer (PT) between Cys145 and His41 in s01-bound SARS-CoV-2 M^{pro} with three different protonation states of His163. (A) Free energy profiles from the neutral dyad. (B) Free energy profiles from the zwitterionic dyad. (C) Combined free energy profiles from the profiles in A and B. All free energy profiles were generated at the DFTB3/MM level of theory. A reaction coordinate value of -1.0 represents the neutral catalytic dyad and a value of 1.0 represents the zwitterionic catalytic dyad. HIE, HID and HIP refer to the Nε, Nδ, and doubly protonated models respectively in the Amber force field naming scheme.

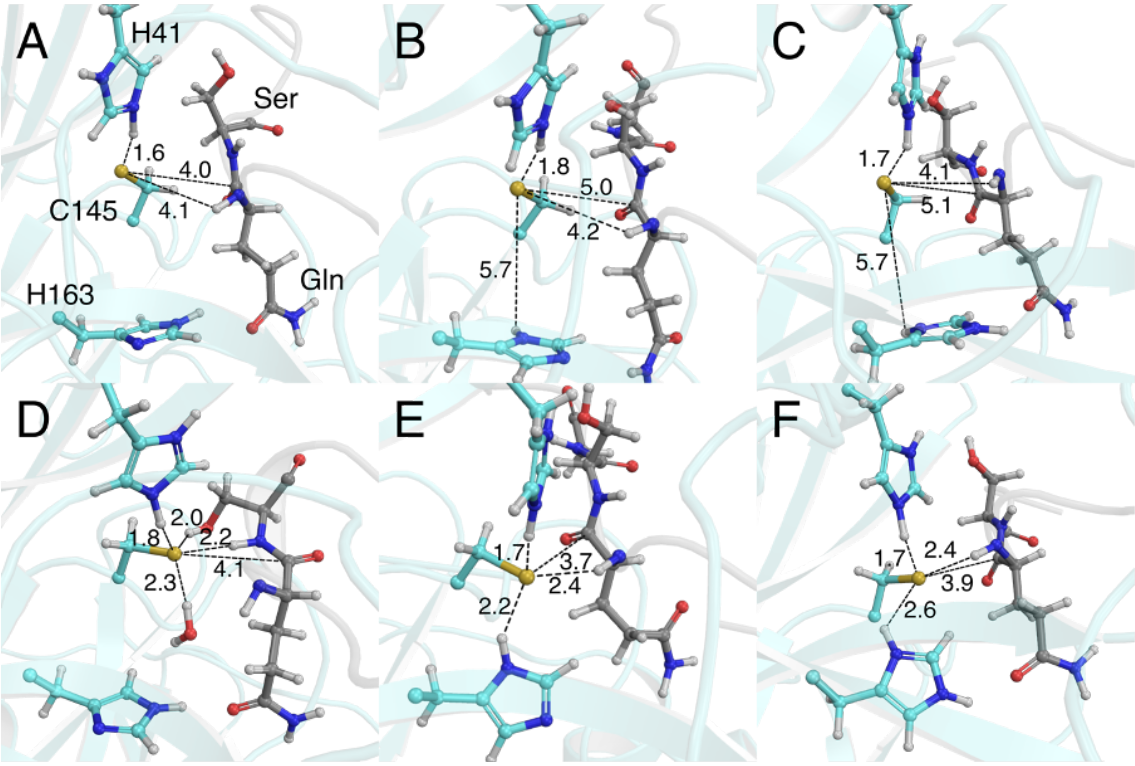


Figure S3.2.2: MD snapshots from umbrella sampling windows of the zwitterionic states of different free energy trajectories showing important interactions with the Cys145 thiol(ate). (A) HIE163, forwards. (B) HID163, forwards. (C) HIP163, forwards. (D) HIE163, backwards. (E) HID163, backwards. (F) HIP163, backwards. Distances are in Angstroms.

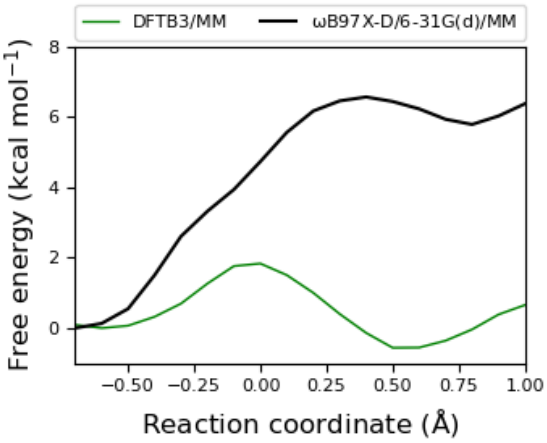


Figure S3.2.3: Free energy profiles of the catalytic dyad PT (simulated backwards) in the HID163 system, at the DFTB3/MM and ωB97X-D/6-31G(d)/MM levels of theory.

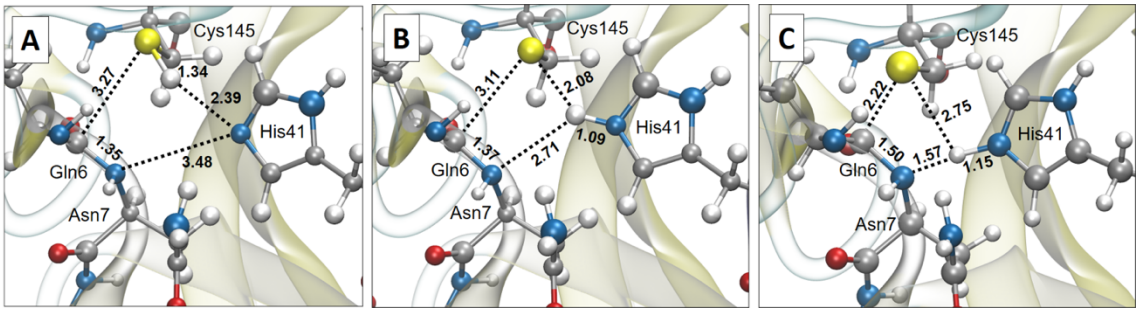


Figure S3.2.4: M06-2X(6-31+G(d,p))/MM structures of the neutral dyad reactant state (A), the pseudo-stable ion pair dyad state (B), and the transition state of the proteolysis (C) that connects the neutral dyad with the covalent bond adduct formed between the protein and the substrate. The structure of the transition state was fully optimised and characterised (imaginary frequency = 349.985*i* cm^{-1}) from coordinates of the most populated clusters of the designed peptide p13, including all water molecules (and counterions). The structures of the ion-pair dyad and the neutral dyad were localised along the IRC (intermediate reaction coordinate) path computed at the same level of theory from the optimised transition state. Importantly, structure of the ion pair dyad does not appear as a minimum in the IRC path, but as a shoulder. Distances are in Ångstroms. The QM sub-set of atoms includes the side chain of His41 (link atom between Cα and Cβ), Cys145 together with carboxyl group of Ser144 and part of Gly146 (link atoms between C and Cα in both of them), and part of the substrate that includes full Gln6 and Asn7 together with carbonyl group of Trp5 and Ser8 (link atoms between C and Cα in both of them), which were treated at M06-2X(6-31+G(d,p)) with Gaussian 09.[14] The rest of the protein, solvent water molecules, and counterions were described by AMBER and TIP3P force fields,[36] as implemented in the fDynamo library.[37, 38]

S3.2b Tautomeric and conformational states of M^{pro} histidine residues

Table S3.2.1: SARS-CoV-2 M^{pro} histidine protonation states adopted in this study, based on the *apo* crystal structure (PDB 6YB7).[15] Using standard nomenclature in the AMBER force fields, δ-protonated His residues are denoted "HID" and ε-protonated His residues "HIE".

Histidine	State	Reason
41	**HID**	General acid/base catalysis, including deprotonation of Cys145 via its unprotonated ε-nitrogen.
64	**HIE**	Solvent exposed (HIE chosen as default).
80	**HID**	δ-NH forms a hydrogen bond with Asn63 side chain oxygen.
163	**HIE**	ε-NH forms a hydrogen bond with substrate P1 Gln side chain oxygen. In the *apo* M^{pro} structure (PDB 6YB7), the sulfoxide oxygen of a DMSO molecule takes the place of this oxygen.
164	**HIE**	Protonation state uncertain based on inspection of crystal structure, but preliminary MD simulations (*vide infra*) suggested HIE is more stable.
172	**HIE**	ε-NH forms a hydrogen bond with Glu166 side chain oxygen.
246	**HIE**	δ-N accepts a hydrogen bond from Thr243 backbone NH.

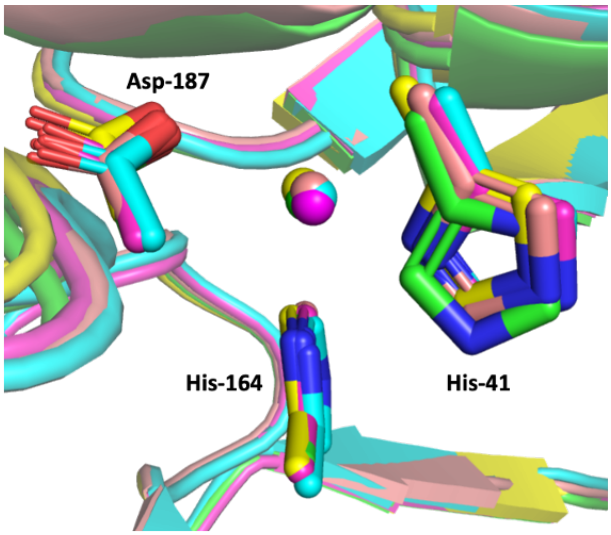


Figure S3.2.5: A conserved water molecule (shown as sphere in identical colour as the protein) is located between the sidechains of His41, His164 and Asp187, in various M^{pro} crystal structures (aligned using chain A), including PDB 6YB7 (HOH 644; green),[15] 6LU7 (HOH 445; cyan),[39] 7BQY (HOH 570; magenta),[39] 6Y2G (HOH 560; yellow)[40] and 6WQF (HOH 417; salmon).[41] Note that His41 in PDB 6YB7 (green) has a unique conformational state.

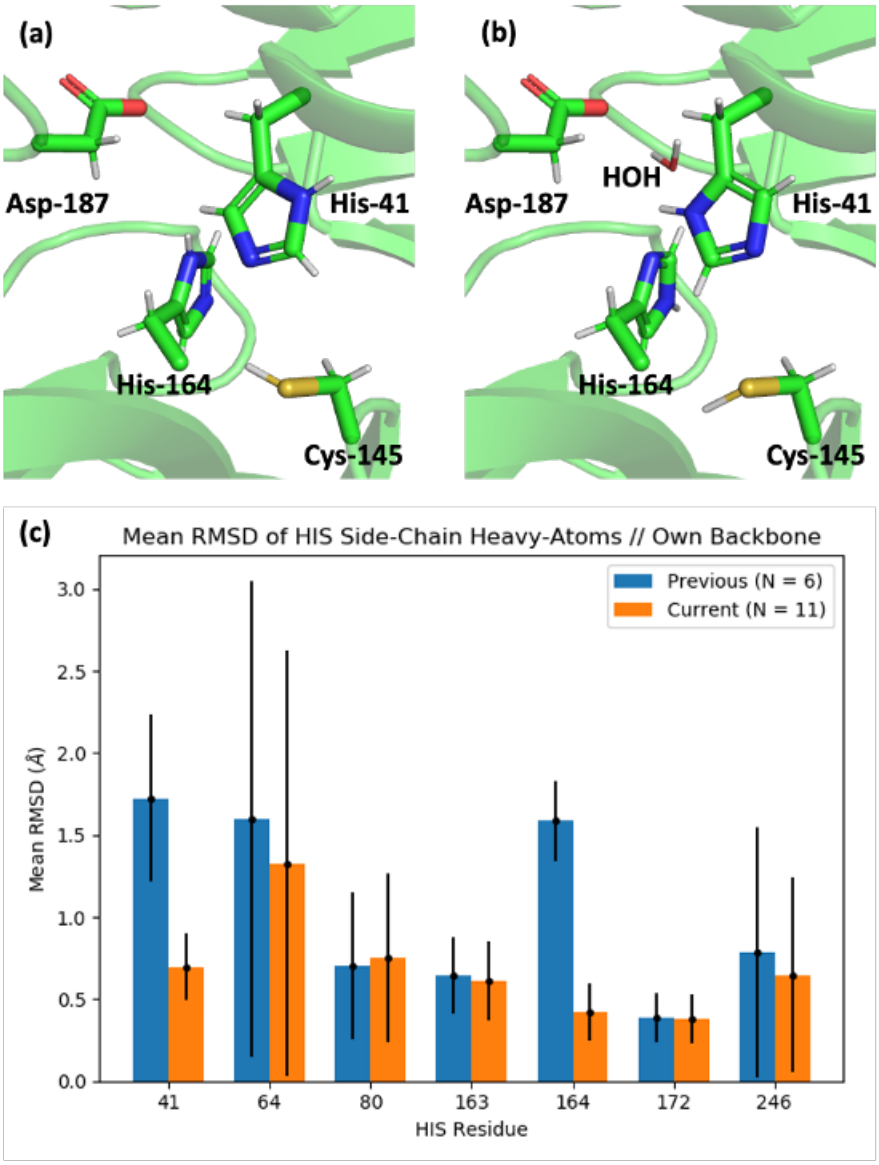


Figure S3.2.6: Comparison of the (a) previous and (b) current setups of His41 and its surrounding residues for MD simulations (PDB 6YB7). The changes involve rotation of His41, ε-protonation on His164, and the retention of water HOH 644. Protons were added automatically by GROMACS (v. 2019.2). (c) The bar plot compares the mean RMSD of the sidechain heavy atoms of each His residue over 100 ns MD relative to their starting positions, after fitting to its respective backbone (N, Cα, C) atoms. Error bar refers to mean standard deviation (N = 6 or 11; see legend). A higher mean RMSD indicates greater deviation from the setup, while a larger error bar reflects higher flexibility during MD. The current setup resulted in lower and less fluctuating RMSD values for both His41 and His164. The previous set of simulations involved the M^{pro} dimer in complex with truncated s01 substrates (Ac-AVLQSG-NHMe) in both active sites, whereas the current simulations involved the 11 M^{pro}-substrate (s01-s11) complexes each with one active site occupied.

S3.2c Models of M^{pro}-substrate peptide complexes

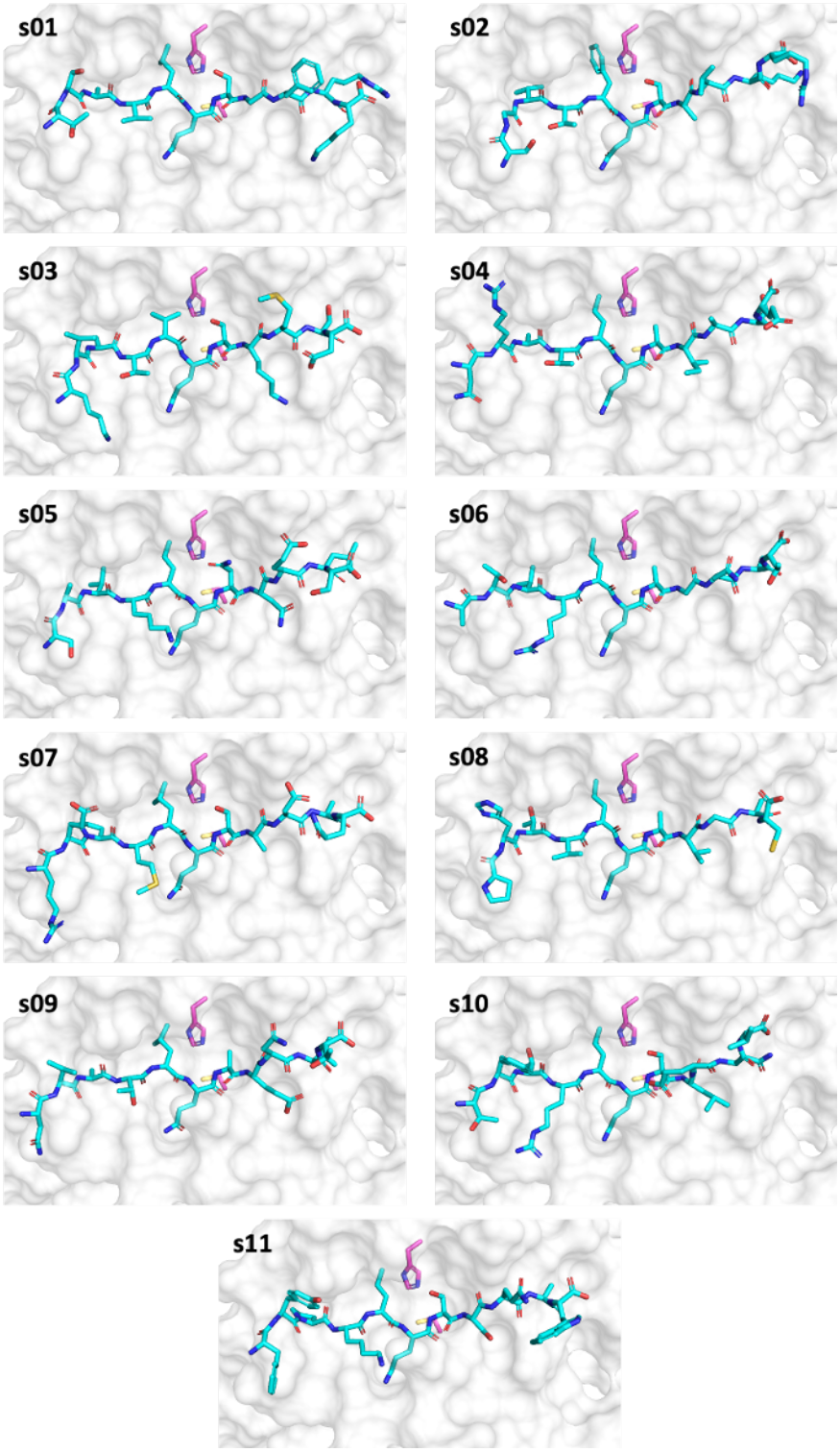


Figure S3.2.7: Starting conformations of the 11 native substrates (s01-s11; cyan) in complex with SARS-CoV-2 M^{pro} (PDB 6YB7;[15] shown as a white surface with the dyad His41 and Cys145 in magenta), constructed by a comparative modelling approach. The crystal structure of the H41A SARS-CoV M^{pro}-s01 complex (PDB 2Q6G) was used as a template.[1]

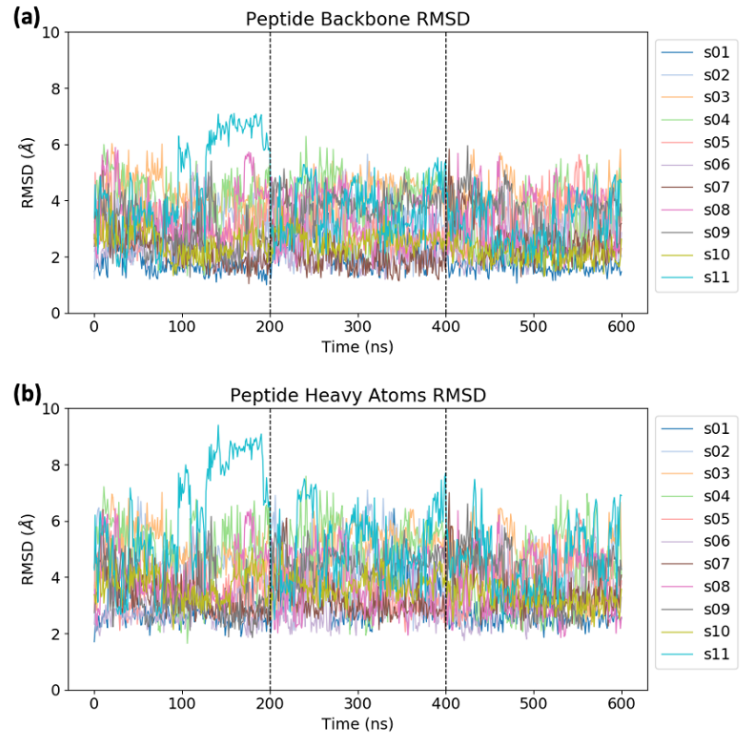


Figure S3.2.8: RMSD of (a) the peptide backbone (N, Cα, C) and (b) all peptide heavy atoms during the concatenated 3 × 200 ns explicitly-solvated MD simulations of the 11 M^{pro}-substrate complexes relative to the initial configuration, with trajectories fitted using the M^{pro} dimer backbone.

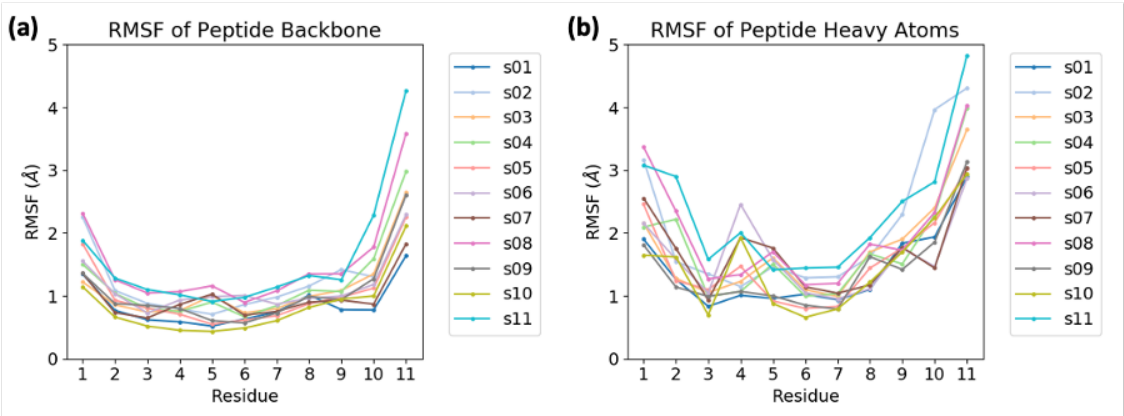


Figure S3.2.9: RMSF of (a) the peptide backbone (N, Cα, C) and (b) all peptide heavy atoms averaged per residue during the explicitly-solvated MD simulations of the 11 M^{pro}-substrate complexes.

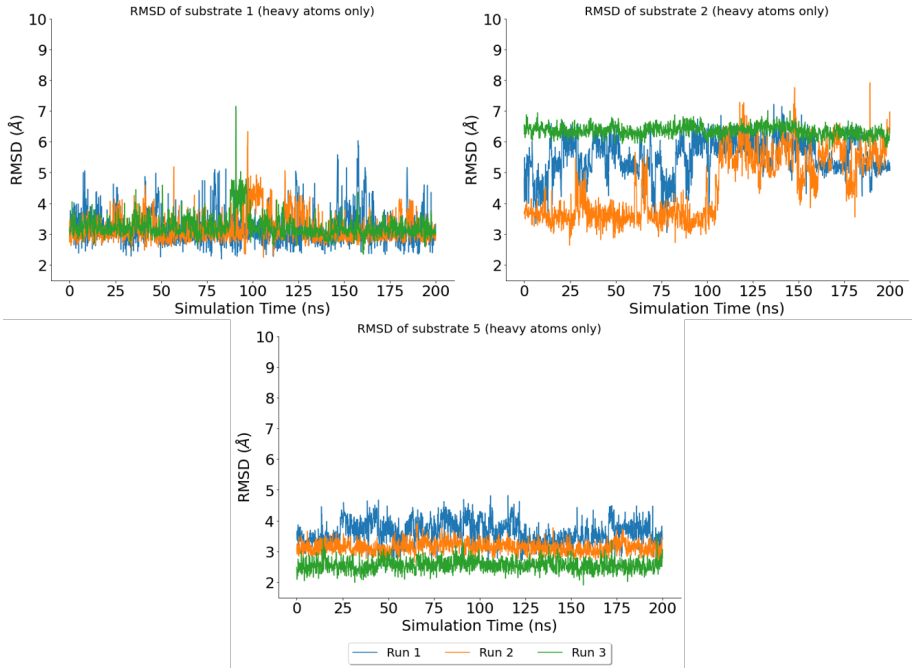


Figure S3.2.10: RMSD of substrate heavy atoms (*i.e.* not including hydrogens) across 3 × 200 ns MD simulation of iMD-VR docked structures, based on comparison to the starting structure (*i.e.* the docked structure from iMD-VR). Substrates s01 and s05 have a lower RMSD on average compared to s02. This could be due to the lack of HBs formed in s02 simulations, compared with s01 and s05 (**Figure S3.3.11**).

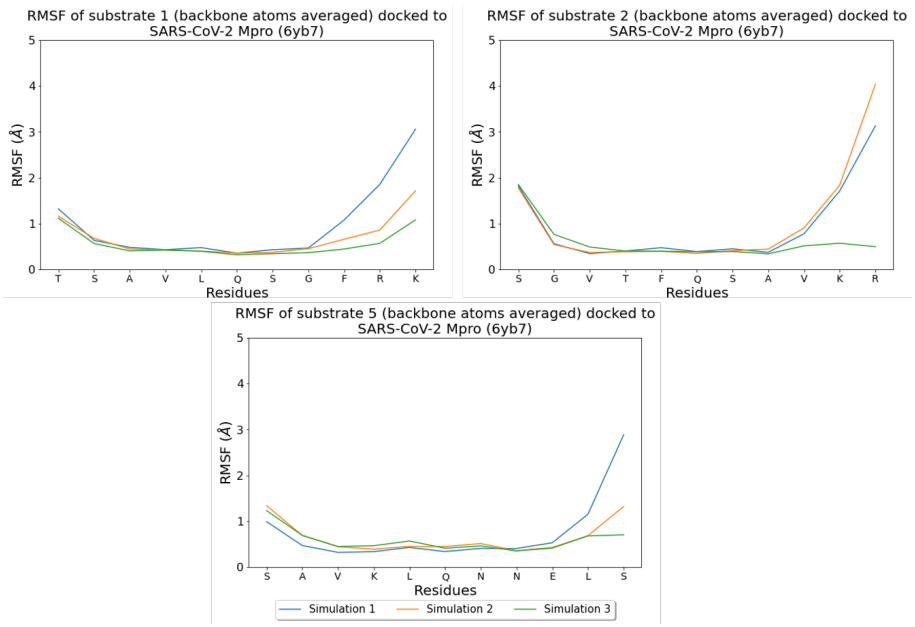


Figure S3.2.11: RMSF of substrate backbone atoms across 3 × 200 ns MD simulation of iMD-VR docked structures. In most cases, the P′ (C-terminal) side of the substrates is more flexible than the P (N-terminal) side.

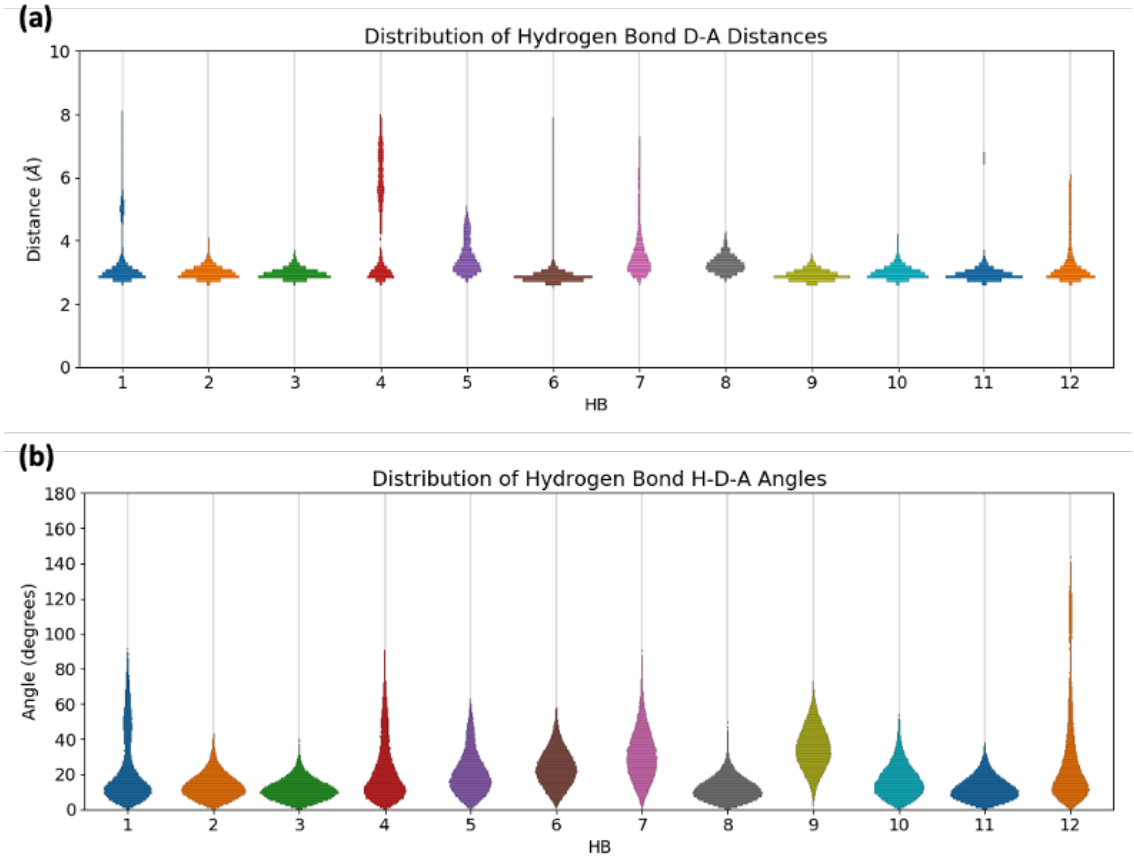


Figure S3.2.12: **Conserved hydrogen bond interactions**. Overall distributions of the (a) donor-acceptor distances ($d_{D\text{-}A}$) and (b) proton-donor-acceptor angles (∠(H-D-A)) corresponding to the 12 monitored M^{pro}-substrate hydrogen bonds (**main text Figure 3.3**), over the 3 × 200 ns explicitly-solvated MD simulations performed on each of the 11 M^{pro}-substrate complexes. A HB is defined using the combined criteria of $d_{D\text{-}A} \leq 3.5$ Å and ∠(H-D-A) ≤ 30°.

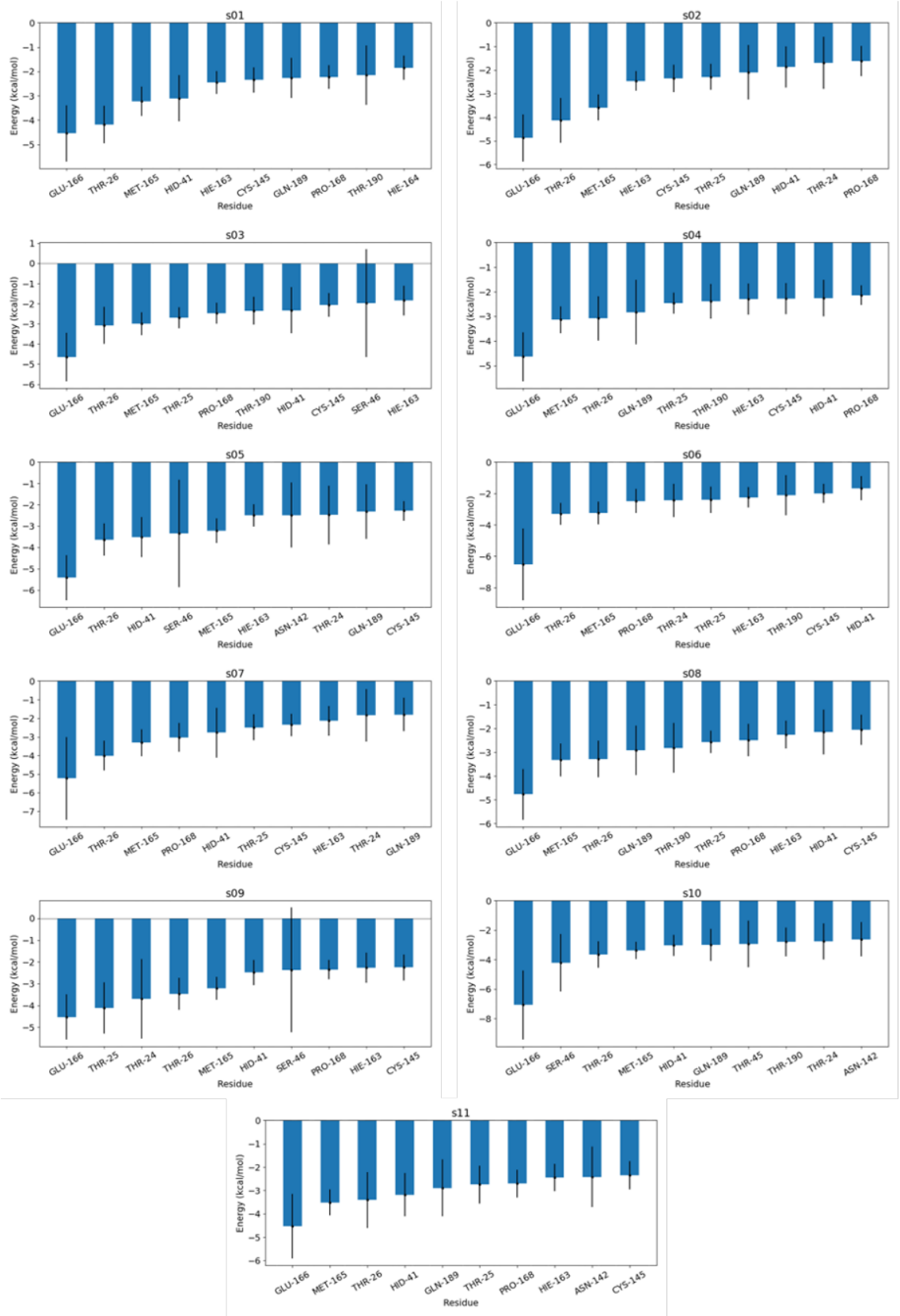


Figure S3.2.13: The ten M^{pro} residues which contribute most to the MM/GBSA binding energy (error bar = standard error of mean over 120 frames) for each of the 11 M^{pro}-substrate complexes.

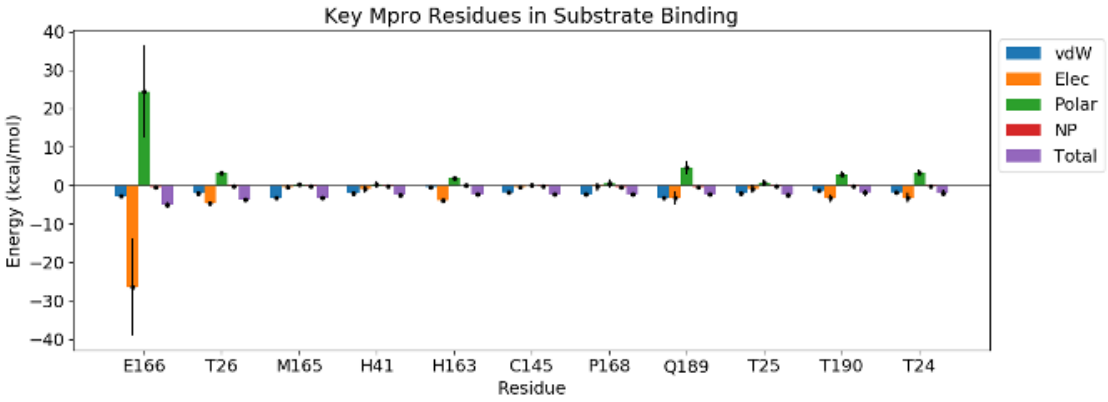


Figure S3.2.14: Contributions to the MM/GBSA binding energy by each hotspot residue (mean ± standard deviation across 11 systems), decomposed into the type of interactions: van der Waals (vdW), electrostatic (Elec), polar solvation from the generalised Born model (Polar), non-polar solvation from surface area calculation (NP), and total contribution. Given the varieties of charges and charged residue distributions across the substrates, there are relatively large variations in the electrostatic and polar solvation contributions by Glu166, which is the only charged residue out of the identified conserved hotspot residues on M^{pro}.

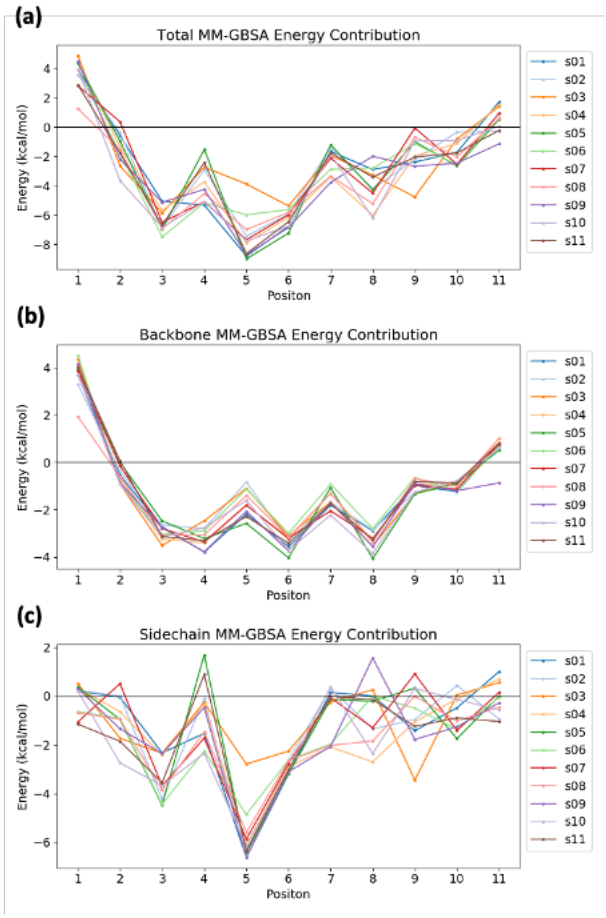


Figure S3.2.15: Contributions to the MM/GBSA binding energy for each residue of the substrate peptide, showing (a) the total contribution, as well as decomposition into (b) backbone and (c) sidechain contributions.

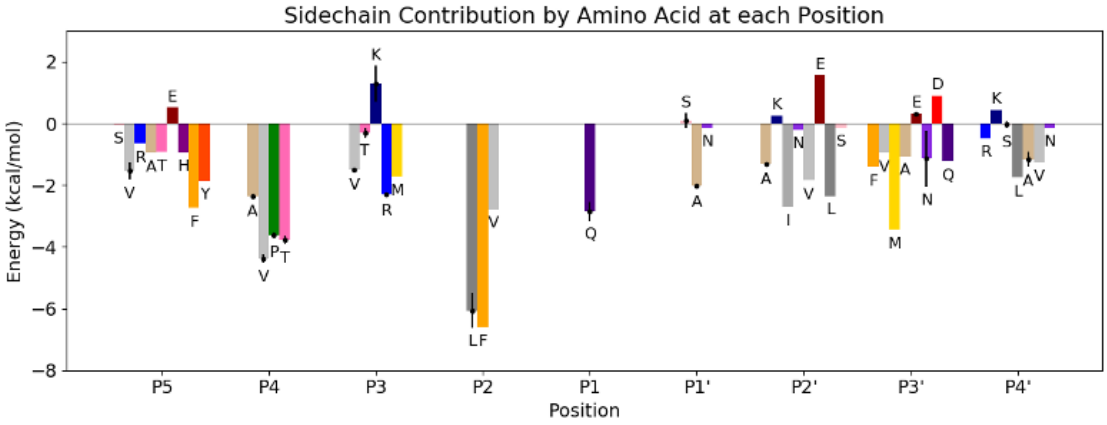


Figure S3.2.16: Average sidechain contributions to the MM/GBSA binding energy by each amino acid identity (except Gly) along the substrate peptide. Error bars (standard deviations across substrates) are shown only if the amino acid occurs in multiple sequences. Terminal residues are omitted due to their small contributions (**Figure S3.2.15**).

S3.2d Density functional theory analysis of the interaction network

Comparison of energetic panorama between DFT and force field. Below we compare the energetic panorama that is provided by DFT and the force field (FF) and show that the results correlate (supporting a reasonable MD configurational sampling). Enzyme-substrate interactions were analysed by post-processing QM results, which enable insight into the possible interaction patterns (or "interaction signatures") for each of the substrates. Note that the ability to treat the whole system with a consistent level of theory enables analysis of both short-range and long-range interactions on an equal footing.

Using this approach, energies were computed for the substrate (S), the protein dimer (P) and the total assembly (A) from which an interaction energy was calculated as follows: $E_{int} = E_A - (E_S + E_P)$; solvent molecules were omitted for easy comparisons. While absolute estimates of physical quantities were not our focus (due to the omission of entropy and solvent contributions), this approach provides an unbiased first-principles approach, complementary to that provided by classical FFs, to identify the main features involved in substrate-protein binding.

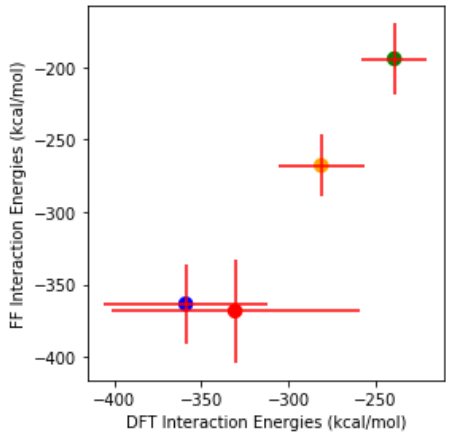


Figure S3.2.17: Comparison of the three-point interaction energies (kcal mol^{-1}) coming from full DFT calculation of the clustered MD snapshots. DFT interaction energies (x-axis) are represented together with FF three-point interaction energies (y-axis). We regroup the data on a per-peptide basis, for peptides s01 (blue), s02 (red), s05 (green), and p13 (orange). Statistic distributions are averaged taking into account the weights of the clusters in the trajectories.

The results (**Figure S3.2.17**) show a good correlation between the different approaches. The energetic panorama offered by the MD of the assemblies is likely similar to what would have been found by employing a first-principle approach, supporting a reasonable conformational sampling by MD. This also indicates that charge-polarisation, which would be captured (at least partially) by DFT, does not play a major role for the peptide-enzyme interaction. This is related to the fact that the most charged peptides are ones which exhibit more attractive interactions. Within this assumption the QM interaction energy can be efficiently approximated by the electrostatic peptide-enzyme interaction. Indeed, we show (**Figure S3.2.18**, left) that there is a remarkable correlation between the three-point interaction and such an approximated term (which is calculated by a multipole expansion of the DFT charge density result). To validate this, we performed the same analysis by employing the polarizable FF Polaris(MD),[42] the results of which show almost perfect correlation between the two FFs (**Figure S3.2.18**, right).

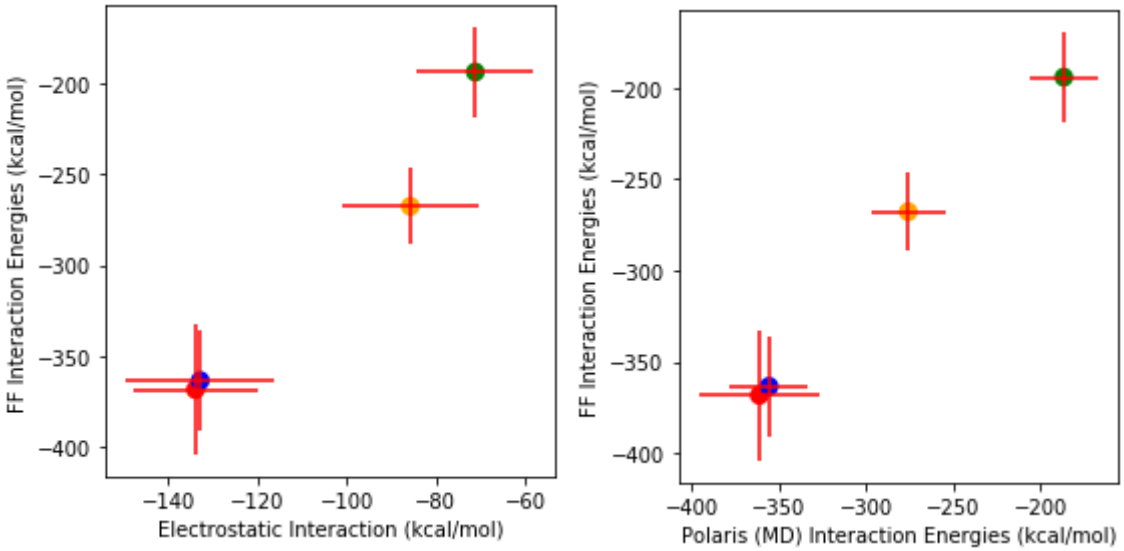


Figure S3.2.18: Comparison of the FF three-point interaction energies (y-axis) in the clustered MD snapshots, and on the x-axis (left) the electrostatic peptide-enzyme interaction or (right) interaction energy calculated with Polaris(MD). Each data point represents a peptide sequence: s01 (blue), s02 (red), s05 (green), and p13 (orange). Statistic distributions are averaged taking into account the weights of the clusters in the trajectories.

Note that the contribution to the interaction energy that is provided by van der Waals terms (the semi-empirical D3 dispersion term) is approximately constant (around 100 kcal mol^{-1}), and therefore the trends for the various interactions are provided by considering only the DFT-PBE contributions.

Contact map derived from electronic structure. A quantity of particular interest in determining the interaction network is the system's Hamiltonian, which gives a measurable indication of the strength of the chemical bonding among the electron clouds of each of the fragments. The partial traces of this operator, once projected on the system's fragments, provide an indication of the fragments which participate in bonding during the trajectories. Such "contact interaction energy", E_{cont}, can be interpreted as short-range sharing of electrons between two fragment residues. For each system, we have plotted this term for the ten most stabilising interactions (**Figure S3.2.19**).

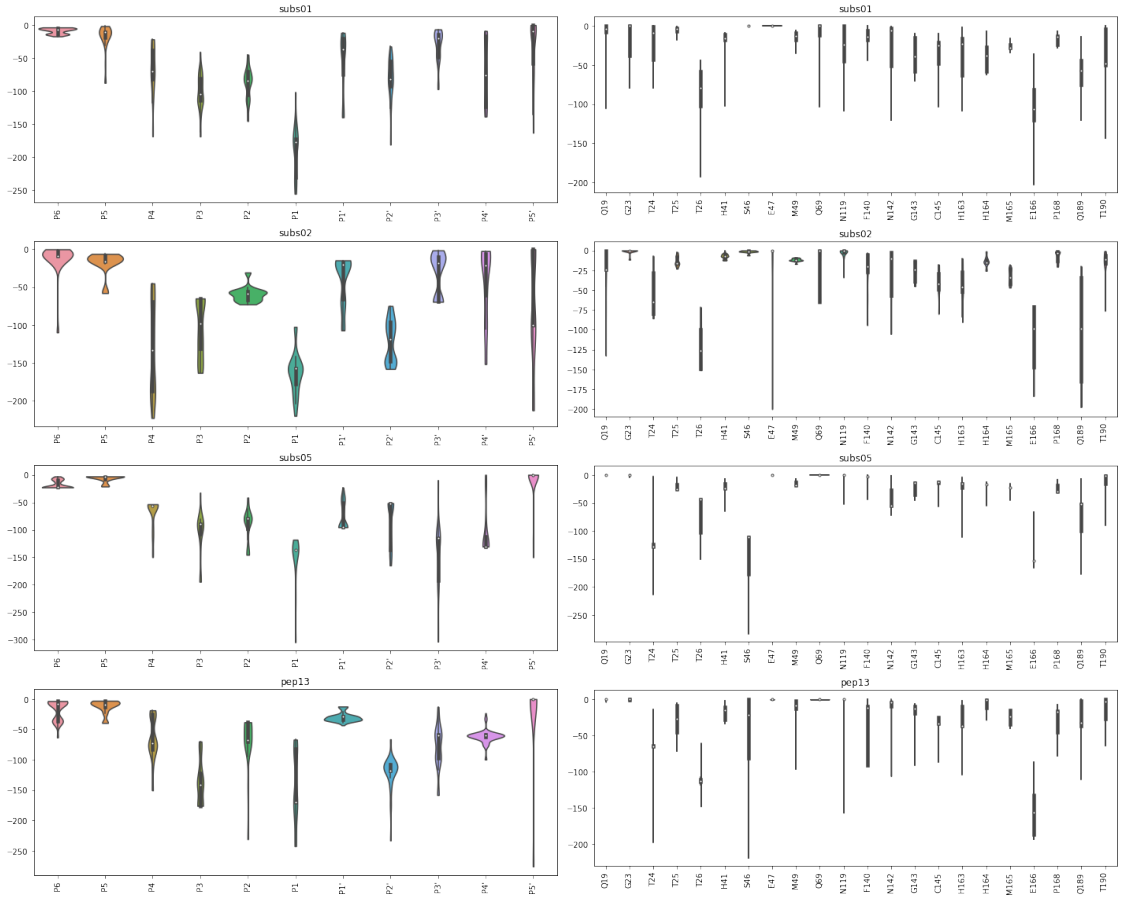


Figure S3.2.19: Violin plot distributions of contact interaction energies (kcal mol^{-1}) between substrate/designed peptides (s01, s02, s05, p13) and selected M^{pro} residues (which rank in the top ten in terms of stabilising interactions with the peptide in at least one of the systems), displayed (left) on the peptide or (right) on the enzyme residues.

Long-range electrostatic interaction patterns. The QM-FF comparison above (**Figure S3.2.18**, left) suggests that peptide binding trends can be estimated, at first approximation, by only considering the long-range electrostatic interactions. As we have employed a full QM calculation on the entire system it is interesting to show which are the emerging patterns of these interactions for the different peptides. This is helpful in defining other "interaction signatures" which are based on long-range patterns. Below (**Figure S3.2.20**) we show how such interactions behave during the dynamics. For clarity, we only represent residues whose magnitude of the (MD-averaged) interaction is larger than 7 kcal mol^{-1} for at least one of the systems. Some residues belonging to the other M^{pro} monomer (chain B) also appear to be relevant, due to their relative geographical proximity with the substrates. Obviously, the main overall pattern is dictated by the total charge of the bound peptide, which disfavours the neutral s05 in comparison with the other positively charged peptides (s01, s02, p13). The strong binding role of Glu166 is clearly visible.

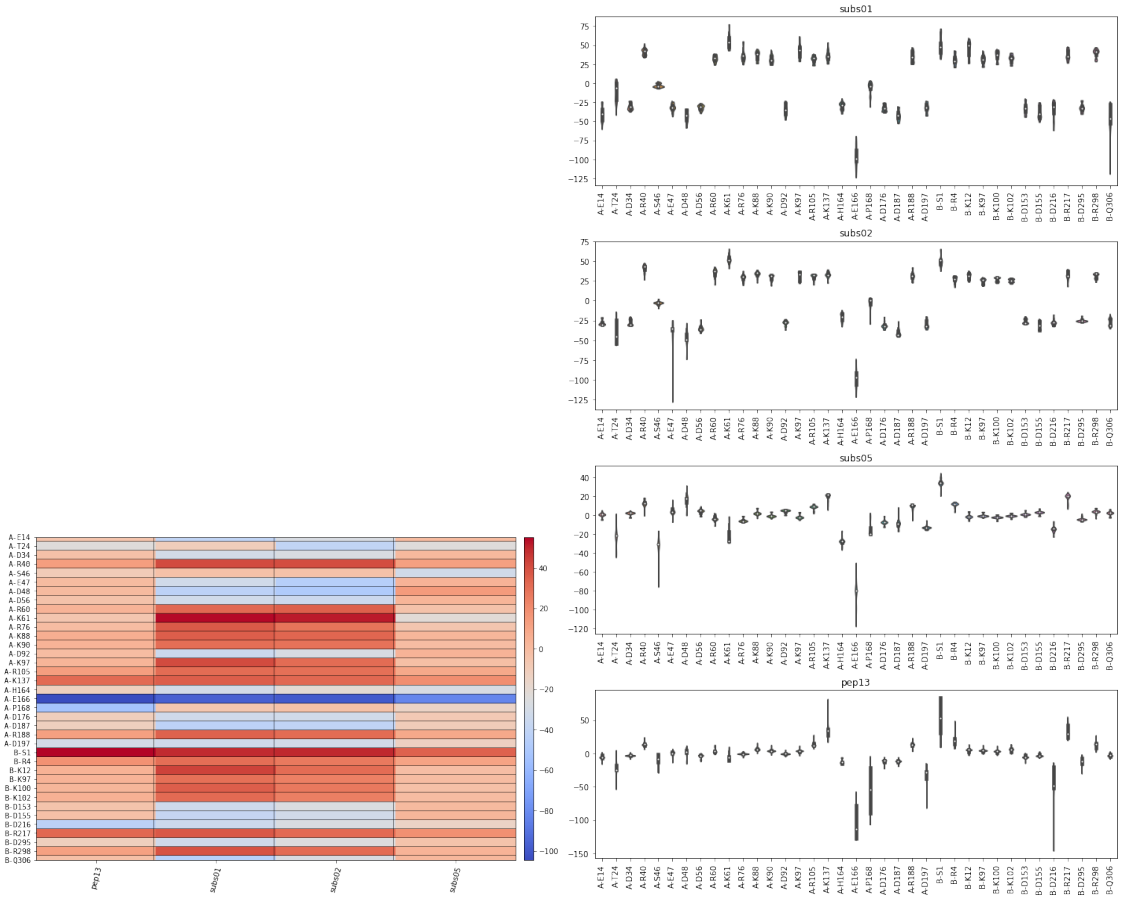


Figure S3.2.20: Distribution of long-range electrostatic interaction energies (kcal mol^{-1}) between substrate/designed peptides (s01, s02, s05, p13) and selected M^{pro} residues (which show a >7 kcal mol^{-1} interaction with the peptide in at least one of the systems), displayed as (left) a heatmap and (right) a box plot.

S3.2e Monitoring of substrate sequence hydrolysis by mass spectrometry

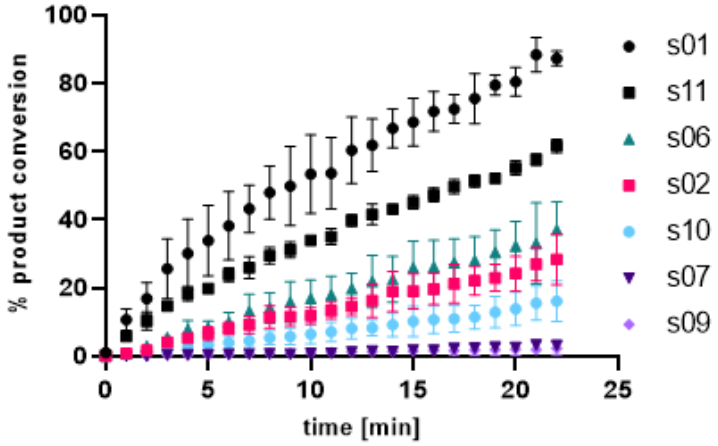


Figure S3.2.21: Ranking of catalysis of native M^{pro} substrate peptides s01-s11 using denaturing MS conditions. No evidence for cleavage was observed for s03, s04, s05 and s08 under these conditions. Conditions: 0.15 µM M^{pro}, 2 µM substrate peptide in 20 mM HEPES, pH 7.5, 50 mM NaCl.

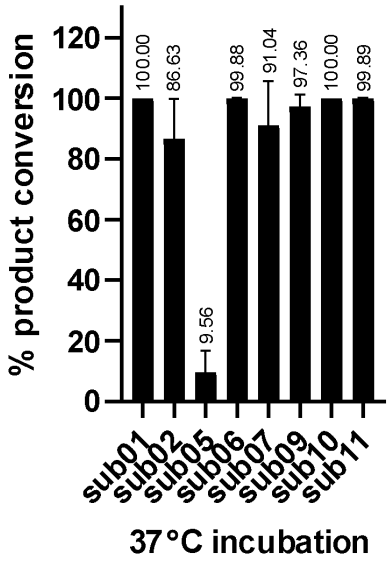


Figure S3.2.22: **Prolonged treatment of substrate peptides with M^{pro} at 37°C**. Note in particular the turnover of s05 as evidenced by observation of a peak corresponding to the +1 charge state of its N-terminal cleaved product; of the analysed substrates, this was the only product not detected at room temperature [20 °C] (**Figure S3.2.21**). Depletion of s04 was also observed, but its likely hydrophilic products were not observed potentially due to weak retention by the SPE C4 cartridge. Conditions: 0.15 µM M^{pro}, 2 µM substrate peptide in 20 mM HEPES, pH 7.5, 50 mM NaCl at 37°C, 300 rpm.

Section S3.3: Supplementary Results – Designed Peptide Inhibitors

S3.3a Synthesis and analysis of designed peptides

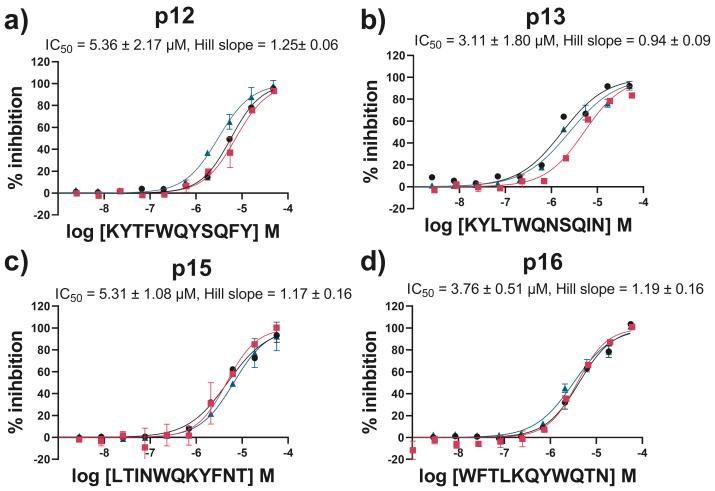


Figure S3.3.1: Designed peptides inhibit M^{pro}. IC_{50}s for a) p12, b) p13, c) p15 and d) p16. Reported IC_{50}s are means of independent repeats each composed of technical duplicates ($N = 3 \pm$ SD). Note: See **Methods Section S3.1e** for assay details. One of the independent repeats of p13 (black) comprised only a single dataset.

Table S3.3.1: The method of Wei *et al*. was used to investigate the mode of inhibition.[43] This approach involves determining the IC_{50} ratio (R) at two substrate concentrations (S1 and S2). The value of R is then used to assign the compound a (likely) competitive (R_1), non-competitive (R_2), uncompetitive (R_3) or mixed (R_4) inhibition mode. Using S1 = 2 µM and S2 = 40 µM (highlighted with a rectangle), the calculated R value for competitive inhibition (R_1) is 3.38. For the designed peptides (p12, p13, p15, p16), the results suggest competitive inhibition with respective R values of (3.61), (6.02), (4.56) and (4.85).

Mode of inhibition			**Competitive**	**Non-competitive**	**Uncompetitive**		**Mixed**
S1	**S2**	**K_M**	**R_1**	**R_2**	**R_3**	**α**	**R_4**
2	10	14	1.50	0.30	1.00	1.00	1.00
2	20	14	2.13	0.21	1.00	2.00	1.33
2	25	14	2.44	0.20	1.00	3.00	1.60
2	30	14	2.75	0.18	1.00	4.00	1.85
2	35	14	3.06	0.18	1.00	5.00	2.10
2	**40**	**14**	**3.38**	**0.17**	**1.00**	**6.00**	**2.34**
2	45	14	3.69	0.16	1.00	7.00	2.58
2	50	14	4.00	0.16	1.00	8.00	2.81
2	55	14	4.31	0.16	1.00	9.00	3.05
2	60	14	4.63	0.15	1.00	10.00	3.28
2	65	14	4.94	0.15	1.00	11.00	3.52
2	70	14	5.25	0.15	1.00	12.00	3.75
2	75	14	5.56	0.15	1.00	13.00	3.98
2	80	14	5.88	0.15	1.00	600.00	5.82

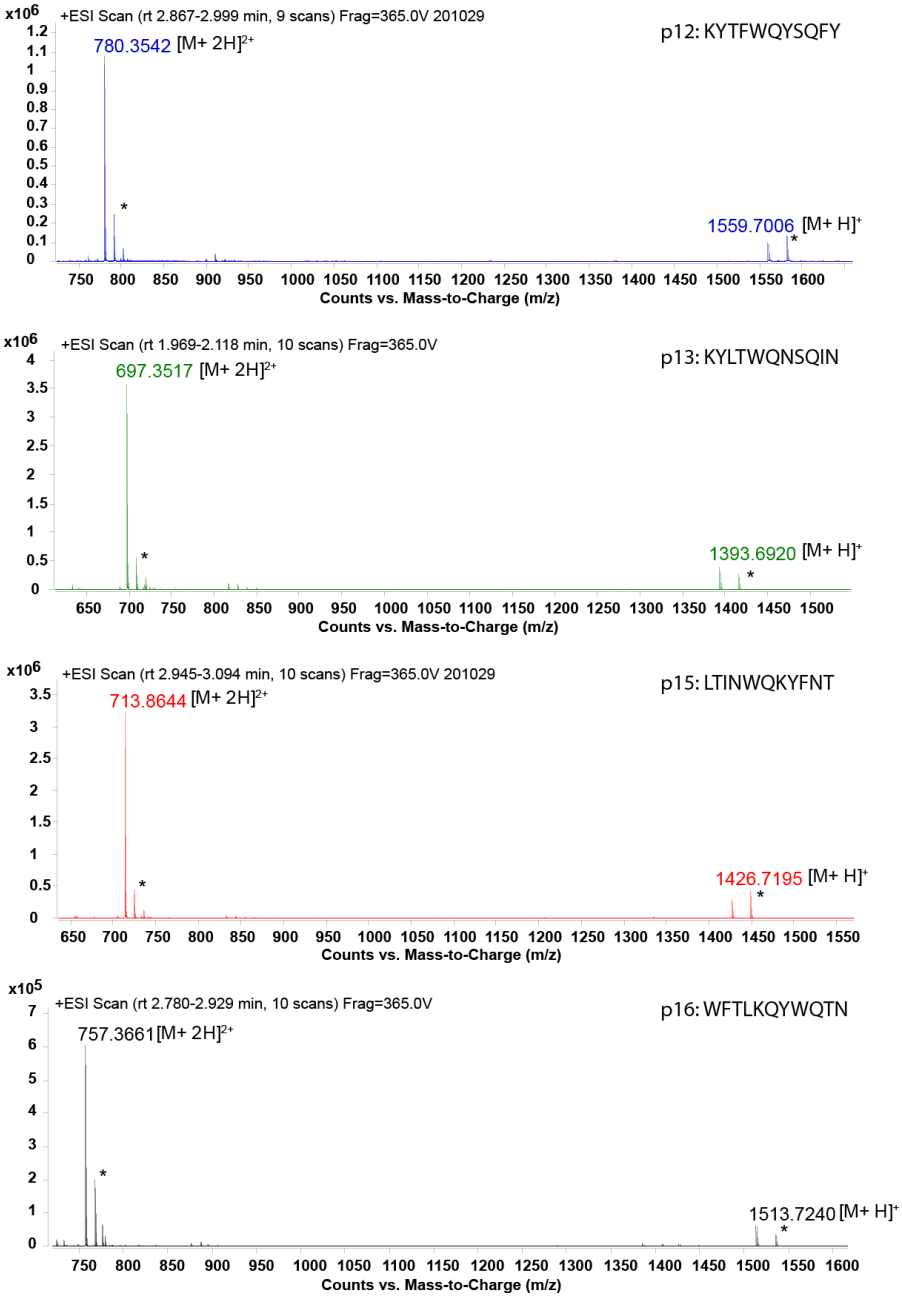


Figure S3.3.2: Potential cleavage products for the designed peptides (p12, p13, p15, p16) (2 µM) were not observed after overnight incubation with M^{pro} (0.15 µM) at ambient temperature. With positive mode ionisation, +2 and +1 charge states for all intact peptides were observed, with the former being the predominantly observed charge state. Sodiated adducts $[M+H+Na]^{+}$ and $[M+2H+Na]^{2+}$ for all peptides were observed, as indicated with an asterisk. See **Methods Section S3.1e** for details.

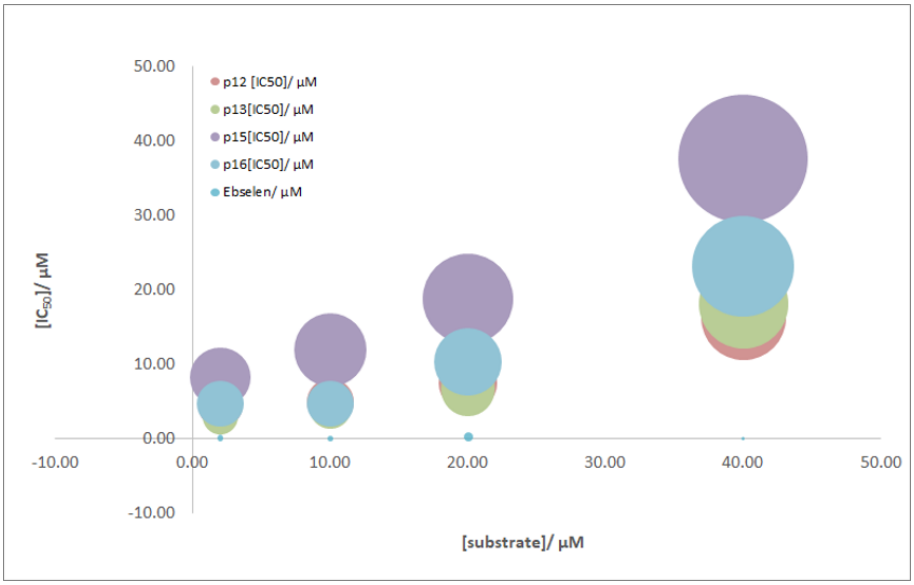


Figure S3.3.3: A linear increase in calculated IC_{50} values was observed in dose response studies with increasing substrate concentrations (2 µM, 10 µM, 20 µM and 40 µM). IC_{50} values are reported in **Table S3.3.2**.

Table S3.3.2: IC_{50} values in µM of designed peptides, and ebselen, at varying substrate (s01: TSAVLQ↓SGFRK-NH_2) concentrations of 2 µM, 10 µM, 20 µM and 40 µM. Each IC_{50} was calculated from the mean of technical duplicates.

[S]/ µM	p12 [IC_{50}]/ µM	p13 [IC_{50}]/ µM	p15 [IC_{50}]/ µM	p16 [IC_{50}]/ µM	Ebselen/ µM
2	4.53 ± 0.90	2.97 ± 0.21	8.26 ± 0.06	4.79 ± 0.08	0.09 ± 0.00
10	4.99 ± 0.22	4.28 ± 0.80	11.98 ± 0.34	6.66 ± 0.41	0.07 ± 0.00
20	7.49 ± 0.04	6.77 ± 0.46	18.83 ± 0.18	10.38 ± 0.53	0.21 ± 0.06
40	16.29 ± 0.52	18.09 ± 0.96	37.65 ± 2.02	23.20 ± 0.42	0.03 ± 0.06

S3.3b Explicitly-solvated MD and implicit solvent iMD-VR

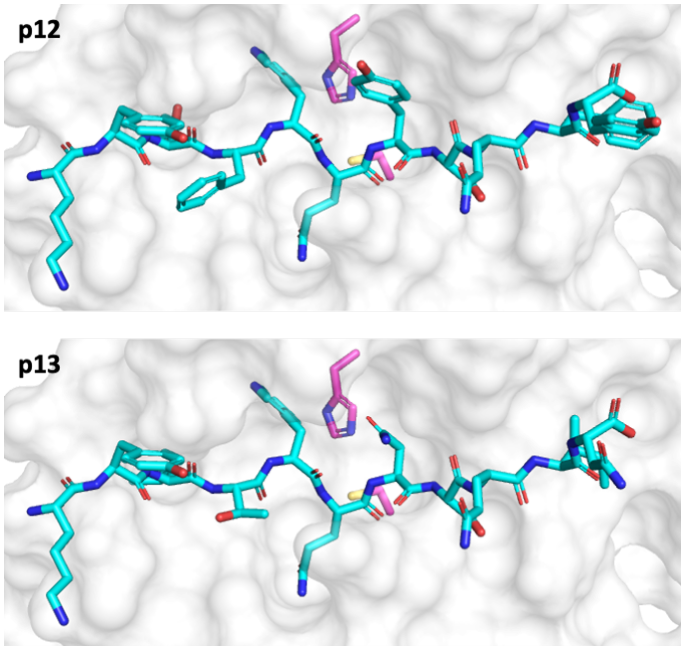


Figure S3.3.4: Starting configurations of the designed peptides p12 and p13 (cyan) in complex with SARS-CoV-2 M^{pro} (PDB 6YB7;[15] white surface with the catalytic dyad residues His41 and Cys145 in magenta), constructed by a comparative modelling approach based on the starting configuration of the M^{pro}-s02 complex.

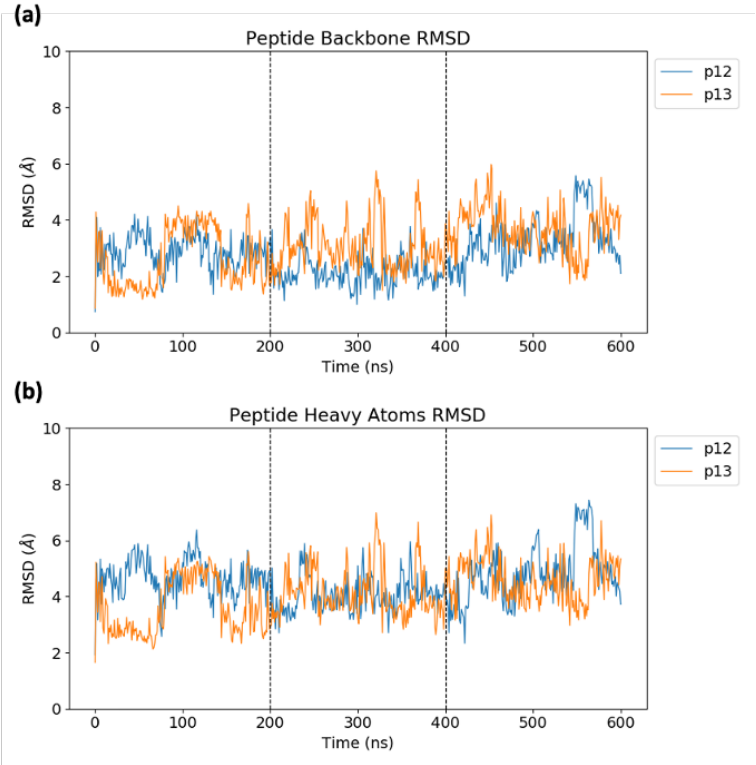


Figure S3.3.5: RMSD of (a) the peptide backbone (N, Cα, C) and (b) all peptide heavy atoms during the concatenated 3 × 200 ns explicitly-solvated MD simulations of M^{pro} in complex with p12 and p13 relative to the initial configuration, with trajectories fitted using the M^{pro} dimer backbone.

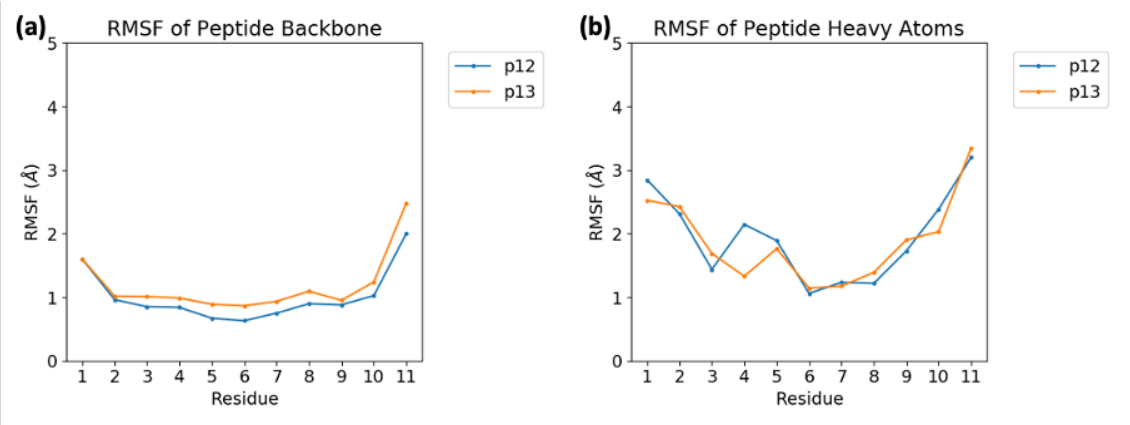


Figure S3.3.6: RMSF of (a) the peptide backbone (N, Cα, C) and (b) all peptide heavy atoms averaged per residue during the explicitly-solvated MD simulations of M^{pro} in complex with p12 and p13.

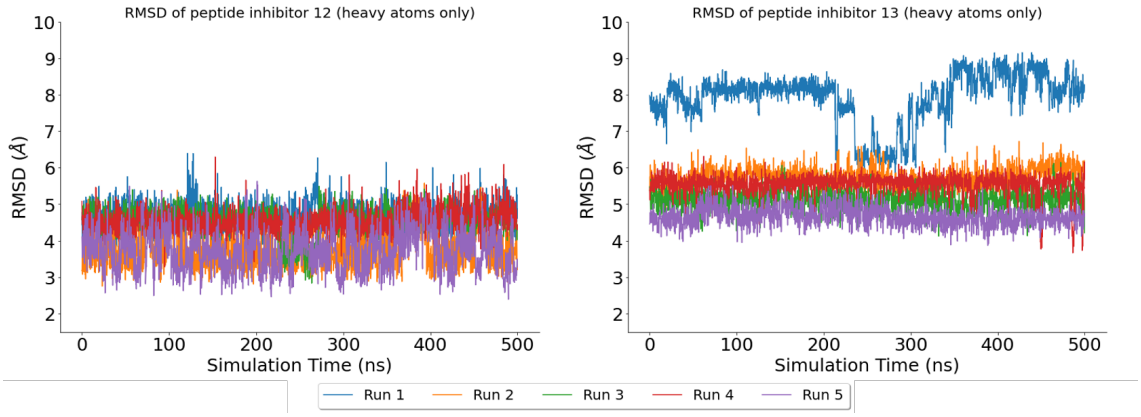


Figure S3.3.7: RMSD of p12 and p13 heavy atoms (*i.e.* not including hydrogens) across 5 × 500 ns MD simulations of iMD-VR docked structures, compared with the starting structure (*i.e.* the docked structure). The RMSD of p12 stays around 2.5–5.5 Å across all simulations, indicating that a common, stable bound structure has been found. By contrast, the RMSD of p13 is higher, with values ranging 4.5–8.5 Å.

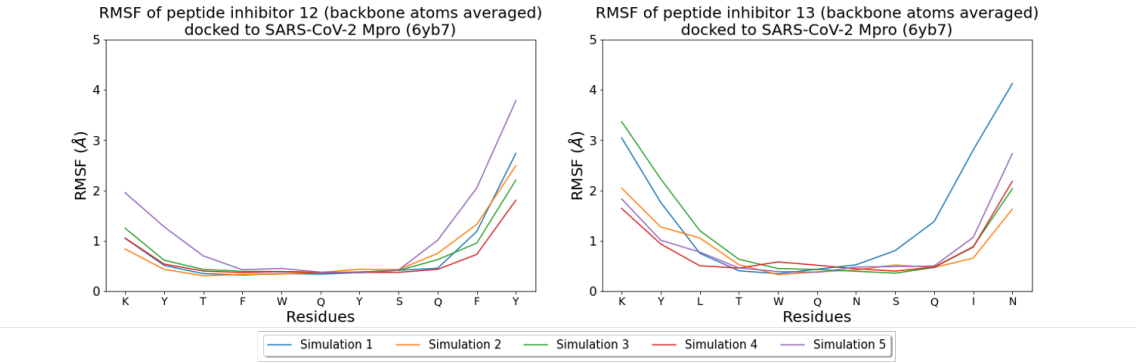


Figure S3.3.8: RMSF of p12 and p13 backbone atoms across 5 × 500 ns MD simulations of iMD-VR docked structures. In simulations of p12, the P′ side of the peptide (residues -YSQFY) is more flexible than the P side (residues KYTFWQ-). In simulations of p13, the P′ and P sides of the peptide seem equally flexible, with residues near the catalytic dyad (residues -WQN-) having a low RMSF.

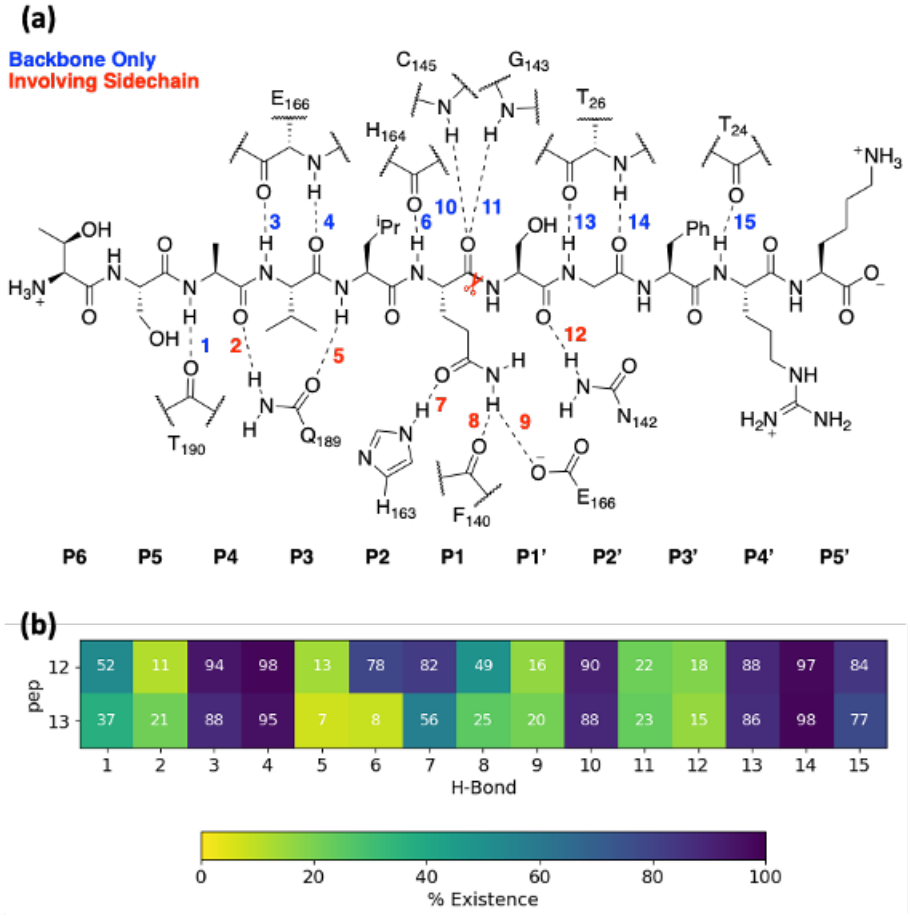


Figure S3.3.9: (a) Hydrogen bonds (HBs) analysed in the MD simulations of M^{pro} complexed with p12 and p13, with s01 illustrated as an example. (b) An annotated heatmap displaying the percentage to which each HB is observed in the simulations of M^{pro} in complex with p12 and p13 in explicit solvent. Frames were extracted every ns from 600 ns of cumulative MD conducted per system.

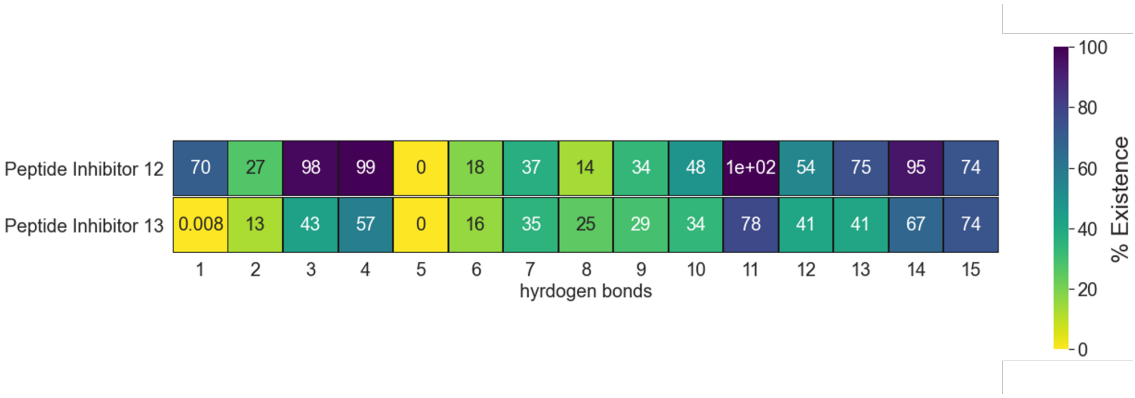


Figure S3.3.10: Averaged existence of HBs present throughout the 5 × 500 ns MD simulations of each of the iMD-VR-docked designed peptide-M^{pro} structures. Like substrates s01, s02 and s05 (**Figure S3.3.11**), the most well-maintained interactions in simulations of p12 are HBs 3, 4, 11, 13, and 14. By contrast, p13 does not maintain these HBs as well. HB 5 is not observed in any of the simulations.

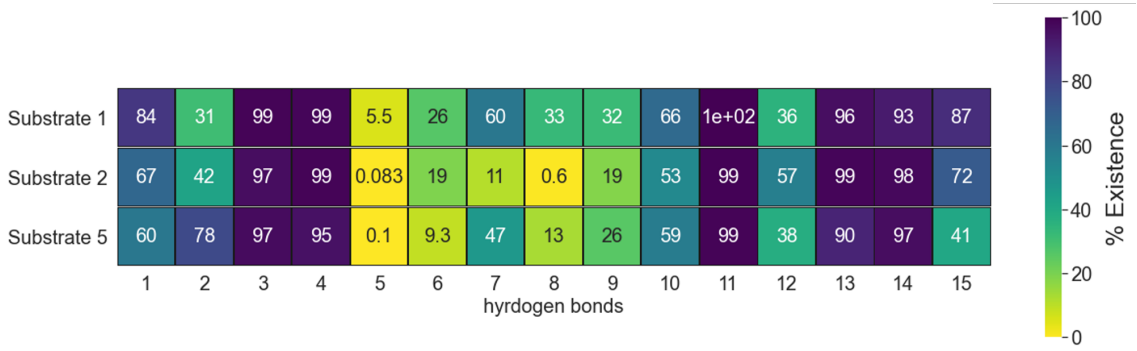


Figure S3.3.11: Averaged existence of HBs present throughout the 3 × 200 ns MD simulations of each of the iMD-VR-docked substrate-M^{pro} structures (s01, s02, s05) from iMD-VR, in comparison with the designed peptide-M^{pro} complexes (p12 and p13, **Figure S3.3.10**). The most well-maintained HBs across all three substrates considered here are HBs 3, 4, 11, 13, and 14.

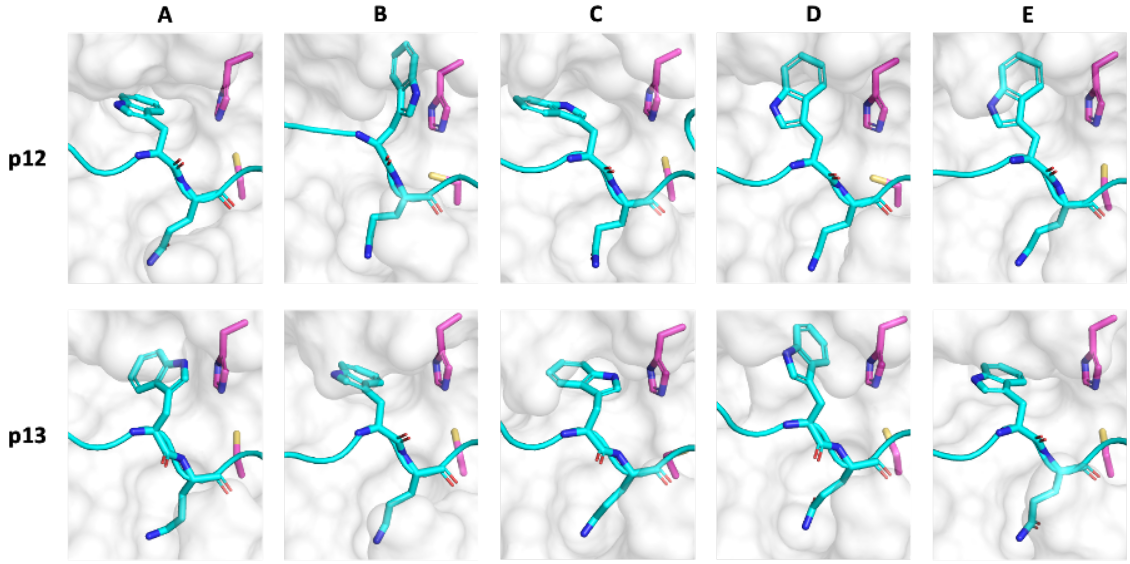


Figure S3.3.12: Conformations adopted by the P2 Trp side chain during MD simulations of M^{pro} complexed with p12 (top row) and p13 (bottom row) in explicit solvent, as visualised with representative snapshots obtained from RMSD clustering. Frames extracted every ns were aligned using the backbone of surrounding M^{pro} residues (His41, Thr45, Ser46, Met49, Pro52, Tyr54, Asn142, Cys145, His164, Met165, Glu166, Phe181, Val186, Asp187, Arg188, Gln189), and then clustered (gromos method, GROMACS v. 2019.2)[44] with a 2 Å cut-off on the RMSD of the P2 Trp side chain heavy atoms, resulting in 5 clusters for each of p12 and p13 (named A-E ranked by population; see **Figure S3.3.13**). M^{pro} is presented as a white surface, with His41 and Cys145 shown in magenta. Only the P2 and P1 residues on the peptide are shown with the rest of the peptide backbone displayed as a ribbon (cyan). Hydrogens are omitted for clarity. In certain conformations (*e.g.* p12 conformation B; p13 conformations A and C) the P2 Trp side chain is in close proximity to the M^{pro} residue His41 which forms part of the catalytic dyad.

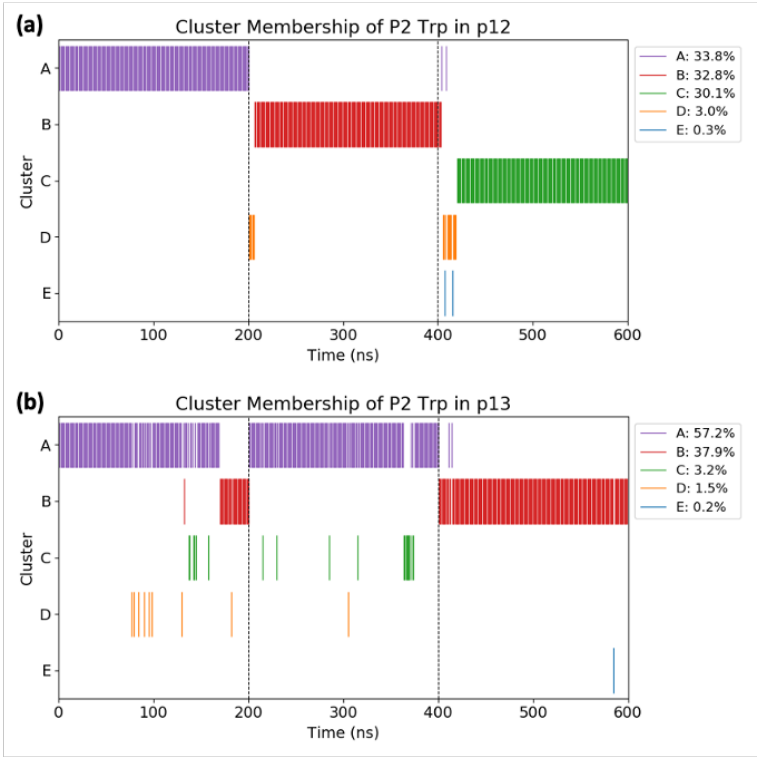


Figure S3.3.13: Evolution of the cluster membership (see **Figure S3.3.12** for representative structures) during the concatenated 3 × 200 ns MD simulations of M^{pro} in complex with (a) p12 and (b) p13 in explicit solvent. The percentage population of each cluster for each system is displayed in the legend.

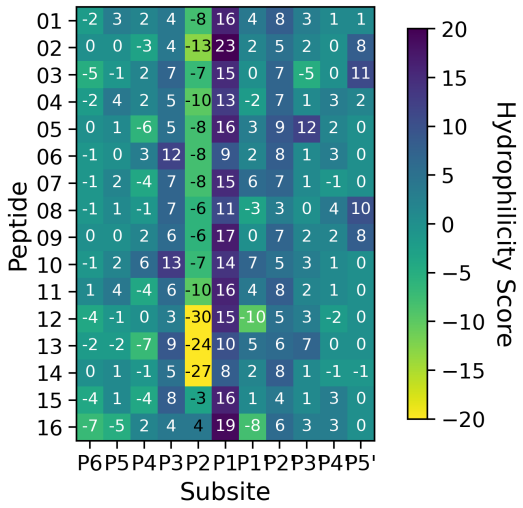


Figure S3.3.14: Hydrophilicity map of the eleven 11-mer substrate peptides (s01-s11) and the designed peptides (p12-p16). Hydrophilicity scores were calculated as a sum of all hydrophilic contacts subtracted by the sum of all hydrophobic contacts of each residue in the peptide, as identified by Arpeggio.[19] More positive scores correlate to hydrophilic sites and negative scores to hydrophobic pockets. Note the yellow at the P2 Trp for p12, p13, and p14 show that the P2 residue can become more deeply buried within S2 than the analogous P2 residues in the natural substrates, forming more than double the number of hydrophobic contacts in the cases of p12 and p13.

S3.3c Comparative peptide docking

Table S3.3.3: Details of the AutoDock CrankPep (ADCP) search space in Cartesian coordinates as identified by AutoSite (v. 1.1),[45] using the protein pdbqt file prepared with ADFRsuite from the corresponding PDB entry.

Receptor	**2Q6G (chain A)**	**7BQY (dimer)**	**7JOY (dimer)**
Centre (Å)	95.708 0.477 3.342	-10.000 0.189 -22.500	44.694 22.055 23.850
Box length (Å)	22.500 23.250 22.500	20.250 25.500 23.250	20.250 21.000 29.250
Size (0.375 Å spacing)	60 62 60	54 68 62	54 56 78
Number of fill points	76	82	97

Table S3.3.4: Assessment of the 10 highest ranked binding poses from ADCP redocking of the s01 sequence in the crystal structure of H41A SARS-CoV M^{pro} (PDB 2Q6G, chain A).[1] These poses are compared with the original position of the peptide by measuring the positional deviations of the 11 peptide Cα atoms. A pose is considered to pass the filter if the deviation is lower than 2 Å for at least three Cα atoms.

Rank	Score (kcal/mol)	Cluster Size	Cα Deviation (Å) at Position											Filter	Cα RMSD (Å)
			1	2	3	4	5	6	7	8	9	10	11		
1	-23.0	34	1.97	0.86	0.64	0.54	0.51	0.63	0.78	3.30	5.51	11.88	16.27	Y	6.42
2	-21.5	9	2.67	1.27	0.96	1.08	0.81	0.63	1.53	5.36	5.96	13.13	15.44	Y	6.67
3	-21.5	56	10.09	7.43	9.61	9.62	13.41	12.67	12.99	12.34	12.33	14.52	13.99	N	11.92
4	-21.3	10	5.38	1.43	0.75	0.70	0.70	0.60	1.46	5.07	4.42	4.37	6.31	Y	3.56
5	-21.0	14	2.06	0.56	0.29	0.50	0.49	0.68	1.48	4.07	5.95	12.99	16.06	Y	6.65
6	-20.6	29	12.95	12.54	12.24	10.68	9.29	6.19	9.27	12.97	15.77	21.98	24.28	N	14.42
7	-20.5	15	18.31	12.78	6.37	1.23	6.30	11.08	15.61	18.46	17.80	19.15	16.65	N	14.30
8	-20.5	21	5.66	1.22	0.81	0.84	0.48	0.19	1.20	6.05	4.47	5.40	3.49	Y	3.50
9	-20.5	22	16.08	12.80	10.71	6.98	10.03	9.77	10.03	9.31	9.76	9.68	9.88	N	10.69
10	-20.3	25	14.76	11.41	9.30	6.81	7.17	5.02	5.52	5.98	6.16	11.39	16.88	N	9.89

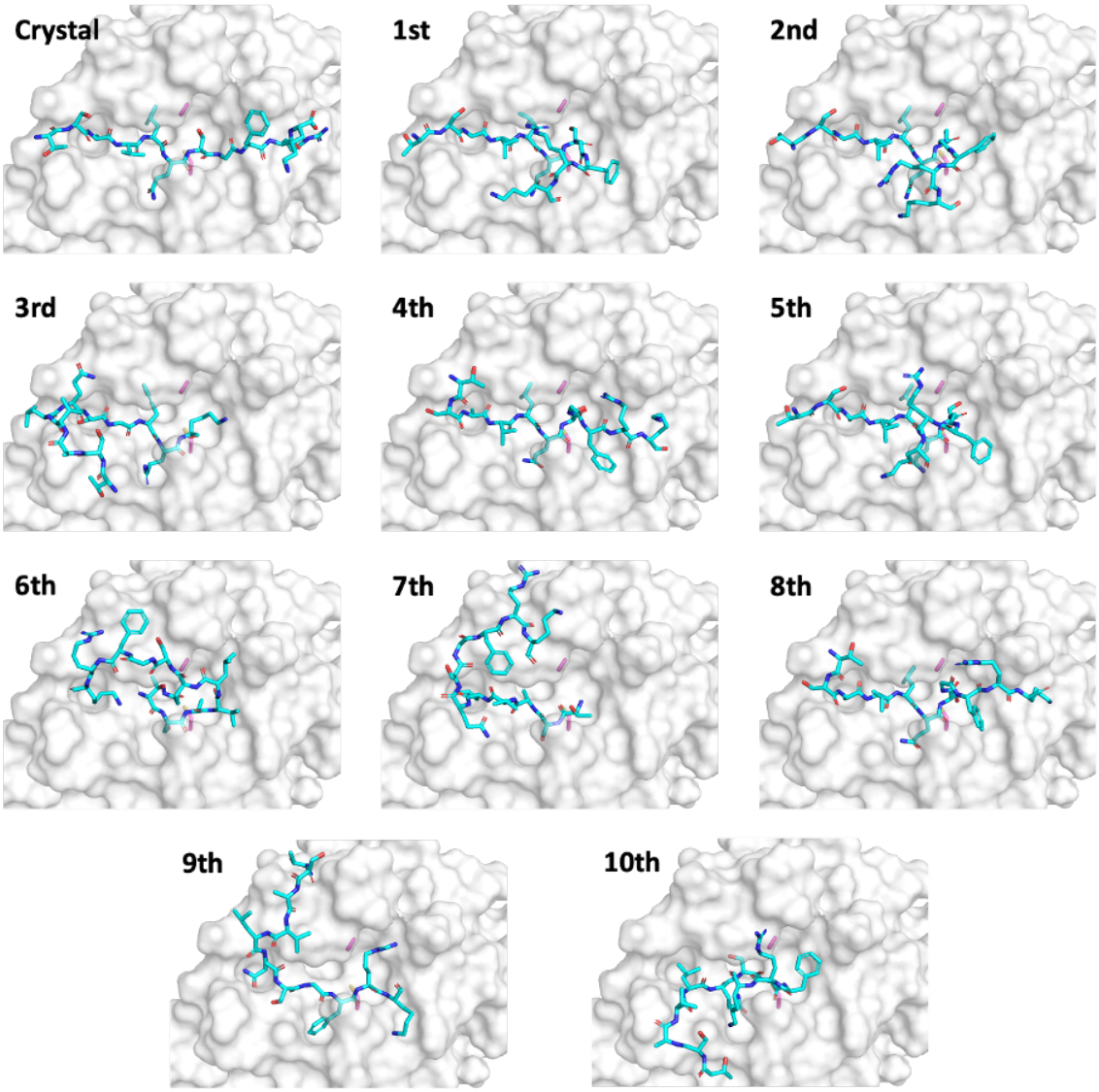


Figure S3.3.15: The 10 highest ranked binding poses from ADCP redocking of the s01 sequence (cyan) in the crystal structure of H41A SARS-CoV M^{pro} (PDB 2Q6G), compared to the original position in the crystal structure (top left).[1] Ala41 (in place of His that is present in the catalytically competent protein) and Cys145 are in magenta.

Table S3.3.5: Assessment of selected binding poses from ADCP docking of each substrate (s01-s11) and designed (p12, p13, p15, p16) sequence in the crystal structure of SARS-CoV-2 M^{pro}, originally in complex with the N3 inhibitor (PDB 7BQY).[39] For each peptide sequence, the 10 highest ranked solutions were evaluated by comparison to the M^{pro}-peptide complex resulting from comparative modelling followed by MM minimisation. For every sequence, the highest ranked solution that passed the filter of < 2 Å deviation in at least three Cα atoms, or if none of the solutions passed the filter, the solution with the lowest Cα RMSD, was presented below.

Seq	Pose Rank	Score (kcal/mol)	Cluster Size	Cα Deviation (Å) at Position											Filter	Cα RMSD (Å)
				1	2	3	4	5	6	7	8	9	10	11		
01	1	-22	55	2.19	1.49	1.16	1.60	1.67	2.36	7.95	9.34	13.47	19.87	24.00	Y	10.96
02	6	-19.9	19	7.32	3.87	1.60	2.25	3.23	2.85	3.99	3.59	4.21	4.56	7.58	N	4.47
03	7	-20.5	27	5.83	3.10	1.31	1.52	1.37	1.38	7.22	10.34	13.72	20.06	23.77	Y	11.14
04	1	-21	65	5.62	1.42	0.91	0.91	0.80	0.85	0.87	2.97	4.46	6.28	5.53	Y	3.52
05	7	-21.1	10	5.24	3.29	1.03	1.76	0.99	0.40	0.58	0.87	3.11	5.93	8.52	Y	3.84
06	1	-21	15	1.96	1.26	0.87	1.46	0.81	0.77	0.84	0.94	4.86	10.76	9.86	Y	4.75
07	2	-19.7	25	0.75	0.94	1.25	2.32	1.17	0.84	0.24	0.64	5.15	10.78	10.89	Y	4.98
08	1	-20.3	33	5.64	2.98	1.01	1.07	0.39	0.42	0.45	1.31	3.20	10.20	13.65	Y	5.61
09	2	-20.3	27	7.94	1.84	1.42	1.71	2.82	2.60	7.45	9.81	12.79	18.27	20.35	Y	10.22
10	4	-21	26	2.82	1.66	0.92	1.03	1.60	3.84	4.74	5.49	2.01	2.09	2.06	Y	2.94
11	1	-22.5	55	2.45	2.07	1.29	1.84	1.94	2.11	7.87	11.78	14.77	21.15	25.07	Y	11.75
12	1	-26	39	12.02	9.00	4.42	1.87	8.42	10.96	10.80	8.39	6.22	7.43	13.85	N	9.11
13	9	-23.3	6	10.24	7.29	7.43	6.88	6.79	7.65	5.91	5.70	10.52	15.64	19.69	N	10.34
15	1	-24	46	10.46	5.99	4.72	3.51	4.13	4.74	4.81	3.73	3.89	4.16	6.90	N	5.53
16	5	-24	32	8.09	3.32	4.37	1.95	0.91	1.17	2.40	3.84	5.64	13.00	15.83	Y	7.22

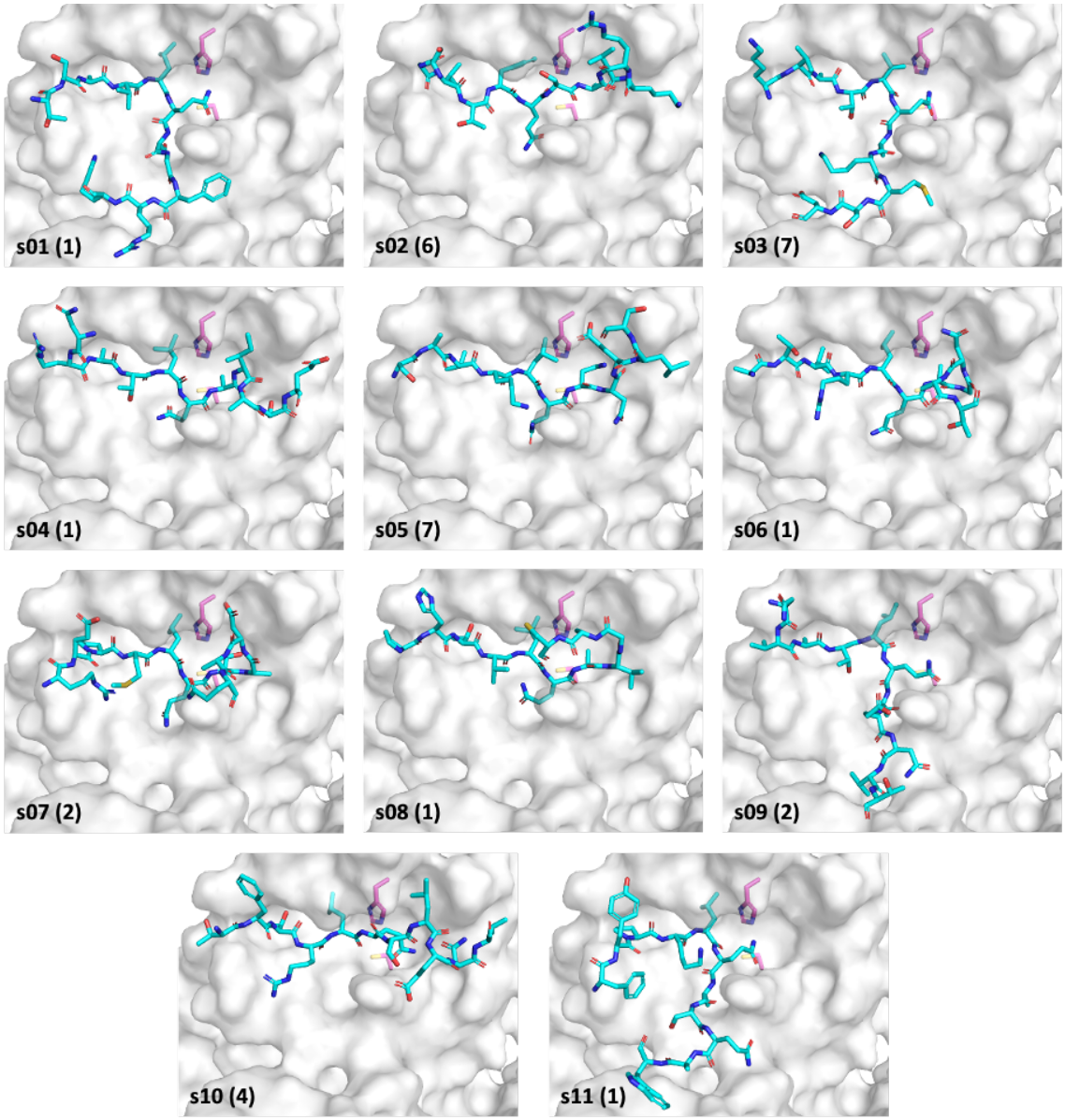


Figure S3.3.16: Selected binding poses (see **Table S3.3.5**; number in parentheses = pose rank) from ADCP docking of each native substrate (s01-s11) sequence in the crystal structure of SARS-CoV-2 M^{pro}, originally in complex with the N3 inhibitor (PDB 7BQY). The catalytic dyad His41 and Cys145 are in magenta. Docked structures with the P4 and P2 residues positioned correctly in their respective S4 and S2 pockets were consistently obtained. However, greater variation was observed in the positioning of residues from P1 onwards, likely due to the poorer definition of the S′ subsites.

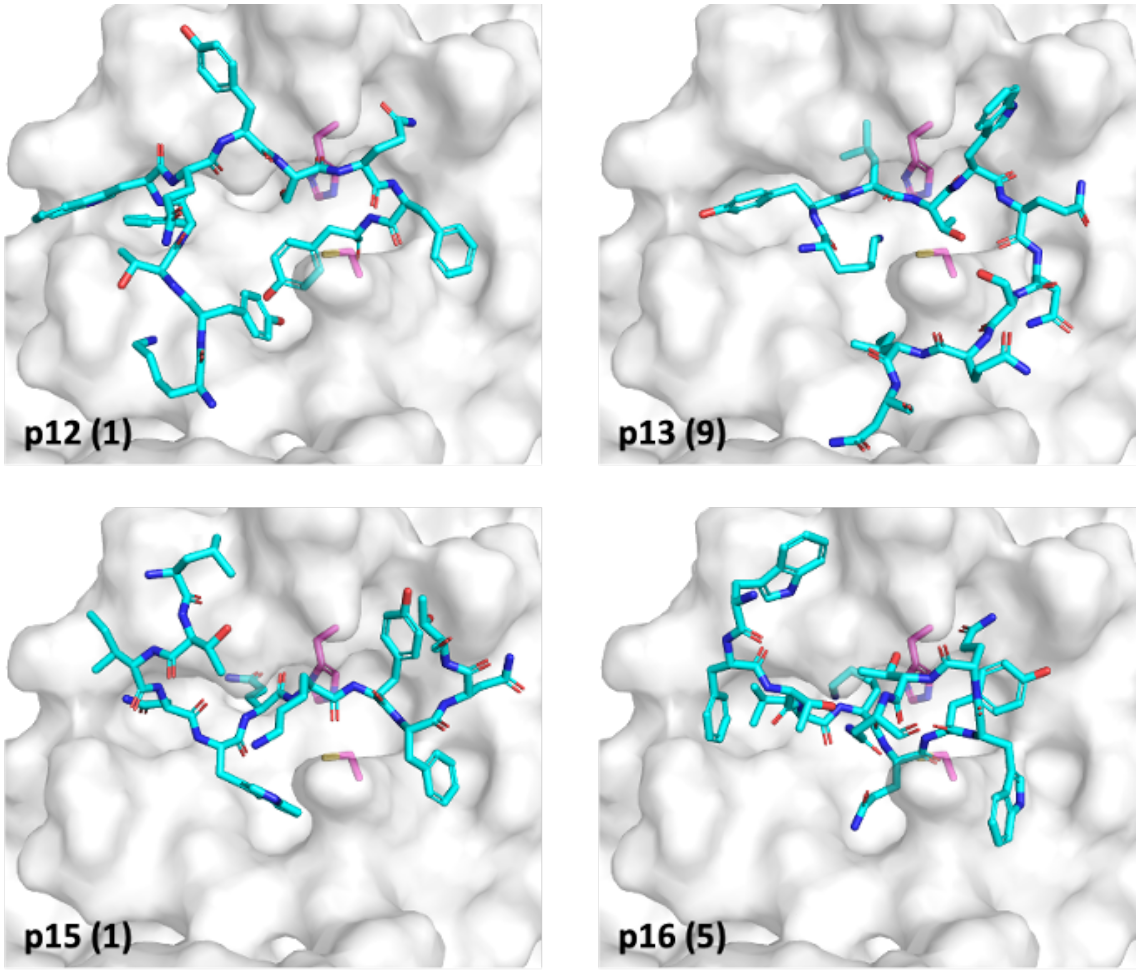


Figure S3.3.17: Selected binding poses (see **Table S3.3.5**; number in parentheses = pose rank) from ADCP docking of each designed (p12, p13, p15, p16) sequence in the crystal structure of SARS-CoV-2 M^{pro}, originally in complex with the N3 inhibitor (PDB 7BQY). The catalytic dyad His41 and Cys145 are in magenta.

Table S3.3.6: Assessment of selected binding poses from ADCP docking of each designed (p12, p13, p15, p16) sequence in the crystal structure of C145A SARS-CoV-2 M^{pro}, originally in complex with the s02 cleaved product (PDB 7JOY).[46] For each peptide sequence, the 10 highest ranked solutions were evaluated by comparison to the M^{pro}-peptide complex resulting from comparative modelling followed by MM minimisation. The highest ranked solution that passed the filter of < 2 Å deviation in at least three Cα atoms, or if none of the solutions passed the filter, the pose with the lowest Cα RMSD, was presented below.

Seq	Pose Rank	Score (kcal/mol)	Cluster Size	Cα Deviation (Å) at Position											Filter	Cα RMSD (Å)
				1	2	3	4	5	6	7	8	9	10	11		
12	1	-29.6	63	5.75	1.95	1.83	1.20	0.99	1.06	1.46	1.80	3.81	4.99	6.94	Y	3.53
13	4	-23.3	35	8.06	4.05	4.76	4.22	3.83	1.74	3.62	3.22	4.36	5.04	6.80	N	4.80
15	2	-25.7	29	4.23	5.33	4.02	4.61	4.58	3.87	3.76	4.37	3.89	7.05	5.78	N	4.78
16	1	-27.7	69	4.57	1.60	1.85	0.79	0.65	0.85	1.06	1.62	3.94	5.64	6.84	Y	3.39

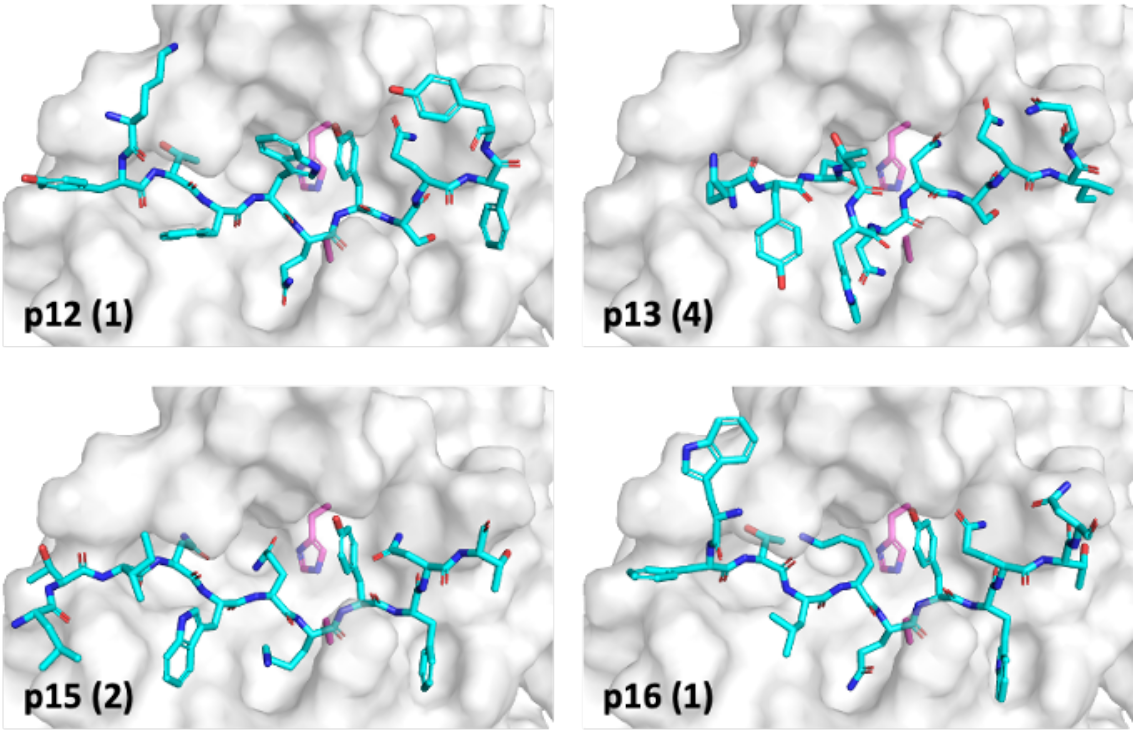


Figure S3.3.18: Selected binding poses (see **Table S3.3.6**; number in parentheses = pose rank) from ADCP docking of each designed (p12, p13, p15, p16) sequence in the crystal structure of C145A SARS-CoV-2 M^{pro}, originally in complex with the s02 product (PDB 7JOY). The catalytic dyad His41 and Ala145 (in place of Cys145 that is present in the catalytically competent protein) are in magenta.

Section S3.4: Supplementary Results – Analysis of Results from Fragment Crystallography

S3.4a Interaction analysis of the XChem fragments

Fragment clustering by interaction fingerprints

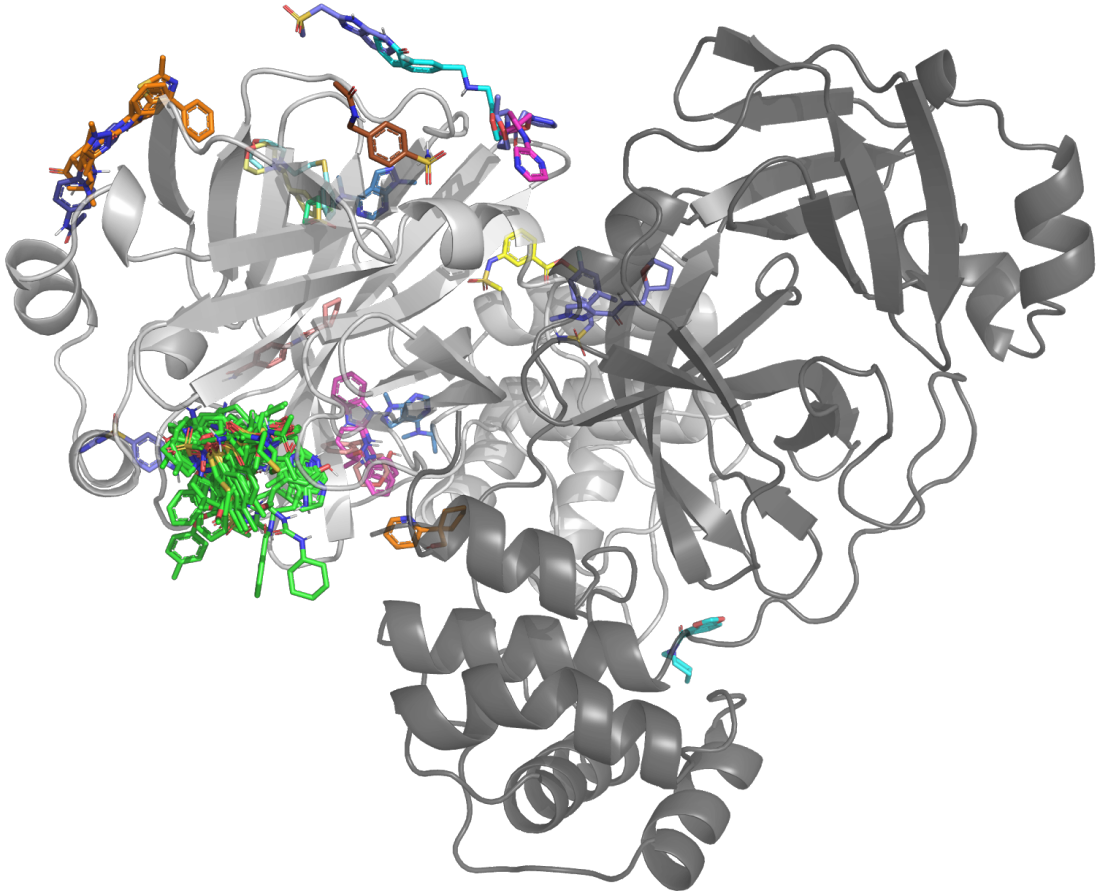


Figure S3.4.1: Views from crystal structures of all XChem fragments and their binding site on the M^{pro} dimer.[23] All binding sites are on chain A (white). As a representative structure the fragment x0830 co-crystal structure was used. There are 66 fragments that bind into the active site (green fragments) and 25 allosteric fragments (fragment 1101 binds in two different allosteric sites).

A fingerprint bit-vector was constructed for every active-site binding fragment, with each bit denoting the presence or absence of an interaction with every M^{pro} protein residue found to interact with either a substrate or fragment. A distance matrix was created by calculating the Tanimoto distance[24] between interaction fingerprints. Fragments were then clustered using the interaction Tanimoto similarity index. A similarity score of 1 corresponds to perfect overlap of all fragment-residue level contacts, and 0 denotes no overlap. Note that the Tanimoto index is computed between Arpeggio-derived contact fingerprints rather than ligand structure extended-connectivity fingerprints (ECFPs). For this analysis, a tighter (0.7) and broader (0.5) threshold were chosen. The former was found to be better at distinguishing distinct binding modes, while the broader threshold was more useful in grouping binders around major significant interactions (see **Table S3.4.1** and **Figures S3.4.2-3** for further details).

Table S3.4.1: Relationship between Tanimoto similarity threshold and cluster sizes for the XChem fragment crystal structures. Only active-site binders were considered.

Tanimoto threshold	Number of clusters	Number of single molecule clusters	Average cluster size
0.1	2	0	33.0
0.2	2	0	33.0
0.3	5	1	13.2
0.4	9	0	7.3
0.5	11	2	6.0
0.6	16	4	4.1
0.7	29	20	2.3
0.8	37	27	1.8
0.9	53	46	1.2

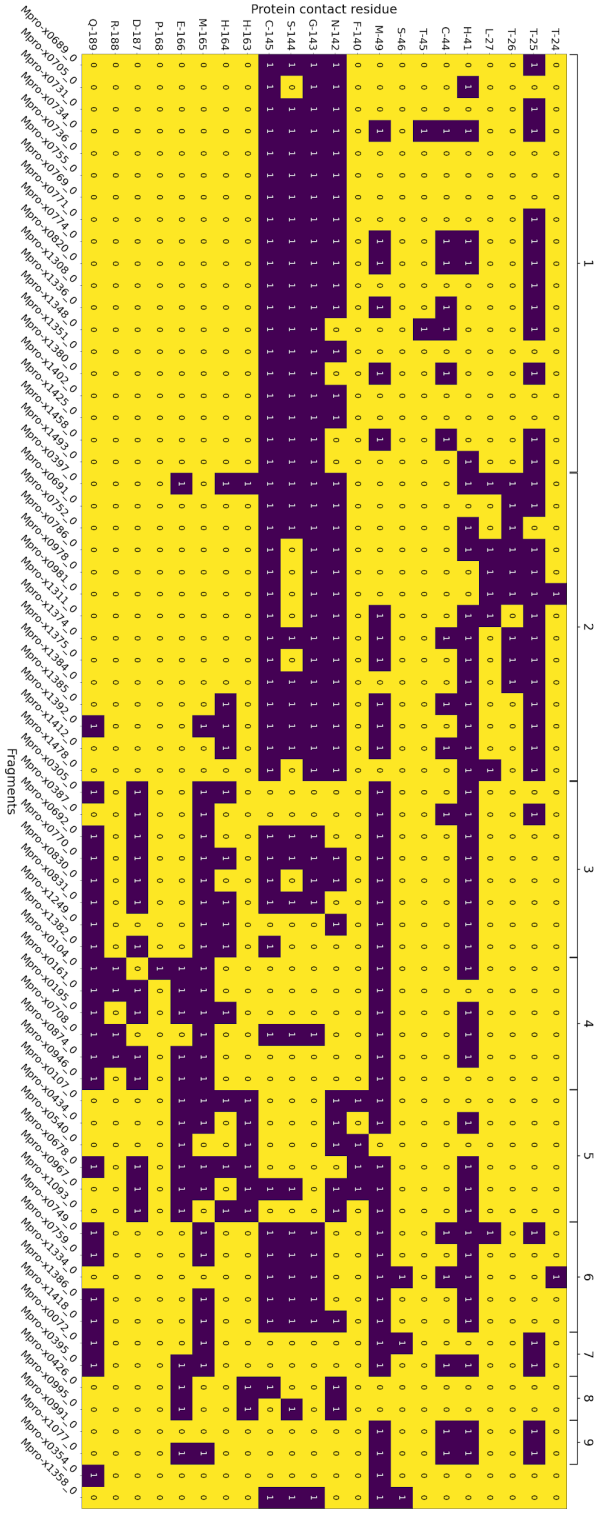


Figure S3.4.2: Contact matrix for the 66 active site XChem fragments sorted by their assigned cluster (threshold 0.5) based on the protein contacts which are indicated by "1". Clusters 1 through 9 are indicated by brackets. Clusters with more than 1 molecule are marked by brackets (except x0354 and x1358).

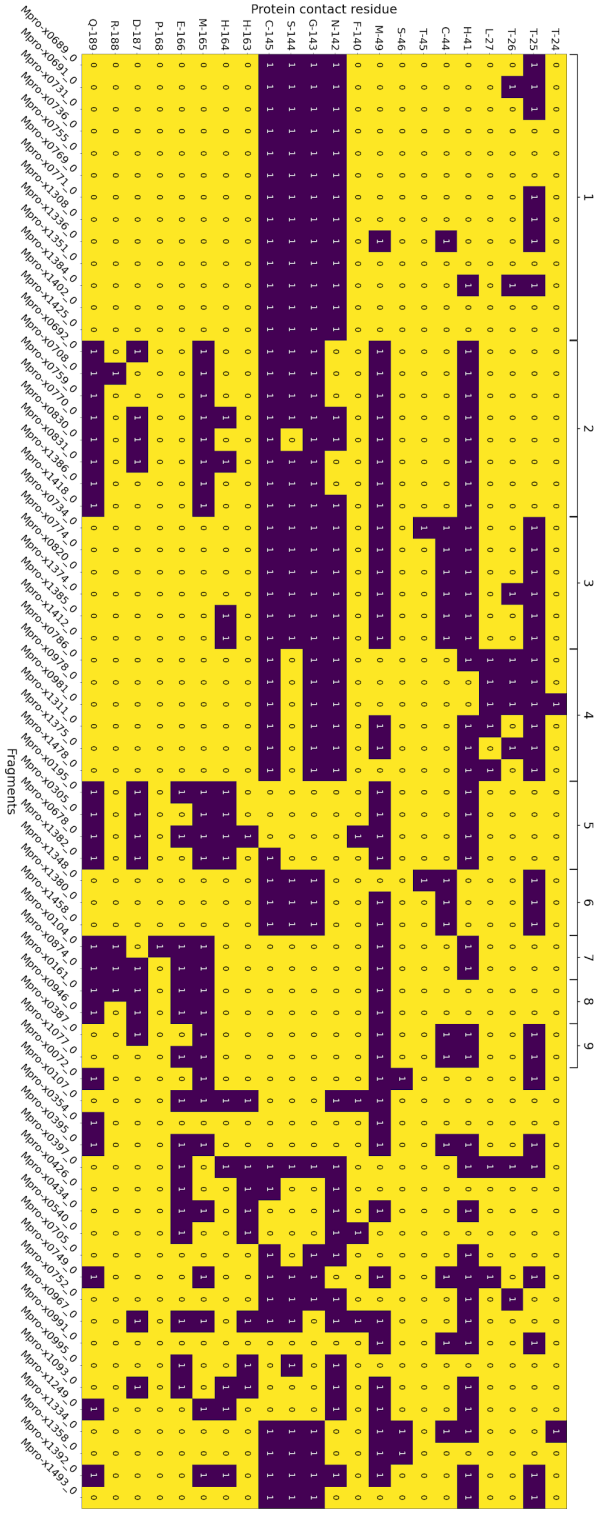


Figure S3.4.3: Contact matrix for the 66 active site XChem fragments sorted by their assigned cluster (threshold 0.7) based on the protein contacts which are indicated by "1". Clusters 1 through 9 are indicated by brackets. Only clusters with more than 1 molecule are marked by brackets.

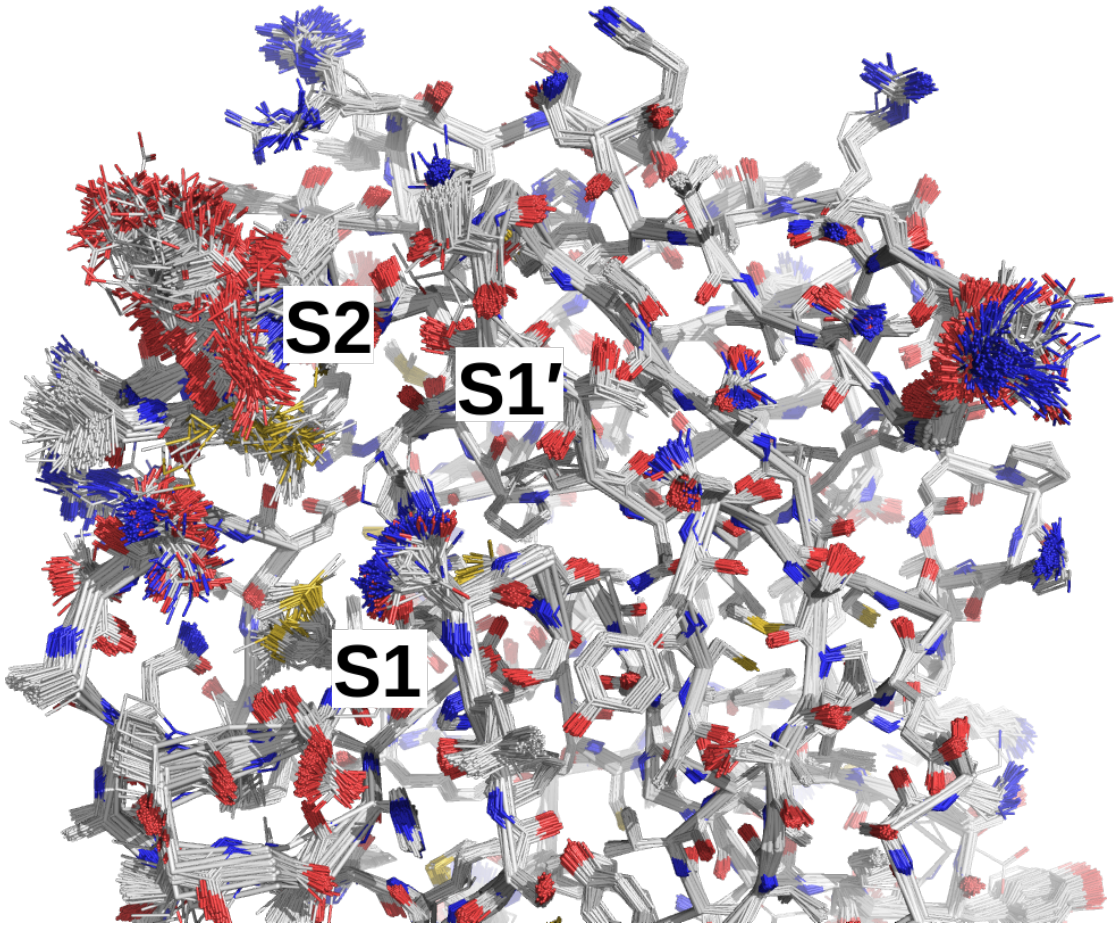


Figure S3.4.4: Overlay of the 333 M^{pro} co-crystal structures published by Fragalysis.[30] Subsite S2 shows extremely large changes in conformation between structures, while S1 is more structurally conserved.

Among the clusters using the 0.5 threshold clustering method (**Figure S3.4.5**), cluster 5 stands out as it the only major cluster with fragments that bind deeply into the S1 pocket, having one of the main conserved contacts identified for the substrate peptides. This cluster shows a distinct binding motif primarily driven by (i) hydrogen bonding between a carbonyl oxygen on the fragment and the Glu166 backbone NH-group, and (ii) a strong polar interaction between His163 and the fragment. Notably, the position of the hydrogen on the imidazole sidechain of His163 appears to depend on the HB fragment partner.

Based on the presence or absence of either a HB donor or acceptor on the fragment, the protonation state of His163 can be inferred. This suggests that for x0107, x0434, x0540, x0678 and x0967, the His163 ε-nitrogen is protonated, forming a HB to the pyridine nitrogen (x0107, x0434, x0540 and x0678) or phenol oxygen (x0967). For x1093, the δ-nitrogen is protonated, leaving the ε-nitrogen free to form a HB with the indole -NH of x1093, reversing the HB polarity compared to the other fragments in the cluster. Nonetheless, the same binding geometry is observed in both cases and the clustering algorithm correctly assigns the molecules into the same cluster.

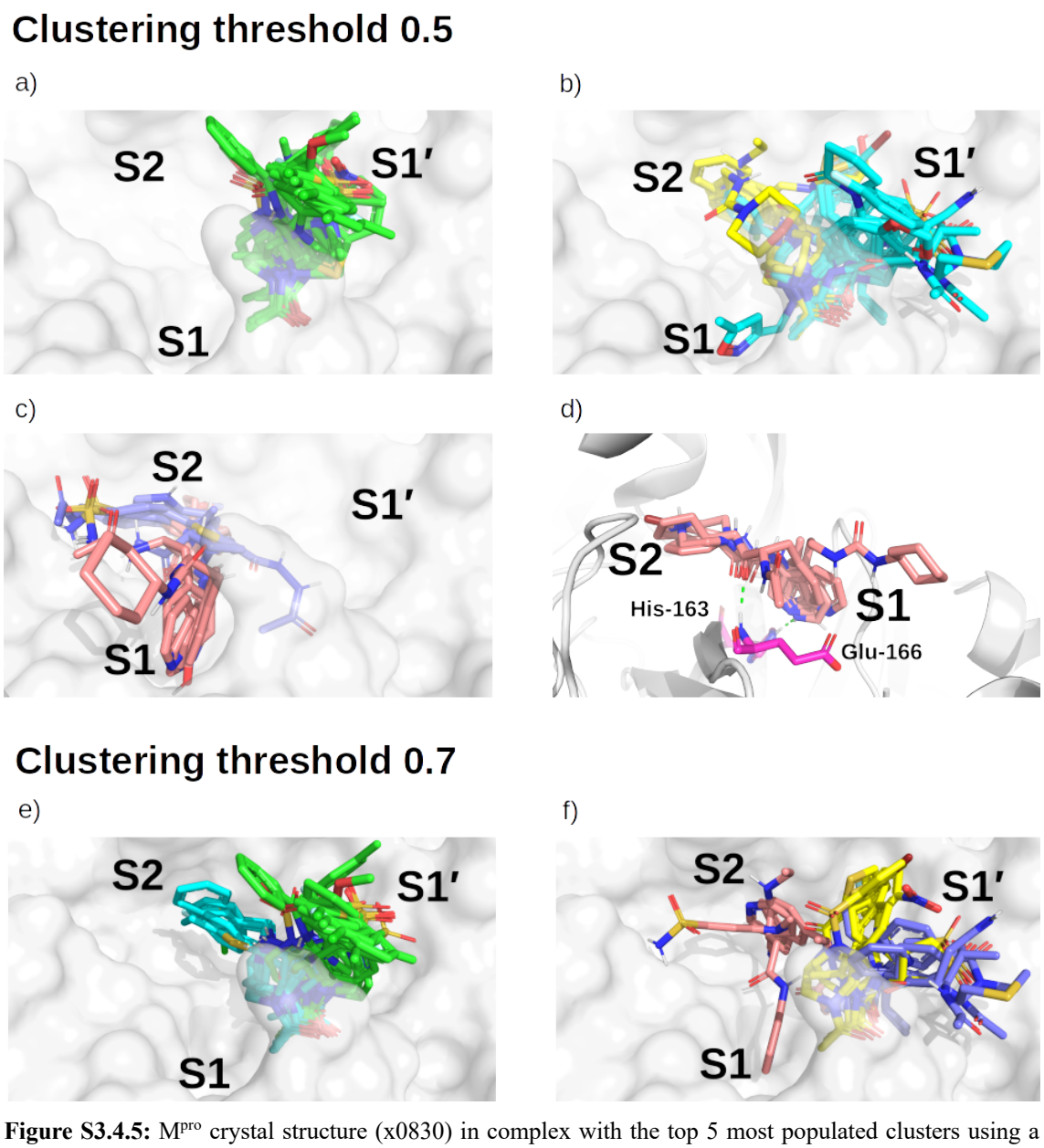


Figure S3.4.5: M^{pro} crystal structure (x0830) in complex with the top 5 most populated clusters using a clustering threshold of 0.5: a) cluster 1 (green); b) clusters 2 (cyan) and 3 (yellow); c) clusters 4 (blue) and 5 (salmon). d) Close-up on the binding pose of cluster 5. Shown in green are the two key HBs between the fragment carbonyl oxygen and the backbone nitrogen of Glu166 (HB 3 as identified in **main text Figure 3.3**), and between the His163 Nε and the nitrogen heterocycle of the fragment (HB 6 as identified in **main text Figure 3.3**). Also shown are the top 5 most populated clusters using a threshold of 0.7: e) clusters 1 (green) and 2 (cyan); f) clusters 3 (yellow), 4 (blue) and 5 (salmon).

Descriptor based on contact and long-range interactions

To further understand how the XChem fragments interact with M^{pro}, we performed linear scaling DFT calculations, as done for the substrate/designed peptides. Both short-range and long-range interaction terms were analysed and the computed coupling strengths between fragments and enzyme residues were compared to those obtained for the natural substrates, with the aim to identify potentially interesting inhibitor candidates. We have made these results publicly available in the following links:

Contact Interaction:

https://maayanlab.cloud/clustergrammer/viz/603f6484d0867e01721c4820/contact_interactions_subs_XChem.tsv

Electrostatic Interaction:

https://maayanlab.cloud/clustergrammer/viz/603f624cd0867e01721c47e5/electrostatic_interactions_subs_XChem.tsv

A hierarchical visualisation of contact and long-range interactions enables identification of families of different inhibitors (**Figure S3.4.6** below; left and right panels respectively). Within this clustering approach, the fragments/inhibitors belonging to the same family as the native substrates (s01, s02, s05) are highlighted in red and the cluster 5 compounds identified with Arpeggio in bold red (**main text Section "Interaction analysis of XChem fragments"**), by considering the short-range DFT interaction as a descriptor. Remarkably, all cluster 5 compounds belong to the same family as the native substrates. Although this analysis is at a preliminary stage, we believe that this is potentially a powerful direction of investigation as it enables agnostic comparison between compounds of different size and nature based on first-principle considerations.

Using long-range electrostatic interactions as a descriptor, cluster 5 compounds split into two main subfamilies. One of the compounds (x0540) belongs to the same subgroup as the substrate peptides, while the remaining compounds are grouped into another subfamily with compound x0874. Compounds that show similar long-range interaction patterns but different contact interaction patterns as substrates s01/s02 are highlighted in purple.

This analysis can be interpreted in two ways: on one hand, it enables a criterion to single out some XChem compounds which provide interaction patterns similar to those exhibited by native substrates. We can thus generalise the cluster 5 group of compounds into a larger family, which also includes the compounds highlighted in red (**Figure S3.4.6**). On the other hand, when applying such a criterion to long-range interactions, other compounds (in purple) with interaction patterns similar to those shown by the native substrates can be identified. It is of interest to further investigate such interaction pattern categorisation with the aim of identifying potent inhibitors with novel modes of action.

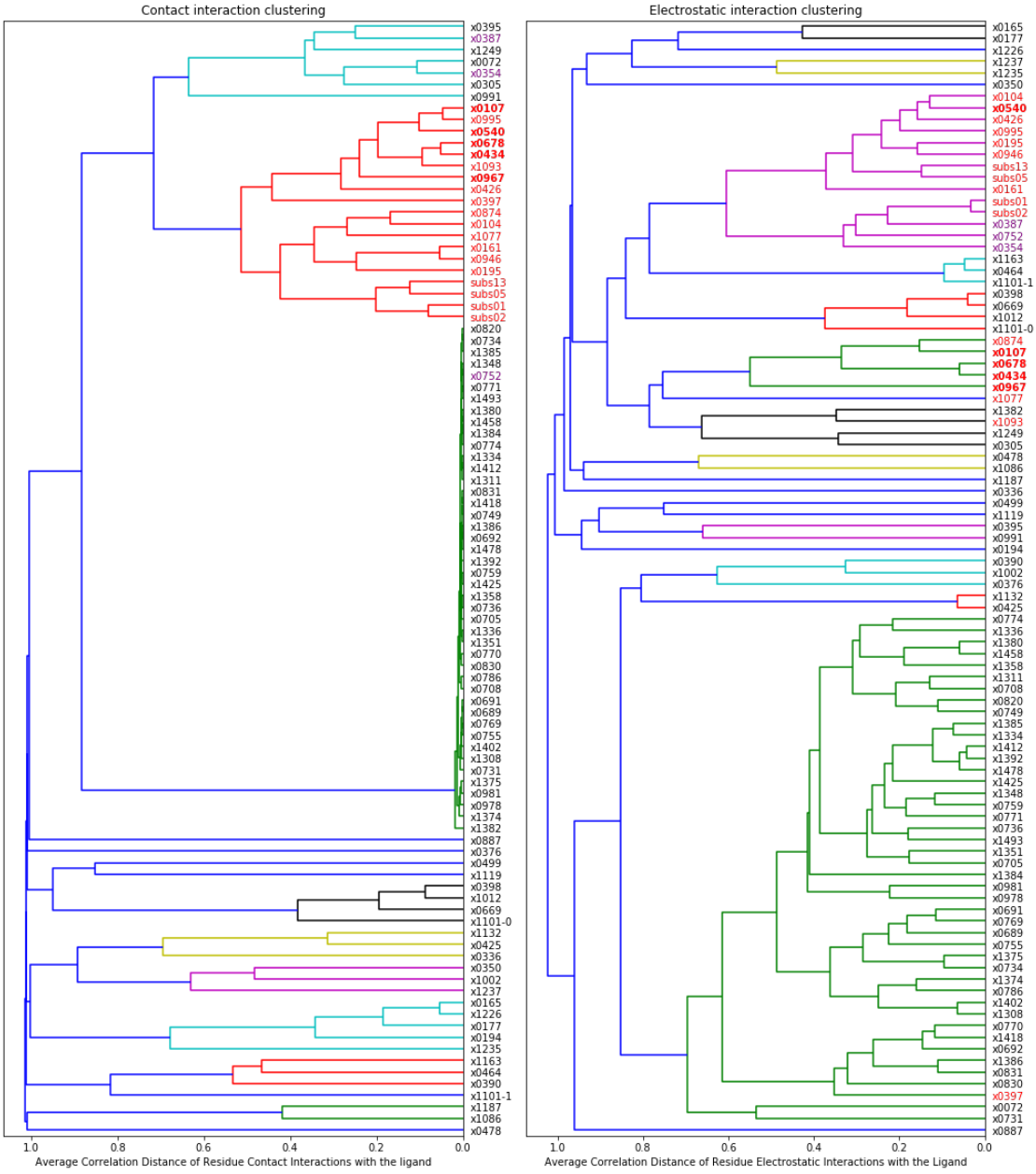


Figure S3.4.6: Dendrograms showing the hierarchical clustering of XChem fragment inhibitors, substrate peptides (s01, s02, s05), and designed peptide p13, considering descriptors defined on (left) short-range contact and (right) long-range DFT electrostatic interactions. The average correlation distance between interaction patterns is indicated on the x-axis. A family of compounds (red) which contains the peptides (s01, s02, s05, p13) as well as cluster 5 binders (bold) is identified using short-range interactions (left). Fragments that belong to the same family as s01 and s02 based on long-range (right) but not short-range interactions are highlighted in purple.

S3.4b Interaction analysis of COVID Moonshot compounds

As of the 11th January, 2021, the COVID Moonshot M^{pro} inhibitor project had reported fluorescence assay data for 798 inhibitors and SPE MS assay data for 784 inhibitors;[21, 22] X-ray crystal structures for 245 inhibitors in complex with M^{pro} were reported. To test our hypothesis that the interactions identified in the substrate and fragment analyses are important for inhibitor design, the Moonshot structures were processed with Arpeggio and ligand-M^{pro} interaction fingerprints were generated and compared to the most important interactions identified in fragment cluster 5. A Moonshot compound was classified as a cluster 5 binder if it shared at least 70% of the atom contacts found in cluster 5 binders. This analysis resulted in 101 assayed cluster 5 binders (**main text Figure 3.17**). Since the sensitivity of the activity assays are capped at 99 µM, weak binders could not be accurately quantified and were labelled with IC_{50} values of 99 µM.[22] In the following analysis the fluorescence assay results for 798 inhibitors were used.

When comparing the average IC_{50} of cluster 5 binders with the rest of the dataset, cluster 5 binders bind slightly more tightly with an average IC_{50} of 42.2 µM (95% confidence interval: 34.55-49.87 µM), while the average IC_{50} of all compounds is 54.0 µM (95% confidence interval: 50.95-57.07 µM). However, given the bimodal distribution of the data (**Figure S3.4.7**), a better metric is the hit rate of good compounds and the number of 'inactive' compounds in the dataset: 2 of the 10 best Moonshot compounds are cluster 5 binders, as are 10 of the top 10% (81 compounds) of Moonshot compounds. Out of the 101 cluster 5 binders, only 15 were identified as weak binders, while out of all 798 assayed compounds 263 (33%) were weak or non-binders. As a result, the observed IC_{50} difference between cluster 5 binders and other inhibitors might be higher than currently calculated. Note, only 245 of the 798 assayed compounds had been crystallographically analysed so it is unknown how many of the unsolved compounds are also cluster 5 binders.

Thus, by analysing fragment and substrate interactions with M^{pro}, we were able to extract and cluster structural interaction fingerprints and found in cluster 5 a promising non-covalent binding mode in the M^{pro} active site. Analysis of the large set of assayed Moonshot compounds reveals that cluster 5 binders are potent inhibitors with a higher hit rate of actives than the rest of the dataset. Cluster 5 binders represent highly promising starting points for further optimisation. The kernel density estimation (KDE) and histogram of the distribution of the COVID Moonshot dataset for the fluorescence and RapidFire dataset are shown in **Figure S3.4.7**.[21]

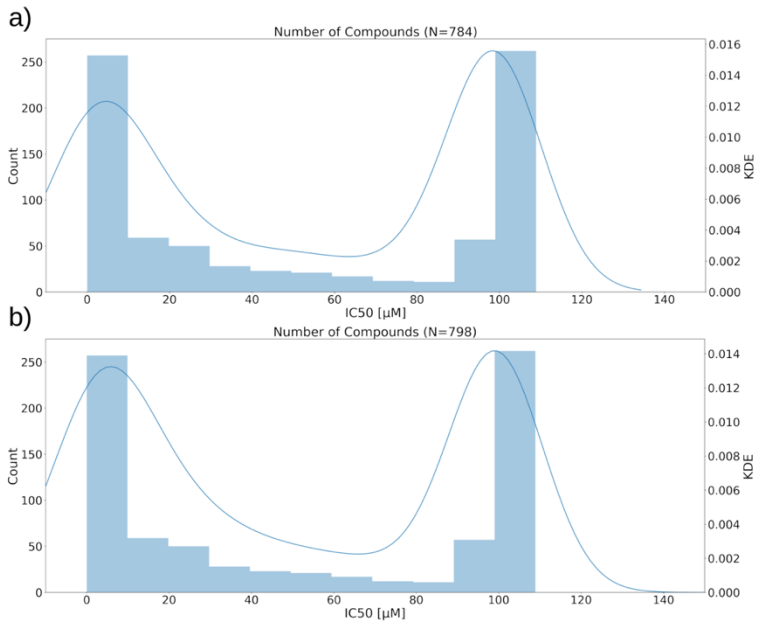


Figure S3.4.7: a) KDE and histogram for the RapidFire dataset. b) KDE and histogram for the fluorescence dataset.

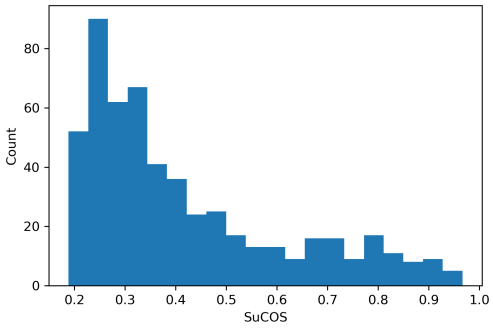
S3.4c Covalent docking of COVID Moonshot compounds

We selected the 540 covalently reacting compounds of the 10,001 designed Moonshot compounds and docked them using the knowledge-based pre-alignment docking method using AutoDock4 (**Section S3.1f**).[32] These compounds were designed by the Moonshot consortium, usually using one or more of the 44 covalent fragments as a core structure.[22] Only compounds with a matching covalent warhead to the inspiration fragment that also cite a single covalent fragment as their inspiration were selected to form the dataset of 540 compounds. To take advantage of the many diverse induced-fit conformations of M^{pro}, each designed compound was docked into the M^{pro} structure of the corresponding covalent "inspiration fragment".[22] The lowest energy docked pose in the highest populated cluster of each docking run was used to identify contacts using Arpeggio and generate the interaction Tanimoto distance matrix using a Tanimoto similarity threshold of 0.5 or 0.7 between poses of each cluster as described above for the XChem fragments. The broader clustering threshold of 0.5 leads to a total of 5 clusters, with the first cluster containing 477 of the 540 poses (88%) and no single-pose clusters; while the tighter threshold of 0.7 results in 46 clusters with 12 single molecule clusters. As expected, the diversity of binding modes of these compounds is much lower than in the original XChem fragment set, due to the limited number of fragments (44) and the reduced structural diversity of the designs, all being covalent S1/S1′ binders.

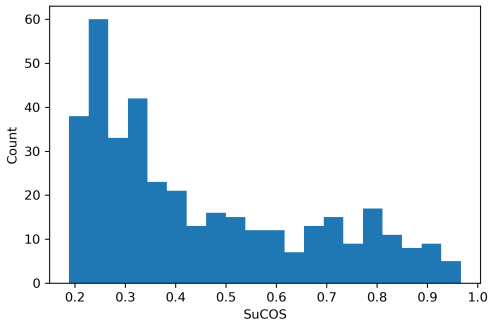
We subsequently analysed the ability of the procedure to recapitulate the binding pose of the parent fragment when docking the fragment-based Moonshot designs. We compared the shape and pharmacophoric (SuCOS) overlap of the lowest energy pose of the highest populated cluster for each Moonshot compound with the inspiration covalent XChem fragment referenced by the designers (**Figure S3.4.8**). A SuCOS score of 0.5 and higher was sufficient to consider the binding poses of the crystallographic fragment and docked design as conserved.[35] Due to creative freedom in the design process, some of the designed compounds do not overlap significantly with the inspiration fragments and in some extreme cases only have the covalent warhead in common. When controlling for the smallest maximum common substructure (MCS) that encompasses at least the covalent warhead and one additional atom in the compound, 379 docked designs remain, from which 132 (34.8%) recovered the binding mode of the inspiration fragment. Given the high similarity between the fragments and the docked designed compounds, it is likely that these binding modes are more representative of the actual binding mode of the ligand. Furthermore, by selecting docked compounds that did not have significant modifications to the original fragment (less than 10 atoms difference to the MCS between fragment and docked compounds), the number of compounds regaining the binding pose was 87 (54.7%) of the 159 compounds. Interestingly, the distribution of SuCOS over the data subsets resembles a bimodal distribution in all cases, with one peak around a SuCOS of 0.25 and a second peak between 0.7-0.8 depending on the subset (**Figure S3.4.8**). This suggests that the AD4 docking process performs well generally in identifying binding poses similar to the original fragment (especially in cases where few modifications were made), but fails completely in some cases, resulting in almost no overlap between the pose and the fragment and an extremely low SuCOS (<0.3). This observation is in line with the expected changes in binding mode of a molecule when subjected to large changes in structure and does not necessarily correspond to incorrect docked poses. As a result, when filtering for compounds where only minor changes were made to the inspiration fragment and the docking pose generated regained the same binding pose as the fragment, we increased the probability of gaining relevant poses. However, thorough validation of this hypothesis is yet to be done.

At the point of analysis, only 6 of the 540 docked covalent compounds have been analysed crystallographically.[30] These structures were used as a limited benchmark for the docking method; an overlay of the crystallographic conformation, the lowest energy pose of the highest populated cluster from docking, and the corresponding crystallographic structure of the inspiration fragment is shown (**Figure S3.4.9**). x10899 (**Figure S3.4.10**) was excluded from further analysis since it binds via a crystal contact to a third symmetry-related M^{pro} molecule, rather than the biologically relevant dimeric state.[40] The binding modes of two compounds, x3077 and x10306 (**Figure S3.4.9a** and **S3.4.9e** respectively), were reproduced perfectly. The method places the aromatic sidechain of x3324 (**Figure S3.4.9b**) correctly into the S2 pocket of M^{pro} but varies on placement of the linker when compared to the crystal structure. However, in this case, the original fragment that was cited as inspiration has no overlap with the designed compound. As a result, the induced fit shape of the active site of the M^{pro} structure used for docking complements a completely different molecule and the alignment before docking pointless. For x3325 and x10172 (**Figure S3.4.9c** and **S3.4.9d** respectively), the selected lowest-energy pose of the highest populated cluster did not match the binding pose of the crystal structure.

a)



b)



c)

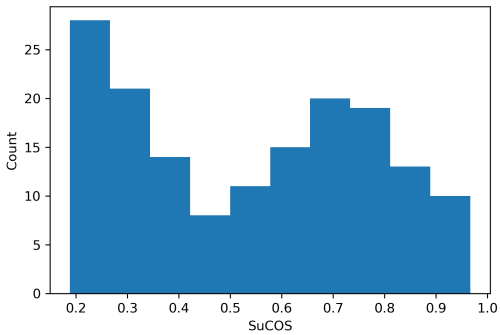


Figure S3.4.8: Distribution of SuCOS between the docked pose of the covalent Moonshot design and the original covalent fragment used as a basis for design. a) SuCOS scores of the 540 docked compounds; b) SuCOS scores of the 379 docked compounds with significant MCS overlap (*i.e.* more than 8 atoms MCS match) to the fragment; c) SuCOS scores of the 159 docked compounds with substantial MCS overlap to the fragment and only small changes to its structure (<10 atoms difference between the compound and its MCS with the inspiration fragment).

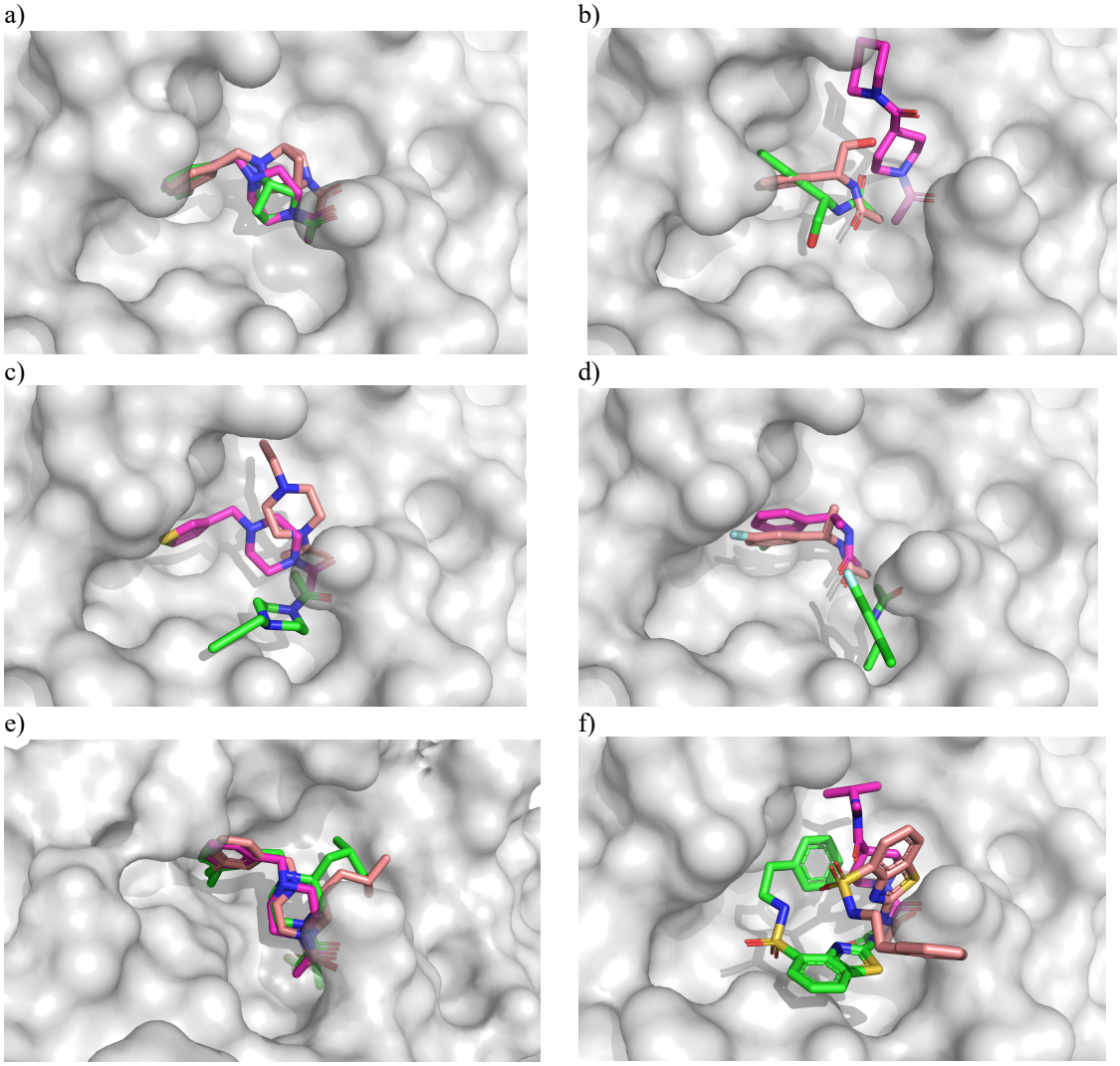


Figure S3.4.9: Overlay of the lowest energy pose in the highest populated cluster of the AD4 covalent docking procedure (green) with the original inspiration fragment (pink) and the crystal structure (salmon). For every dock, the M^{pro} protein structure of the corresponding inspiration fragment co-crystal structure was used. a) View from a crystal structure of Moonshot design X3077 (salmon) with fragment X0770 (pink) and the docked pose of X3077 (green). b) View from a crystal structure of Moonshot designed compound X3324 (salmon) with fragment X1380 (pink) and the docked pose of X3324 (green). c) View from a crystal structure of Moonshot design X3325 (salmon) with fragment X1386 (pink) and the docked pose of X3325 (green). d) View from a crystal structure of Moonshot design X10172 (salmon) with fragment X1382 (pink) and the docked pose of X10172 (green). e) View from a crystal structure of Moonshot design X10306 (salmon) with fragment X0770 (pink) and the docked pose of X10306 (green). f) View from a crystal structure of Moonshot design X10899 (salmon) with fragment X1458 (pink) and the docked pose of X10899 (green).

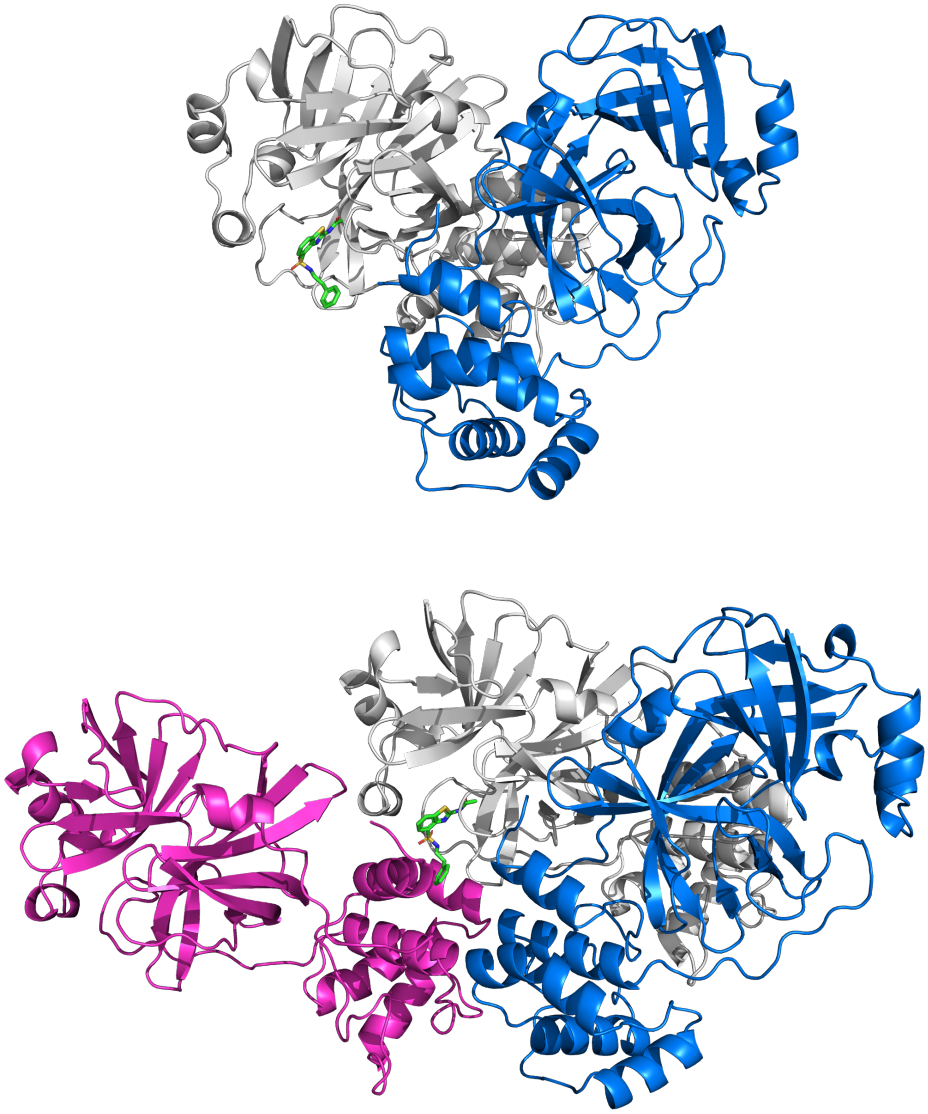


Figure S3.4.10: View of the crystal structure of X10899 (green) from the perspective of the biologically relevant dimer (top), with one additional crystal packing symmetry mate (bottom). Chains A, B (dimer) and C (crystal-contacting chain) are coloured white, blue and pink, respectively. The small molecule X10899 is bound in the active site of chain A (white); but its terminal aromatic sidechain interacts with the symmetry-related chain C (magenta, bottom), leading to what would appear to be an unusual binding mode if only the dimer were considered (top).

S3.4d Implications for future inhibitor design

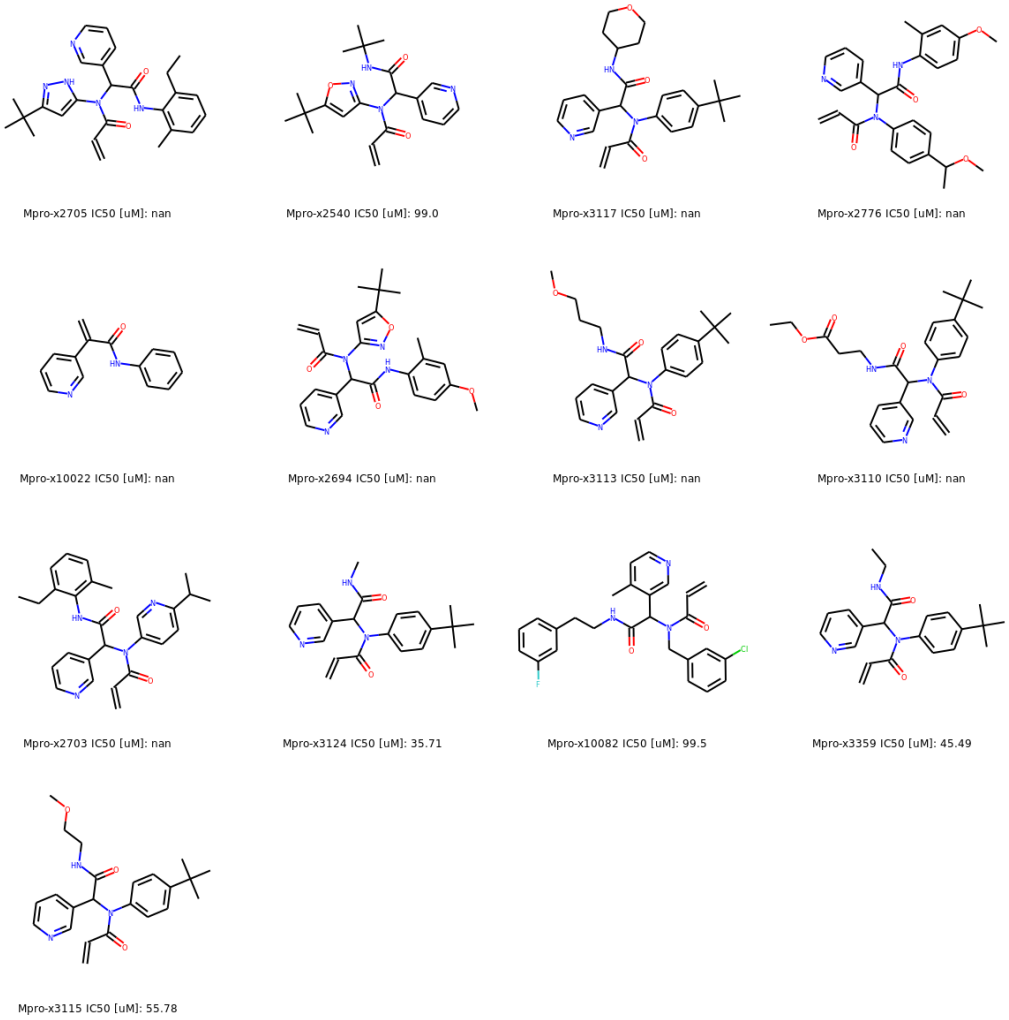


Figure S3.4.11: Selection from all Moonshot compounds in cluster 5 that bind into the oxyanion hole. All compounds are covalent inhibitors, reacting with Cys145 via the acrylamide warhead. IC_{50} values are obtained from the postera.ai GitHub page.[22] Note: "nan" means "not a number", indicating that the compound has not been assayed.

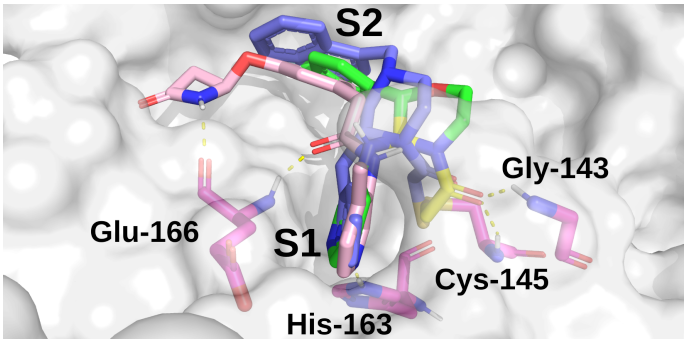


Figure S3.4.12: Overlay of the docked pose of FOC-CAS-e3a94da8-1 (green), the docked pose of MIH-UNI-e573136b-3 (blue), and a crystallographically observed binding mode of X10789 (salmon) with M^{pro} (PDB 5RER).[23] The proposed expansion of x10789 into the oxyanion hole is shown in yellow on compound FOC-CAS-e3a94da8-1.

Section S3.5: Supplementary Information References

1. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.*, et al.* Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **2008,** *82*, 2515.
2. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2016,** *54*, 5.6.1.
3. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525.
4. Schrödinger LLC. *Schrödinger Release 2020-4: Maestro*, New York, NY, 2020.
5. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.*, et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696.
6. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926.
7. Grubmüller, H.; Groll, V.; Tavan, P. *SOLVATE*, Institute for Medical Optics, University of Munich, 1996-2013.
8. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T.E. Cheatham, I.*, et al. AMBER 2019*, University of California, San Francisco, 2019.
9. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23*, 327.
10. Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011,** *7*, 931.
11. Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008,** *10*, 6615.
12. Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chem. Acc.* **1973,** *28*, 213.
13. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.*, et al.* Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982,** *77*, 3654.
14. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.*, et al. Gaussian*, Wallingford, CT, 2016.
15. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
16. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
17. Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004,** *55*, 383.
18. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.*, et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **2017,** *13*, e1005659.
19. Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montaño, B.; Pitt, W. R.; Ascher, D. B.*, et al.* Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J. Mol. Biol.* **2017,** *429*, 365.
20. Jubb, H. C. *PDBTools*, GitHub: https://github.com/harryjubb/pdbtools, 2019.
21. Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **2020,** *12*, 581.
22. Achdout, H.; Aimon, A.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.*, et al.* COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning. *bioRxiv* **2020**, DOI: 10.1101/2020.10.29.339317.
23. Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.*, et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020,** *11*, 5047.
24. Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytologist* **1912,** *11*, 37.
25. Ratcliff, L. E.; Dawson, W.; Fisicaro, G.; Caliste, D.; Mohr, S.*, et al.* Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *J. Chem. Phys.* **2020,** *152*, 194110.

26. Dawson, W.; Mohr, S.; Ratcliff, L. E.; Nakajima, T.; Genovese, L. Complexity Reduction in Density Functional Theory Calculations of Large Systems: System Partitioning and Fragment Embedding. *J. Chem. Theory Comput.* **2020,** *16*, 2952.
27. Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.*, et al.* Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **2021,** *11*, e1493.
28. Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem. Int. Ed.* **2020,** *59*, 15665.
29. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
30. Diamond Fragalysis. https://fragalysis.diamond.ac.uk/ (accessed January 2021).
31. *RDKit: Open-Source Cheminformatics Software*, 2020.
32. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.*, et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30*, 2785.
33. *Molecular Operating Environment (MOE)*, 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.
34. Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. *J. Cheminform.* **2015,** *7*, 26.
35. Leung, S.; Bodkin, M.; von Delft, F.; Brennan, P.; Morris, G. SuCOS is Better than RMSD for Evaluating Fragment Elaboration and Docking Poses. *ChemRxiv* **2019**, DOI: 10.26434/chemrxiv.8100203.v1.
36. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.*, et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003,** *24*, 1999.
37. Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* **2000,** *21*, 1088.
38. Krzemińska, A.; Paneth, P.; Moliner, V.; Świderek, K. Binding Isotope Effects as a Tool for Distinguishing Hydrophobic and Hydrophilic Binding Sites of HIV-1 RT. *J. Phys. Chem. B* **2015,** *119*, 917.
39. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.*, et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
40. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.*, et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368*, 409.
41. Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.*, et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020,** *11*, 3202.
42. Commissariat à l'Energie Atomique *Polaris(MD)*, http://biodev.cea.fr/polaris/index.html, 2018.
43. Wei, M.; Wynn, R.; Hollis, G.; Liao, B.; Margulis, A.*, et al.* High-throughput determination of mode of inhibition in lead identification and optimization. *J. Biomol. Screen.* **2007,** *12*, 220.
44. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.*, et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
45. Ravindranath, P. A.; Sanner, M. F. AutoSite: an automated approach for pseudo-ligands prediction—from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics* **2016,** *32*, 3142.
46. Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.*, et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020,** *11*, 5877.

B. Supplementary Information to Section 3.4: β-lactam inhibitors

Section S3.6: Workflow for identification of inhibitor binding modes

S3.6a Docking

Three-dimensional inhibitor structures were generated from SMILES strings,[1] protonated at pH 7.5 using Open Babel,[2] and optimised with PBE0/6-31G(d) for a maximum of 7 cycles using the Gaussian 16 package.[3, 4] The SMD implicit solvation model (water) and D3BJ dispersion correction were used.[5, 6] Optimised ligand structures were docked against the SARS-CoV-2 M^{pro} crystal structure complexed with the peptidomimetic inhibitor N3 (PDB 6LU7; 2.16 Å resolution).[7] An inhibitor-bound M^{pro} structure was chosen for docking instead of apo structures, as the protein pockets in the inhibitor-bound structure were likely widened for accommodating ligands and more representative of inhibitor-bound complexes.

Docking was performed using GOLD (v. 5.8.1) with the ChemPLP scoring function,[8] with a search space defined as within 6 Å of N3. For each ligand, 10 poses were outputted, before RMSD clustering with a cut-off of 2.0 Å. Docking was also performed using AutoDock Vina (v. 1.1.2)[9] with a cubic search space of length 30 Å centred at the C145 Sγ atom. For each ligand, 9 solutions, which had been clustered internally within the program, were outputted.

S3.6b Selection of poses

For identification purposes, clustered docked poses from GOLD and AutoDock Vina were labelled "G" and "V" respectively, followed by a number indicating its scored ranking. For instance, the top ranked solution from AutoDock Vina would be labelled pose "V1". Considering the conserved interactions formed between M^{pro} and its substrates or potent inhibitors (**Manuscript I),**[10] only poses that displayed constructive binding in at least two of the three key pockets (the hydrophobic S2 pocket, the Gln-accommodating S1 pocket, and the oxyanion hole accommodating electronegative functionalities) were considered for MD.

For inhibitors **1** and **2** (**Manuscript VI**),[11] an "F2" pose was considered. This pose was inspired by the V2 pose of a structurally similar compound (compound **2** with benzyl in place of *p*-nitrobenzyl) which had the benzyl group situated in the S2 pocket. Such a pose was of mechanistic interest and deemed spatially possible for **1** and **2**.

S3.6c Parametrisation and MD

Based on the V1 conformation, ligands were parametrised with Antechamber (v. 19.0) according to the GAFF forcefield.[12, 13] Charges were derived from the AM1-BCC scheme.[14]

The selected poses were transferred to an *apo* M^{pro} structure of higher resolution (PDB 6YB7; 1.25 Å resolution),[15] and the M^{pro}-ligand complex was solvated, neutralised, minimised, and equilibrated as performed for protein-substrate complexes (see **Methods** in **Manuscript I**),[10] before being subjected to 10 ns of MD. For each pose, three independent simulations were performed initially. Provided that at least one of the three simulations showed stable ligand binding (stability defined in the following section), two more independent runs were performed, yielding a total of 5 runs. The runs were labelled "a" to "e".

S3.6d Stability and interaction analyses

The trajectory was fitted based on M^{pro} backbone atoms. Ligand binding was considered stable if at least one of the following two criteria was satisfied:

1. The final non-hydrogen ligand RMSD was below 5 Å.
2. Frames every 100 ps (101 frames in total, including at $t = 0$) were clustered with the *gromos* method,[16] using a cut-off of 2 Å on the RMSD of ligand non-hydrogen atoms. This stability test was passed if the most populated cluster accounted for 80 or more frames. In complement to Test 1, this allowed

consideration of poses which might have undergone relaxation and deviation from the original docked pose but remained stable afterwards.

For MD runs which demonstrated stable ligand binding, protein-ligand interactions were characterised. HBs were detected by the *gmx hbond* module using the combined criteria of $d_{D\text{-}A} < 3.5$ Å and $\angle$(H-D-A) < 30°, and were considered significant if detected in ≥25 of the 101 frames. Significant hydrophobic contacts were identified from the central structure of the most populated cluster, using the PLIP server.[17] Significant interactions are shown in **Figures 3.7.1-2**. Note that the significant interactions shown were not necessarily present simultaneously.

Inhibitor binding energy (omitting the entropic term) was estimated using MM/PBSA method as implemented in the g_mmpbsa program.[18] Frames were extracted in 0.5 ns intervals over the 10 ns MD. An internal dielectric constant ε_{int} of 2 was used. Polar solvation was accounted for by solving the linear Poisson-Boltzmann equation at 300 K, with a 0.15 M salt concentration and bondi radii. Non-polar solvation was considered using the SASA model, with coefficient $\gamma = 0.02267$ kJ mol^{-1} Å^{-2} and offset $c =$ 3.84928 kJ mol^{-1}.[18]

Section S3.7: Pose stabilities of compounds 1 and 2

Table S3.7.1: MD-observed ligand stability. Stability is defined as passing at least 1 of the 2 Tests (T1/T2; **Section S3.6d**). Bound conformations are shown in **Figures S3.7.1-2**.

Compound	*Pose*	*Run*	*Final ligand RMSD (nm)*	*T1 (< 5 Å)*	*Top cluster #Frames*	*T2 (≥80)*	*Stable (T1/T2)?*	*#Runs pass T1/T2*
1	G1	1	0.311	Y	93	Y	Y	3
		2	0.499	Y	77	N	Y	
		3	0.311	Y	44	N	Y	
		4	0.795	N	42	N	N	
		5	0.841	N	21	N	N	
	G6	1	0.285	Y	76	N	Y	2
		2	0.816	N	29	N	N	
		3	0.269	Y	64	N	Y	
		4	0.714	N	24	N	N	
		5	1.338	N	18	N	N	
	V1	1	0.230	Y	49	N	Y	5
		2	0.219	Y	93	Y	Y	
		3	0.293	Y	91	Y	Y	
		4	0.256	Y	72	N	Y	
		5	0.272	Y	83	Y	Y	
	F2	1	0.292	Y	99	Y	Y	5
		2	0.194	Y	101	Y	Y	
		3	0.165	Y	101	Y	Y	
		4	0.174	Y	101	Y	Y	
		5	0.223	Y	96	Y	Y	
2	G3	1	1.729	N	15	N	N	1
		2	0.498	Y	9	N	Y	
		3	0.551	N	12	N	N	
		4	0.801	N	41	N	N	
		5	0.622	N	73	N	N	
	V1	1	0.323	Y	88	Y	Y	3
		2	1.225	N	30	N	N	
		3	0.517	N	36	N	N	
		4	0.407	Y	80	Y	Y	
		5	0.299	Y	69	N	Y	
	V3	1	0.220	Y	97	Y	Y	5
		2	0.359	Y	88	Y	Y	
		3	0.218	Y	87	Y	Y	
		4	0.278	Y	90	Y	Y	
		5	0.218	Y	85	Y	Y	
	V6	1	0.686	N	29	N	N	0
		2	0.519	N	14	N	N	
		3	0.524	N	61	N	N	
	F2	1	0.627	N	82	Y	Y	1
		2	0.584	N	54	N	N	
		3	1.014	N	29	N	N	
		4	0.656	N	66	N	N	
		5	0.872	N	31	N	N	

S3.7a Compound 1

Table S3.7.2: Stability of **1** in M^{pro}. Poses are ranked based on the number of stable runs, followed by average MM/PBSA binding energy (kcal mol^{-1}).

Ranked Pose	*Description*	*# Runs pass T1/T2*	$<\Delta U_{bind}>$	ΔU_{bind} *range*	$<\Delta U_{vdW}>$	$<\Delta U_{Coulomb}>$
F2	Reactive (Bz coiled into S2)	5	-24.5	-27.2 – -21.0	-51.6	-12.0
V1	Reactive (Bz in S1′)	5	-20.7	-22.2 – -19.2	-47.6	-12.7
G1	Spanning S2+S1	3	-14.3	-15.8 – -13.1	-40.2	-12.1
G6	Extended	2	-19.3	-19.9 – -18.7	-43.9	-9.9

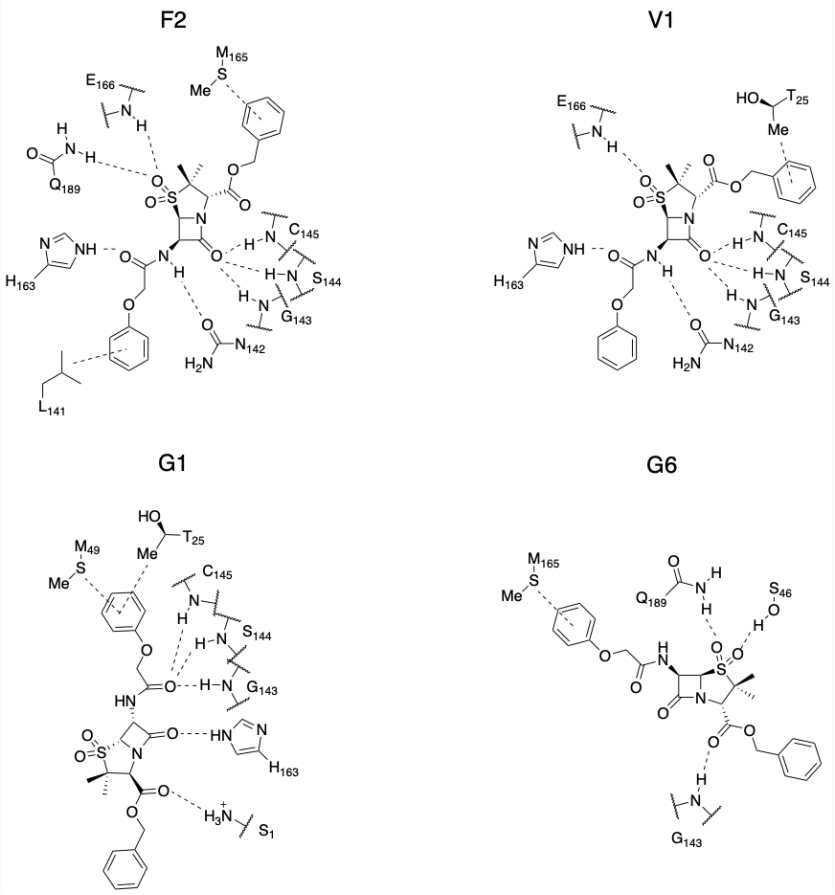


Figure S3.7.1: Poses and identified significant interactions for **1**.

S3.7b Compound 2

Table S3.7.3: Stability of **2** in M^{pro}. Poses are ranked based on the number of stable runs, followed by average MM/PBSA binding energy (kcal mol^{-1}).

Ranked Pose	*Description*	*# Runs pass T1/T2*	$\langle \Delta U_{bind} \rangle$	ΔU_{bind} *range*	$\langle \Delta U_{vdW} \rangle$	$\langle \Delta U_{Coulomb} \rangle$
V3	C3-NO_2Bz in S2 and sulfoxide in oxyanion hole	5	-26.8	-28.5 – -25.8	-49.6	-13.6
V1	Reactive with C3-NO_2Bz in S1′	3	-18.4	-18.9 – -18.1	-45.8	-14.2
G3	Spanning S2+S1, with C6-Bz in S2	1	-13.8	-13.8	-28.0	-5.0
F2	Reactive with C3-NO_2Bz in S2	1	-23.3	-23.3	-41.7	-7.4
V6	Spanning S2+S1, with C3-NO_2Bz in S2	0	-	-	-	-

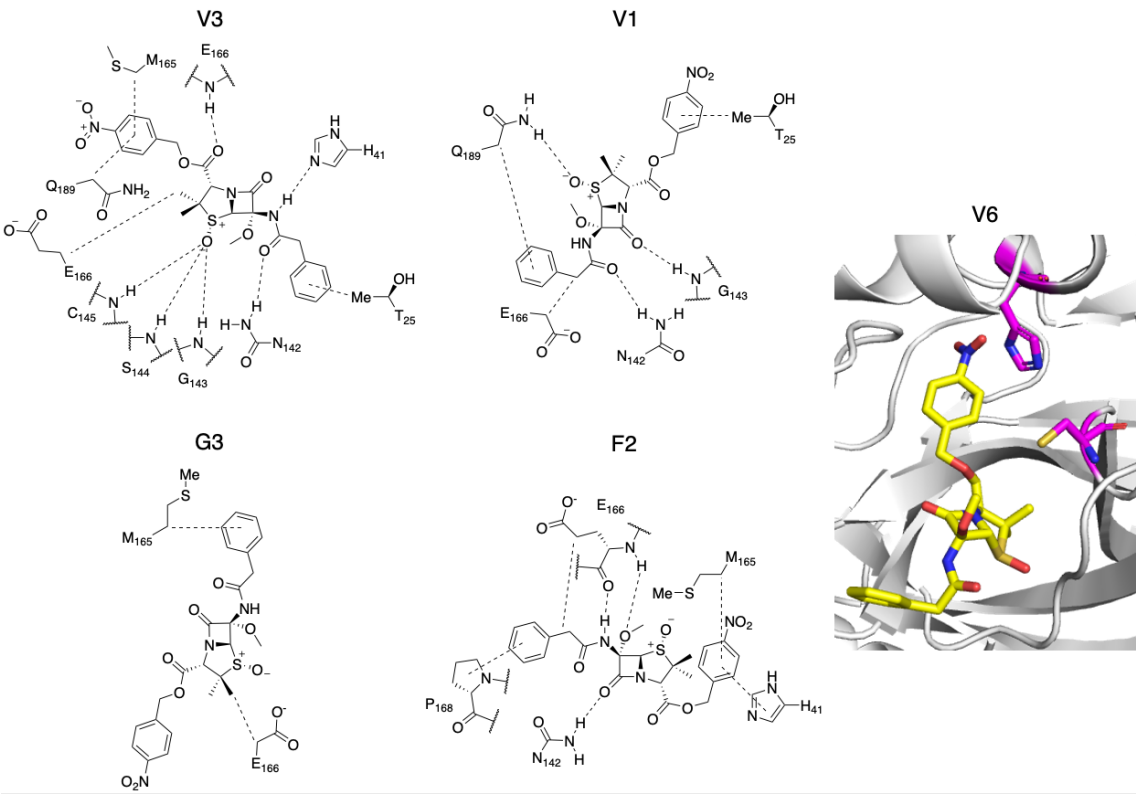


Figure S3.7.2: Poses and identified significant interactions for **2**. None of the MD simulations of the V6 pose were stable; the MM-minimised pose is shown above in yellow (M^{pro} Cys145 and His41 in magenta).

Section S3.8: FEP calculations on compound 1 poses

S3.8a Setup

General setup. FEP calculations were performed using GROMACS (v. 2021.3).[19] Preparation and parametrisation of the solvated M^{pro}-ligand complex and the solvated ligand system were performed as described in **Section S3.6c**. Each λ-window was subjected to the following series of procedures: (i) steepest descent minimisation until maximum force <1000 kJ mol^{-1} nm^{-1}; (ii) 100 ps NVT equilibration (2 fs timestep) at 298.15 K with position restraints on solute non-hydrogen atoms; (iii) 100 ps NPT equilibration (2 fs timestep) at 298.15 K and 1.0 bar; (iv) 5 ns production run (2 fs timestep) under NPT conditions, during which ΔU values to the neighbouring λ-values were calculated every 0.2 ps. Bonds involving hydrogens were constrained using the LINCS algorithm.[20] System temperature control was achieved using stochastic dynamics with $\tau = 1$ ps. Pressure control was achieved using the Parrinello-Rahman barostat with $\tau = 2$ ps.[21, 22] All simulations were performed with three-dimensional PBC. Long-range electrostatic and LJ interactions were calculated using the smooth particle mesh Ewald method with short-range cut-offs of 1 nm.[23] Long-range LJ pair parameters were obtained using geometrical combination rules.[24]

Restraints. To aid convergence, prior to switch-off of protein-ligand interactions, a combination of harmonic restraints (one distance, two angles, three dihedrals) was gradually turned on to restrain the ligand positionally and orientationally relative to the protein.[25] A distance restraint (k = 4184 kJ mol^{-1} nm^{-2}) was applied to the distance between (M^{pro}) Cys145_Cα and (ligand) C7. Two angle restraints (k = 41.84 kJ mol^{-1} rad^{-2}) were applied to three consecutive atoms in the sequence: (M^{pro}) Ser144_Cα, Cys145_Cα, (ligand) C7, N4. Three dihedral restraints (k = 41.84 kJ mol^{-1} rad^{-2}) were applied to four consecutive atoms in the sequence: (M^{pro}) Gly143_Cα, Ser144_Cα, Cys145_Cα, (ligand) C7, N4, C5. Equilibrium values for these restraints were taken from the observed histogram mode from previously run MD simulations (**Table S3.7.2**). $\Delta G_{rest(sol)}$ was calculated analytically as described by Boresch *et al.* (2003).[25]

λ-Windows. Overall, two alchemical transformations had to be considered to complete the ABFE cycle: (i) the decoupling/annihilation of the ligand solvated in water, for evaluation of $-\Delta G_{lig\text{-}sol}$; (ii) the application of restraints followed by the decoupling/annihilation of the ligand in the solvated M^{pro}-ligand complex, for evaluation of $\Delta G_{rest(comp)} + \Delta G_{decoup}$. For both transformations, both the intramolecular and intermolecular non-bonded interactions of the ligand were gradually turned off, *i.e.*, the ligand was gradually annihilated. In (ii), restraints were gradually turned on over 12 windows: λ = 0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1.[26] Then, for both (i) and (ii), electrostatic interactions were turned off linearly over 11 windows in $\Delta\lambda = 0.1$ intervals. This is followed by the switch-off of vdW interactions over 21 windows in $\Delta\lambda = 0.05$ intervals, using a soft-core potential with $\alpha = 0.5$, minimum $\sigma = 0.3$ nm, and power of $\lambda = 1$.[27, 28] In total, transformation (i) was performed over 31 λ-windows, while transformation (ii) was performed over 42 λ-windows.

BAR. ΔG values of the above transformations were calculated by BAR as implemented in *gmx bar* (GROMACS, v. 2021.3),[19, 29] at 298.15 K. Errors were estimated from the average variance over five data blocks.

S3.8b Convergence assessment and symmetry corrections

Initially, entire ΔU datasets from the 5 ns production runs performed in each λ-window were used for BAR. However, considering that earlier data points might not reflect fully equilibrated systems, convergence and equilibration were assessed using methods proposed by Klimovich *et al.* (2015).[30] Briefly, ΔG values were evaluated both in the forward time direction starting from $t = 0$, and in the reverse time direction starting from $t = 5$ ns, with cumulative addition of 10% (0.5 ns) more of the data until the entire dataset was included in both directions; non-equilibrated phases were then identified as a deviation from a steady plateau towards the end of the reverse time series plot. The following plots show the ΔG convergence assessment for transformation (i) ($-\Delta G_{lig\text{-}sol}$) of the solvated ligand system, with simulations initiated from the ligand conformation identical to that when bound in the (left) F2 or (right) V1 pose:

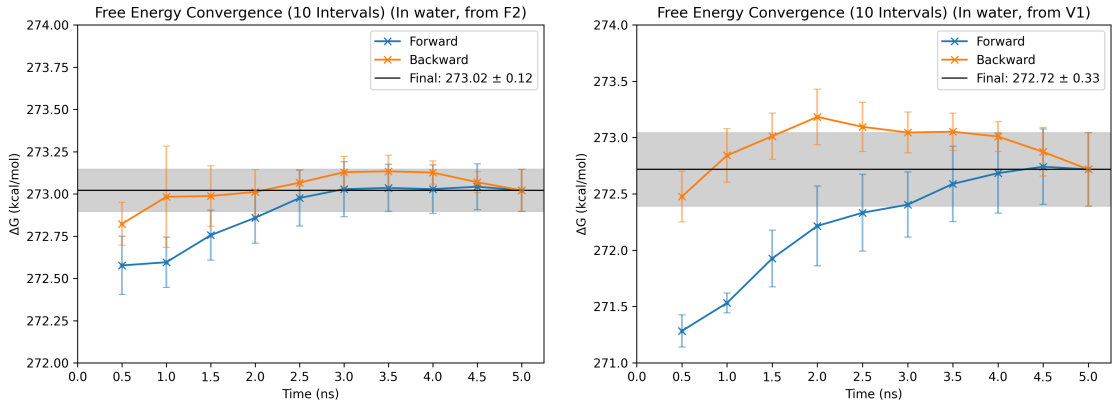


Figure S3.8.1: ΔG convergence for transformation (i) ($-\Delta G_{\text{lig-sol}}$), initiated from the (left) F2 or (right) V1 conformation.

In the F2 case, ΔG was always within ±0.5 kcal mol^{-1} of the final value, and displayed quick convergence to within error bar of the final value. In the V1 case, the reverse time ΔG plateaued around 273.0 kcal mol^{-1} up to ~80% of the time series, before decreasing towards the final value of 272.7 kcal mol^{-1}. This suggests that the first 20% (1 ns) of the forward time series data should be discarded as a non-equilibrated region. Since $-\Delta G_{\text{lig-sol}}$ is independent of the initial ligand conformation, the equilibrated datasets from both the F2- and V1-initated simulations were combined to obtain the final $-\Delta G_{\text{lig-sol}}$ value.

For transformation (ii) ($\Delta G_{\text{rest(comp)}} + \Delta G_{\text{decoup}}$), in the F2 case, results quickly converged to within error bar of the final value, and thus the entire dataset was retained. In the V1 case, similar to the above case for (i), the first 1 ns (20%) of the forward time series was discarded as a non-equilibrated region.

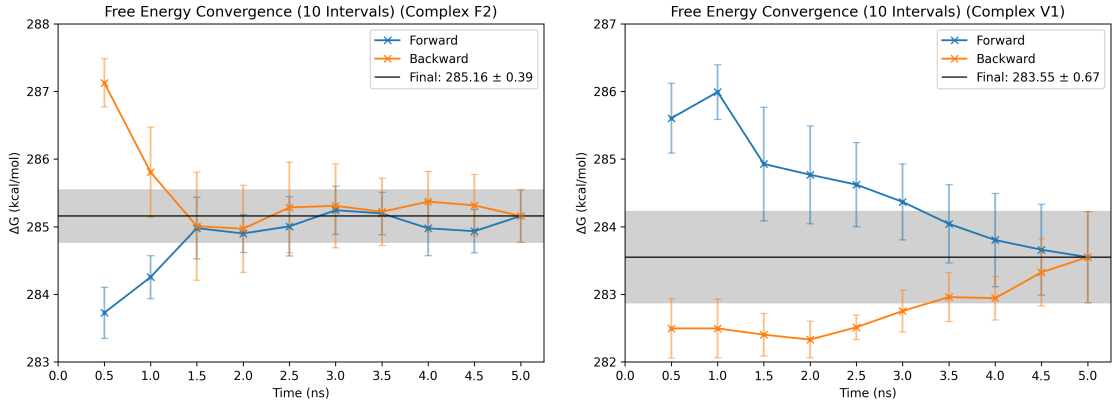


Figure S3.8.2: ΔG convergence for transformation (ii) ($\Delta G_{\text{rest(comp)}} + \Delta G_{\text{decoup}}$) in the (left) F2 and (right) V1 bound conformations.

A symmetry correction to ΔG_{bind} equal to $-k_B T \ln \sigma$, where k_B is the Boltzmann constant, T the temperature, and σ the symmetry number of the ligand or its substituent group that possesses any rotational symmetry, needs to be applied if the alternative rotationally equivalent conformation(s) were not sampled over the course of simulations, *e.g.,* due to applied restraints.[31] In **1**, the aromatic sidechains at C3 and C6 possess C_2 rotational symmetry. However, since none of the applied restraints restricted these rotational degrees of freedom, and the alternative conformers were sampled over the course of simulations, such corrections were not required.

S3.8c Free energy differences over λ-windows

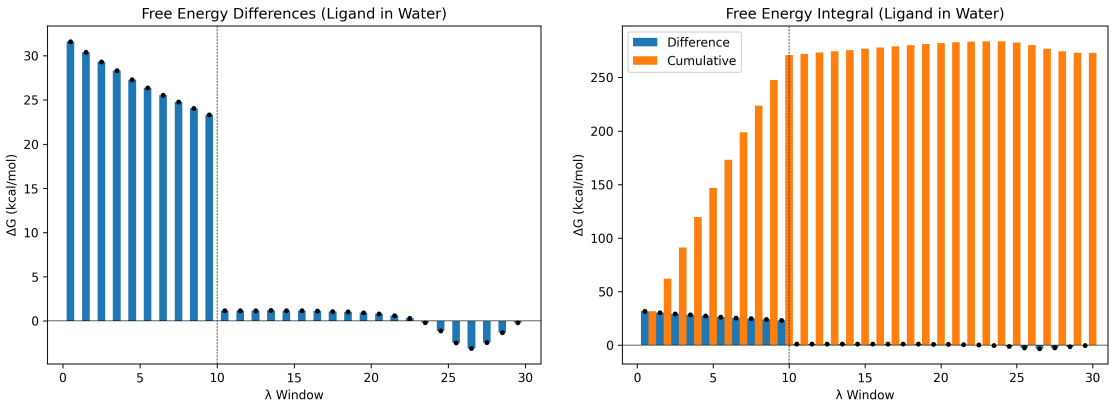


Figure S3.8.3: BAR-derived ΔG values between every neighbouring λ-value, and the cumulative ΔG values for transformation (i) ($-\Delta G_{\text{lig-sol}}$), using the combined F2 and V1 datasets. The dotted line separates the two phases: turning off electrostatics; turning off vdW.

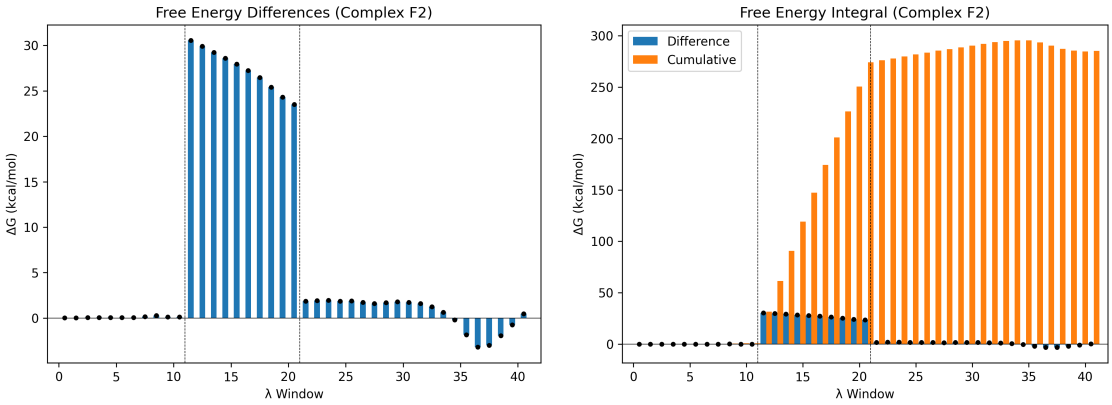


Figure S3.8.4: BAR-derived ΔG values between every neighbouring λ-value, and the cumulative ΔG values for transformation (ii) ($\Delta G_{\text{rest(comp)}} + \Delta G_{\text{decoup}}$), for ligand pose F2. The dotted lines separate the three phases: turning on restraints; turning off electrostatics; turning off vdW.

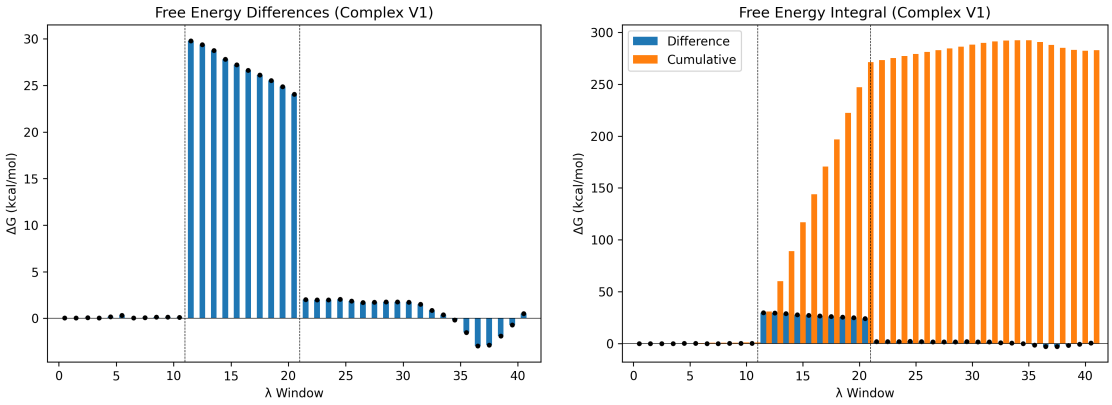


Figure S3.8.5: BAR-derived ΔG values between every neighbouring λ-value, and the cumulative ΔG values for transformation (ii) ($\Delta G_{\text{decoup}} + \Delta G_{\text{rest(comp)}}$), for ligand pose V1. The dotted lines separate the three phases: turning on restraints; turning off electrostatics; turning off vdW.

Section S3.9: References to Supplementary Information: β-lactam inhibitors

1. *RDKit: Open-Source Cheminformatics Software*.
2. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.*, et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **2011,** *3*, 33.
3. Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999,** *110*, 6158.
4. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.*, et al. Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
5. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009,** *113*, 6378.
6. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456.
7. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.*, et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582*, 289.
8. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997,** *267*, 727.
9. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31*, 455.
10. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
11. Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.*, et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **2021,** *57*, 1430.
12. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157.
13. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247.
14. Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23*, 1623.
15. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
16. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.*, et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
17. Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **2015,** *43*, W443.
18. Kumari, R.; Kumar, R.; Lynn, A. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014,** *54*, 1951.
19. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
20. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18*, 1463.
21. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52*, 7182.
22. Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50*, 1055.
23. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.*, et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.
24. Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties. *J. Chem. Theory Comput.* **2013,** *9*, 3527.
25. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003,** *107*, 9535.
26. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016,** *7*, 207.

27. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994,** *222*, 529.
28. Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994,** *100*, 9025.
29. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22*, 245.
30. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* **2015,** *29*, 397.
31. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006,** *125*, 084902.

Appendix for Chapter 4

(Supplementary Information to Manuscript II)

Section S4.1: Equilibrium simulations of the M^{pro}-substrate complex

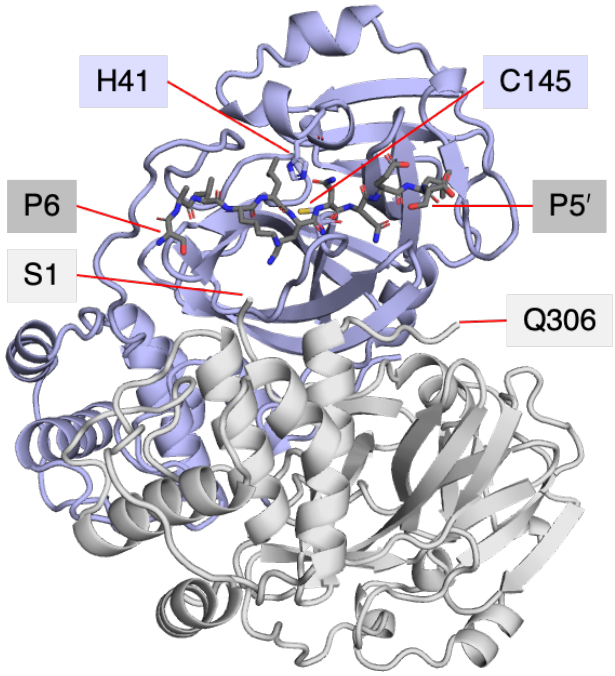


Figure S4.1.1: View of the starting non-covalent complex of dimeric M^{pro} (chains A and B shown as blue and grey cartoons respectively, with the active site His41 and Cys145 labelled and shown as sticks; PDB 6YB7)[1] and the comparatively modelled s05 peptide (shown as dark grey sticks) representing the nsp8/9 native cleavage sequence.[2] Hydrogens are omitted for clarity. Chains A and B are subsequently referred to as ChA and ChB, respectively. The N- and C-terminal residues of ChB are labelled.

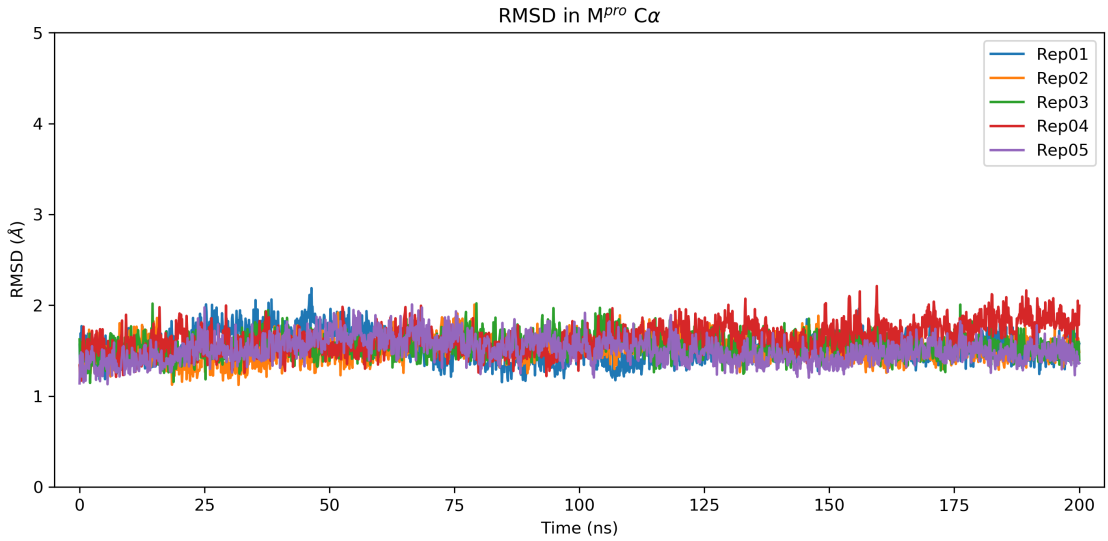


Figure S4.1.2: Time evolution of M^{pro} Cα RMSD in each of the five 200 ns production MD simulations relative to the protein crystal structure. Each replica was subjected to 200 ps NVT equilibration, during which the protein and peptide heavy atoms were restrained, followed by 200 ps NPT equilibration where the restraints were released.

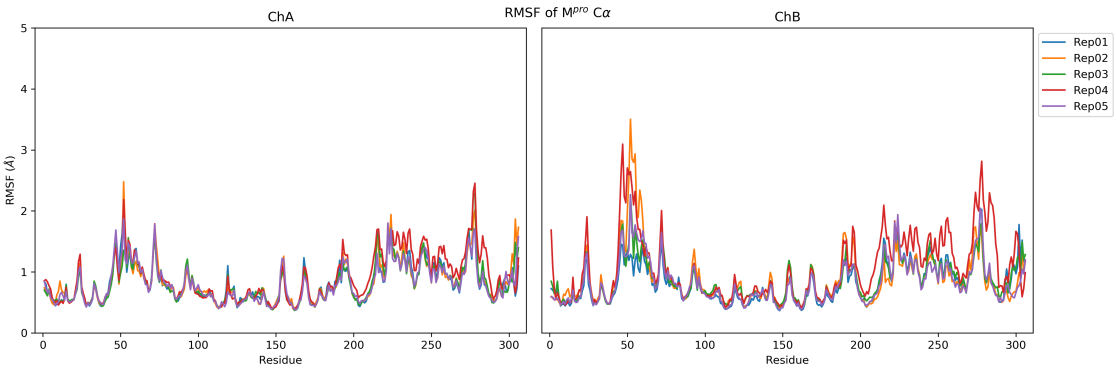


Figure S4.1.3: M^{pro} Cα RMSF in each of the five 150 ns equilibrium MD simulations. The first 50 ns of each 200 ns production MD simulation was discarded as reflecting equilibration.

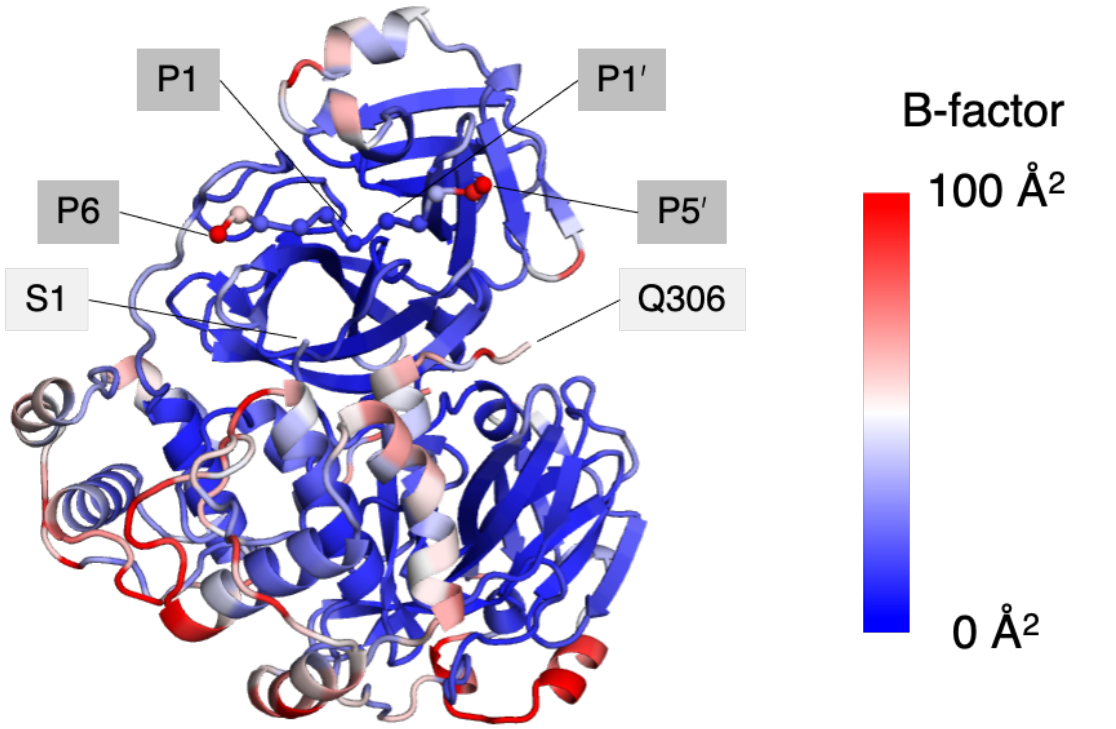


Figure S4.1.4: View of the global average structure of the M^{pro}-s05 complex over the 5 × 150 ns equilibrium MD simulations. The colouring is based on a 0-100 $Å^2$ blue-white-red scale of the MD-derived B-factors, with rigid regions in blue and flexible regions in red. The N- and C-terminal residues of ChB are labelled. The s05 peptide Cα atoms are shown as spheres, with the terminal (P6 and P5′) and the scissile amide (between P1 and P1′) positions labelled.

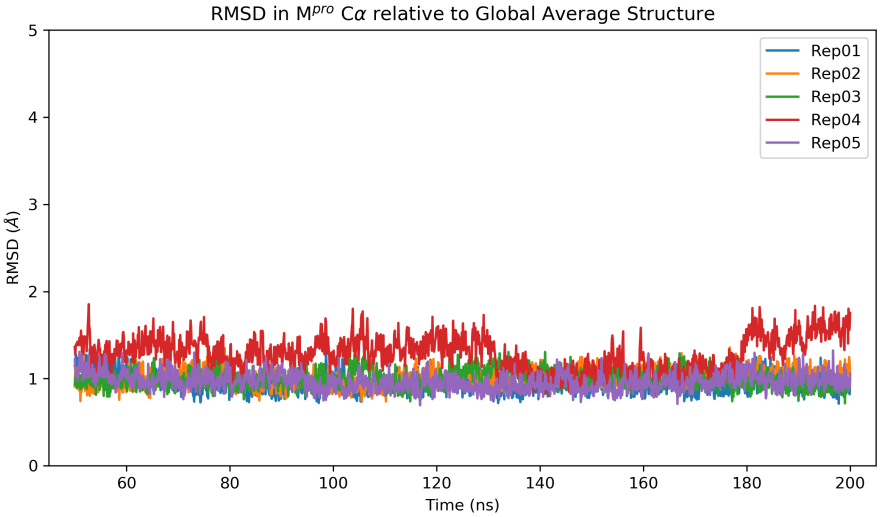


Figure S4.1.5: Time evolution of M^{pro} Cα RMSD in each of the five 150 ns equilibrium MD simulations (starting from t = 50 ns), relative to the equilibrium MD-derived global average structure of the complex.

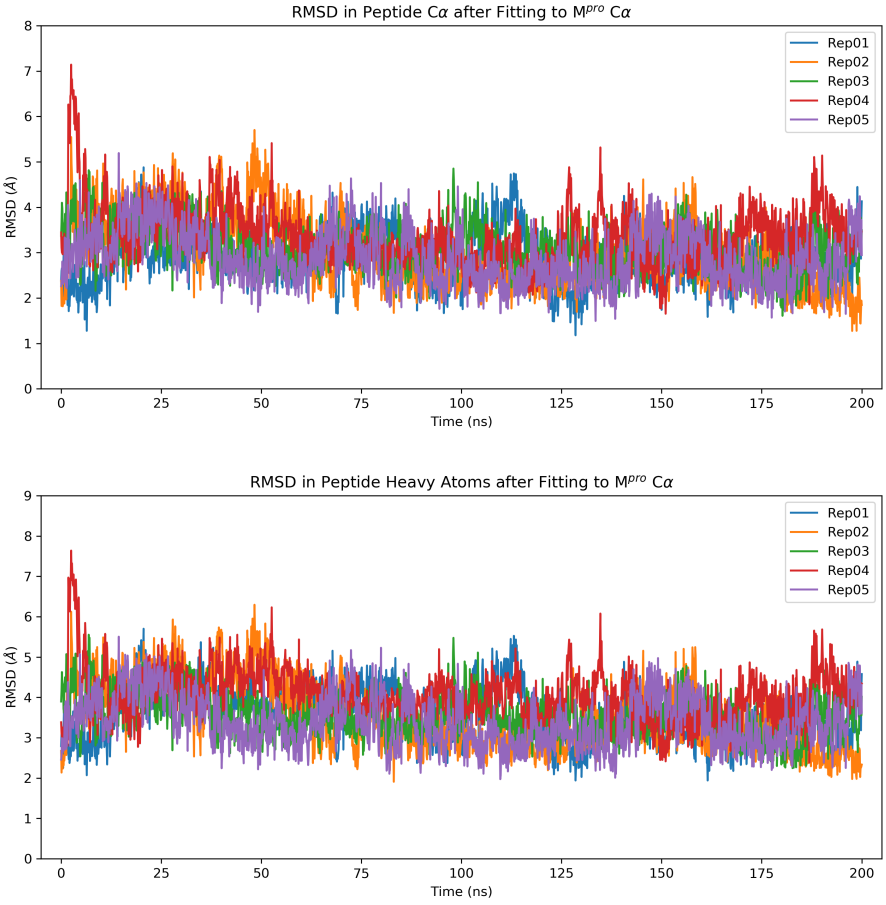


Figure S4.1.6: Time evolution of the s05 substrate peptide (top) Cα and (bottom) heavy atom RMSDs over the 5 × 200 ns production MD simulations relative to the starting conformation, with trajectories fitted based on M^{pro} Cα atoms.

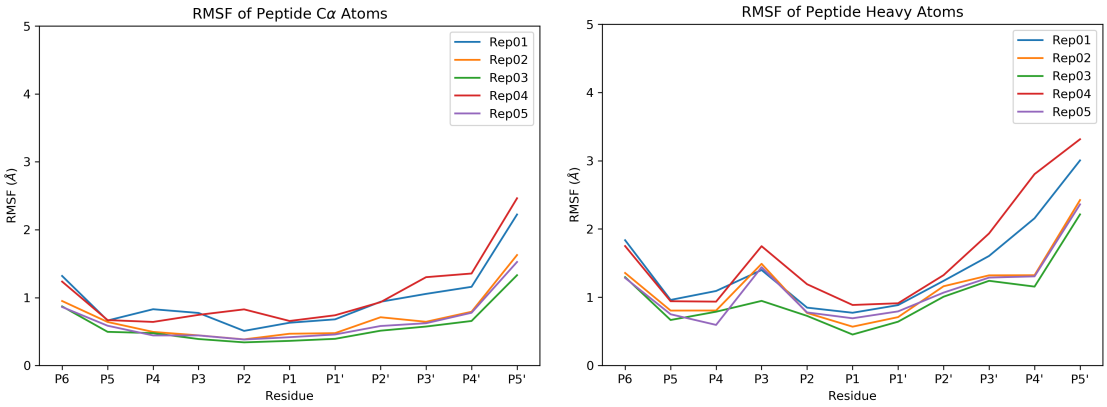


Figure S4.1.7: Per-residue RMSF of the s05 substrate peptide (left) Cα atoms and (right) heavy atoms in the 5 × 150 ns equilibrium MD simulations.

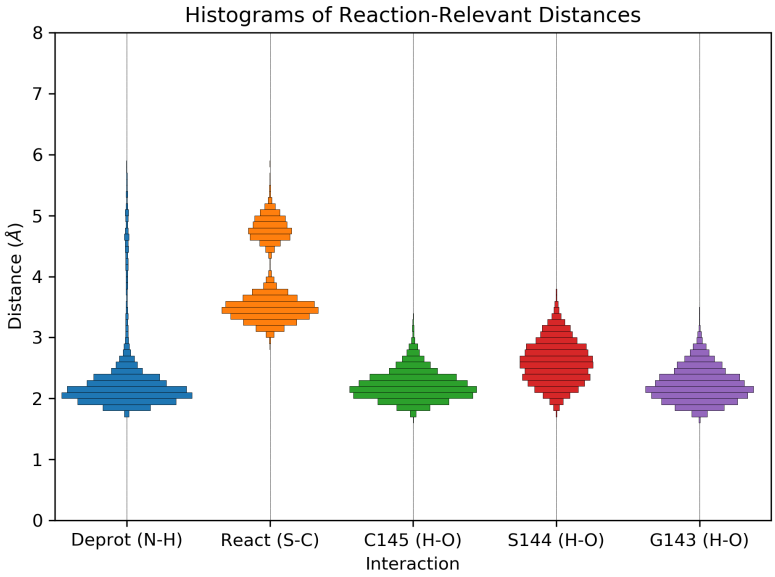


Figure S4.1.8: Distributions of selected distances relevant to the initiation of reaction between M^{pro} Cys145 and the scissile substrate amide bond, observed over the combined 5 × 150 ns equilibrium MD simulations. These include (blue) the His41-NE2 – Cys145-HG distance; (orange) the Cys145-SG – substrate P1-C distance; and the distances from the substrate scissile amide carbonyl P1-O to the backbone NH atoms of (green) Cys145, (red) Ser144, and (purple) Gly143, which together form the oxyanion hole.

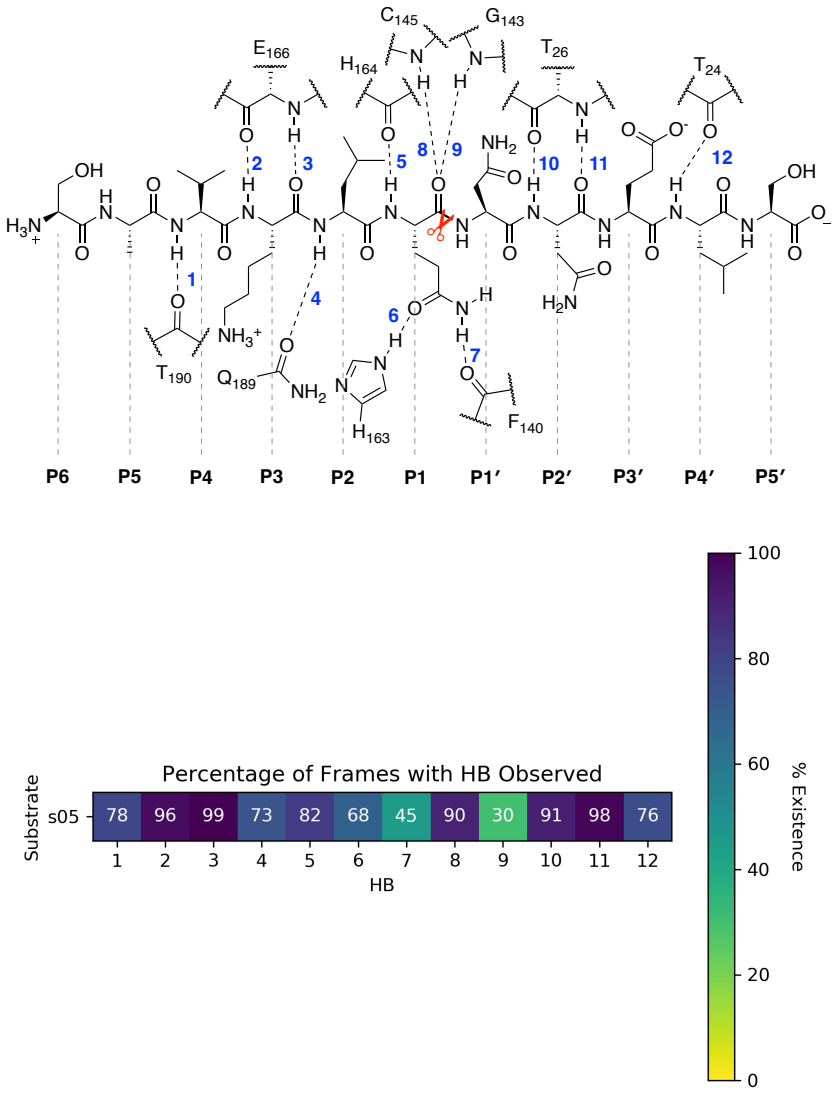


Figure S4.1.9: (top) Illustration of the 12 conserved M^{pro}-substrate hydrogen bonds (HBs)[2] exemplified by s05; (bottom) a heatmap showing the percentage of frames (analysed every ns) where the HB is observed, over the combined 5 × 150 ns equilibrium MD simulations.

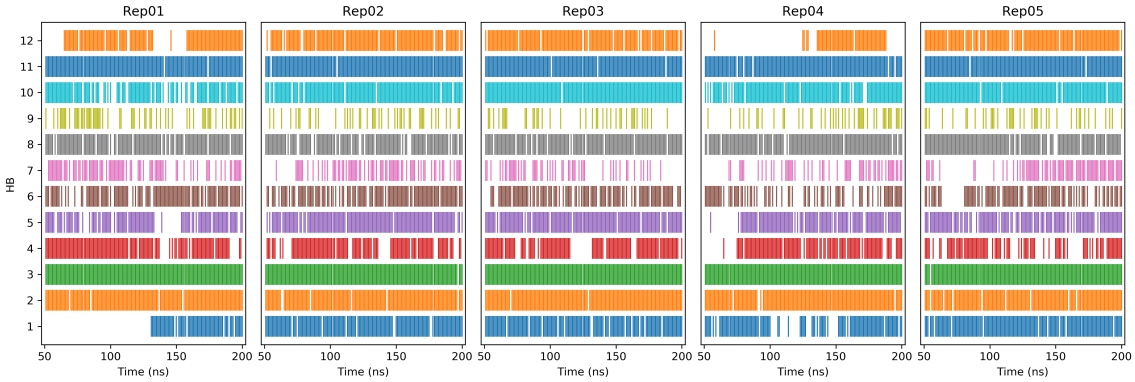


Figure S4.1.10: Event plot showing the existence of the 12 conserved HBs (analysed every ns) in each of the five 150 ns equilibrium MD simulations.

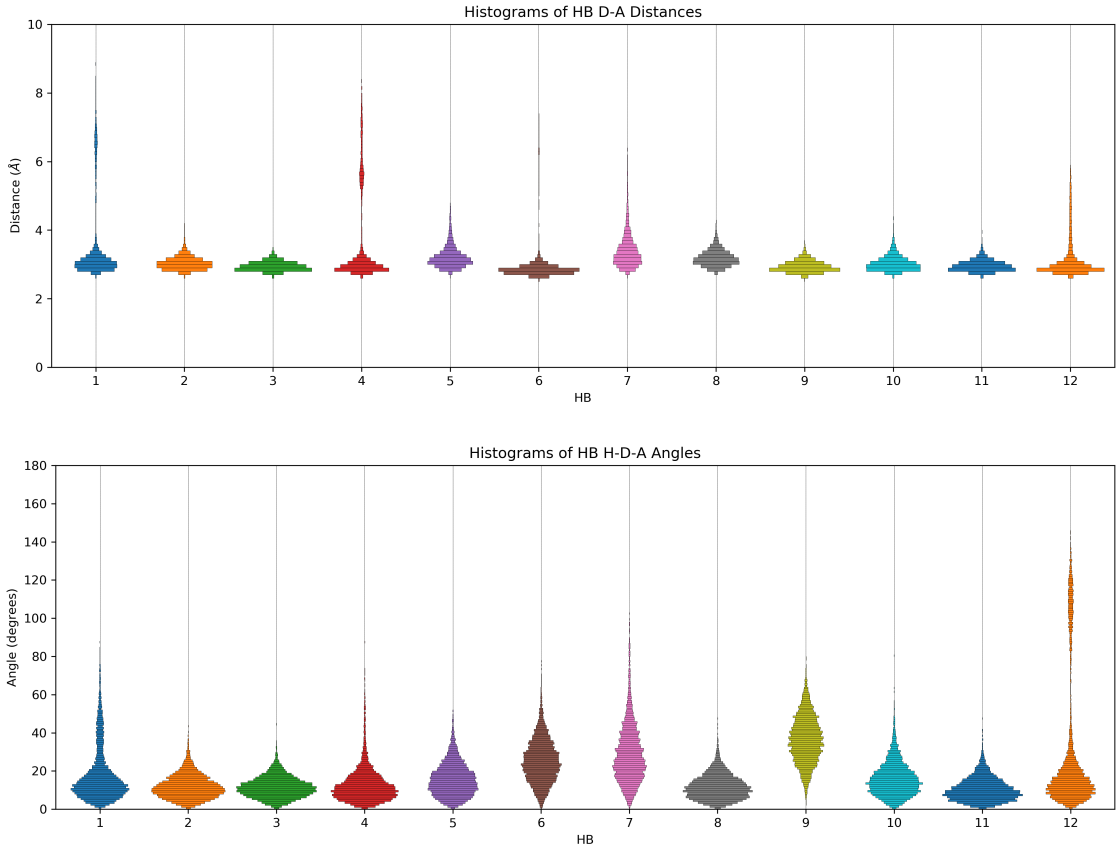


Figure S4.1.11: Distributions of (top) the donor-acceptor distances ($d_{\text{D-A}}$) and (bottom) the hydrogen-donor-acceptor angles (∠(H-D-A)) for each of the 12 conserved HBs, observed over the combined 5 × 150 ns equilibrium MD simulations. HBs are defined based on a combined distance ($d_{\text{D-A}} \leq 3.5$ Å) and angle (∠(H-D-A) ≤ 30°) criteria.

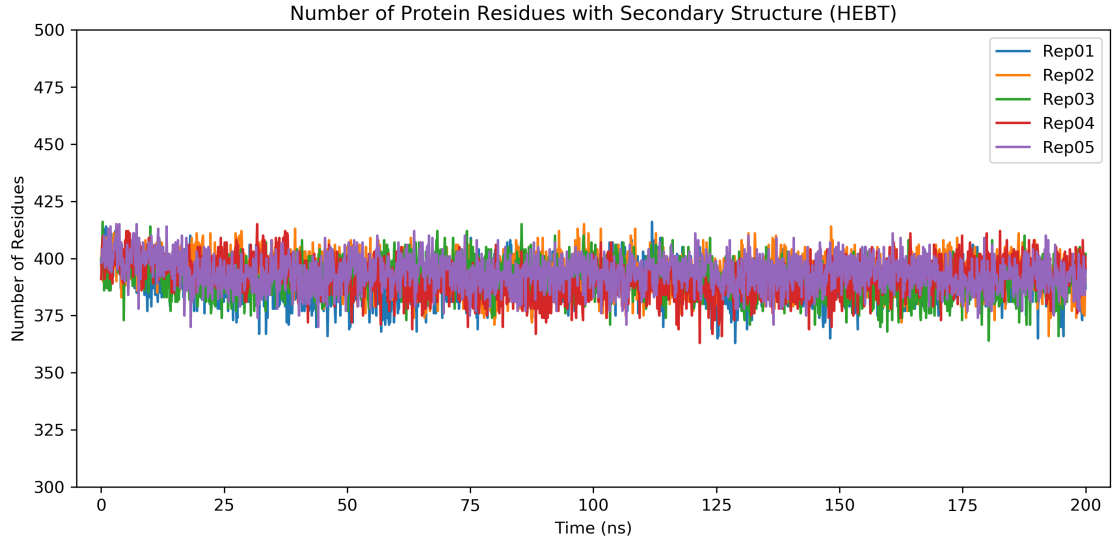


Figure S4.1.12: Time evolution of the number of M^{pro} residues possessing secondary structure (H = α-helix; E = extended strand; B = β-bridge; T = hydrogen bonded turn) in each of the five 200 ns production MD simulations, as defined by DSSP (version 2.0.4).[3, 4]

Section S4.2: Immediate effect of substrate removal

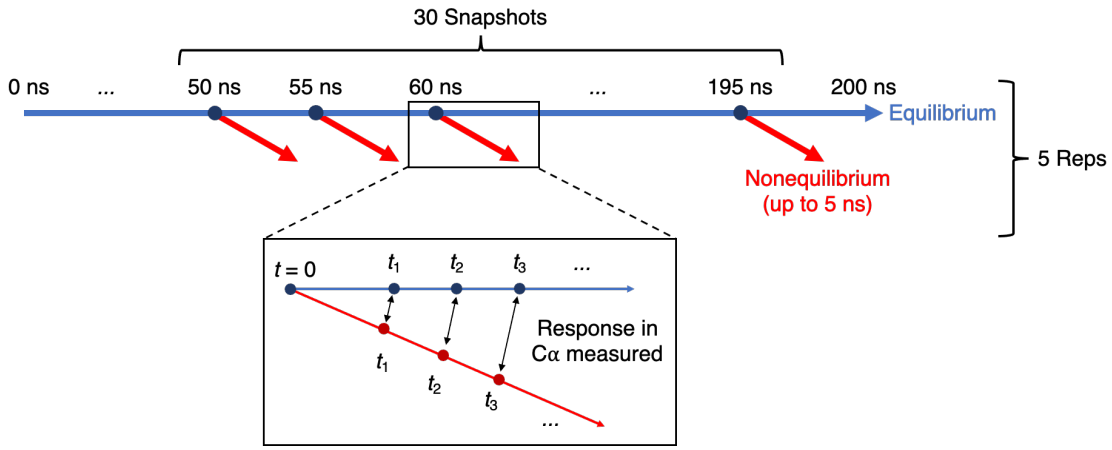


Figure S4.2.1: Scheme of the procedure used to set up and analyse the D-NEMD simulations. Five equilibrium MD simulations, each 200 ns, were performed for the M^{pro}-s05 substrate complex. These equilibrium simulations were used to generate starting structures for the nonequilibrium simulations. From the equilibrated part of each s05-bound simulation (from 50-200 ns), conformations were extracted every 5 ns, and the s05 peptide was removed from the system. The nonequilibrium simulations started from this ensemble of 150 conformations. The Kubo-Onsager approach was used to extract the response of M^{pro} to s05 peptide annihilation.[5, 6] For that, for each pair of equilibrium s05-bound and nonequilibrium *apo* simulations, the change in position for each Cα between the equilibrium and nonequilibrium trajectories at equivalent times was determined and averaged over all 150 simulations.

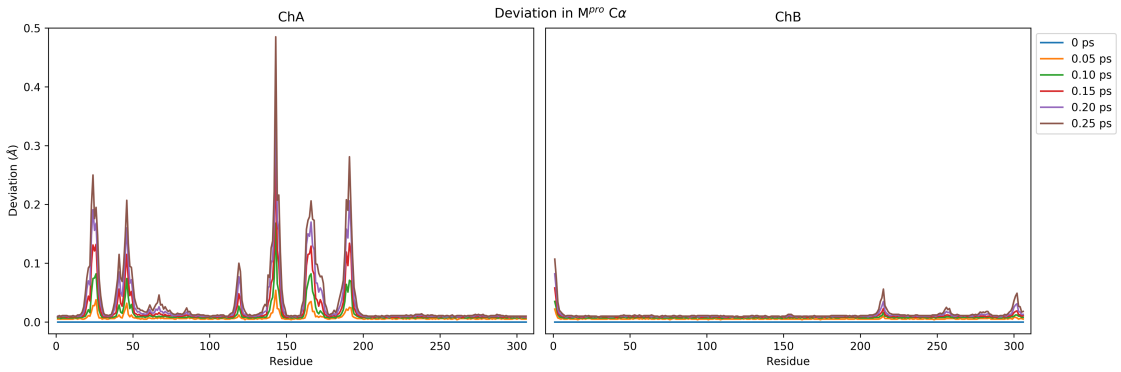


Figure S4.2.2: Cα deviations in (left) chain A and (right) chain B of the M^{pro} dimer as observed in the first six time points (saved in 0.05 ps intervals) up to 0.25 ps following the removal of s05.

Table S4.2.1: The five most perturbed M^{pro} residues, in terms of Cα deviations, at selected time points within the 2 ps following the removal of s05.

Rank	M^{pro} residues with highest Cα deviations (Å) at each time point following perturbation											
	t = 0.10 ps		*t* = 0.25 ps		*t* = 0.50 ps		*t* = 0.75 ps		*t* = 1.00 ps		*t* = 2.00 ps	
1	G143	0.168	G143	0.485	G143	0.623	A191	0.706	A191	0.813	A191	1.132
2	T26	0.082	A191	0.281	A191	0.547	G143	0.693	G143	0.781	T190	0.905
3	E166	0.082	T24	0.250	T24	0.448	G23	0.603	G23	0.684	Q189	0.888
4	M165	0.078	N142	0.222	N142	0.425	T24	0.558	T24	0.651	T24	0.875
5	T25	0.075	C145	0.216	G23	0.409	N142	0.542	S46	0.602	S46	0.850

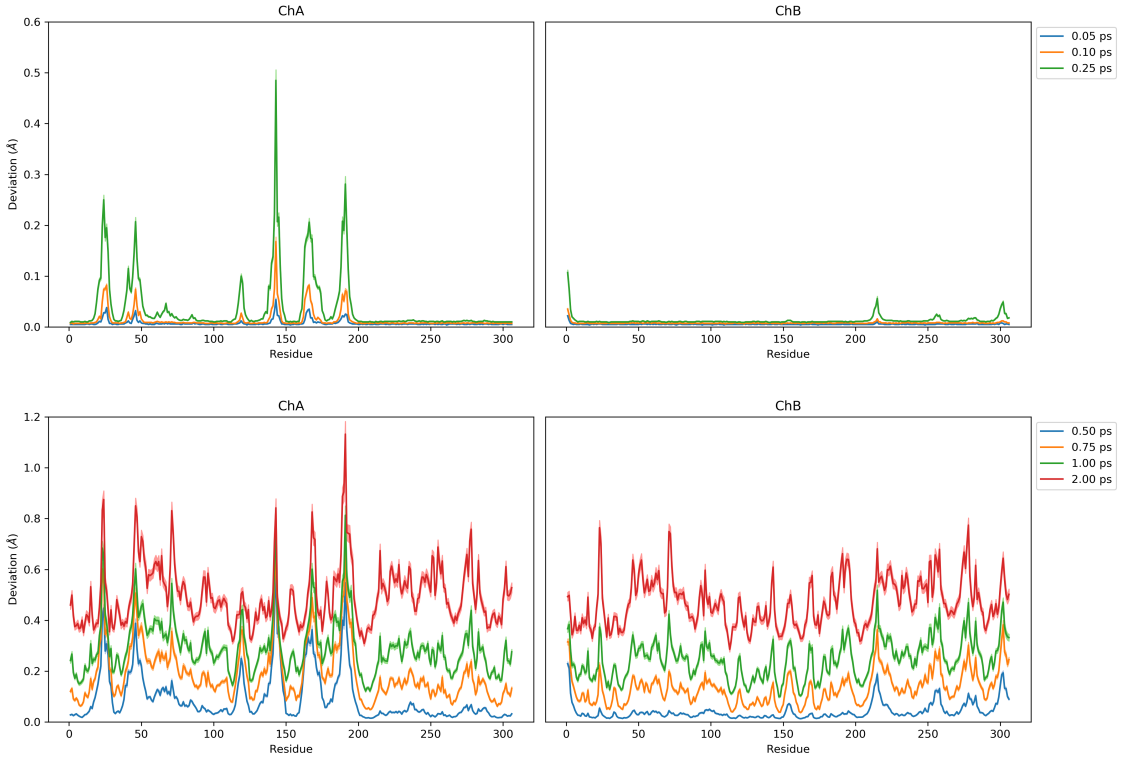


Figure S4.2.3: Cα deviations as observed at different time points up to 2 ps following perturbation, with ± standard error of the mean (SEM) shown as shaded regions. Note that the two sets of plots are shown at different scales for clarity.

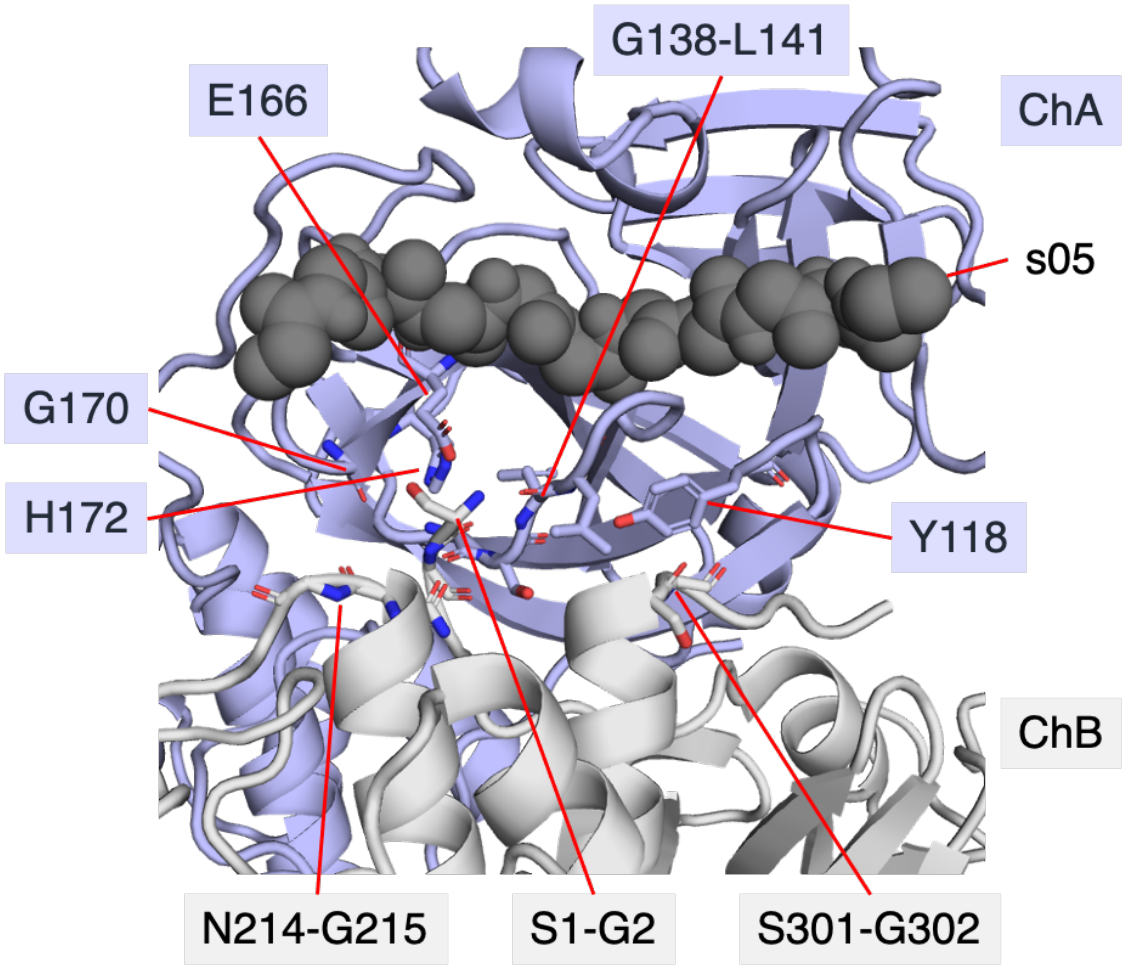


Figure S4.2.4: View of the dimer interface near the s05 binding site in the M^{pro}-s05 complex prior to MD simulations, with chains A and B shown in blue and grey, respectively, and the s05 peptide backbone shown as black spheres. Shown as sticks are chain B residues displaying substantial initial responses to s05 removal (deviation >15 pm at t = 0.50 ps), as well as their surrounding (within 3 Å) chain A residues, with labels in the corresponding chain colours.

Section S4.3: Propagation of signal to allosteric sites

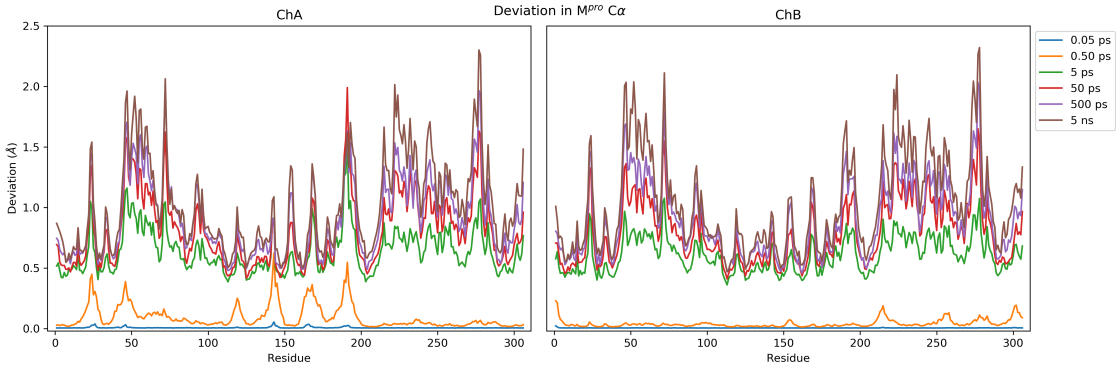


Figure S4.3.1: Cα deviations over the full range of the sampled timescales.

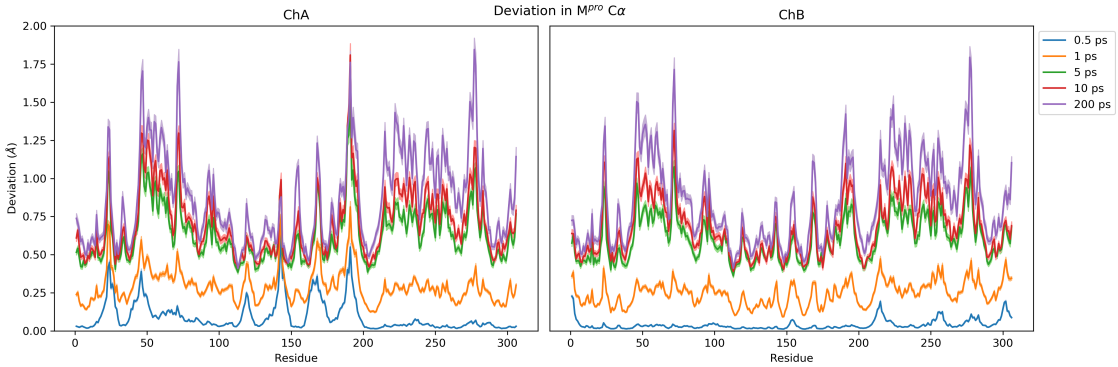


Figure S4.3.2: Cα deviations up to 200 ps, with ± SEM shown as shaded regions.

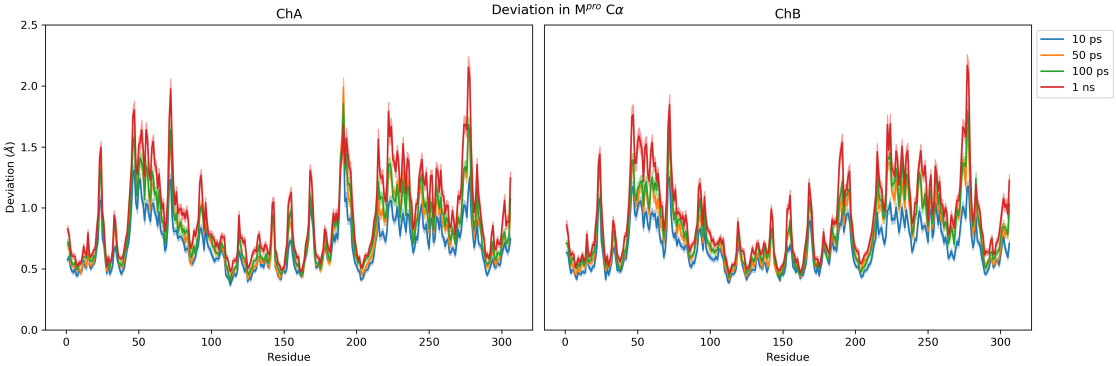


Figure S4.3.3: Cα deviations up to 1 ns, with ± SEM shown as shaded regions.

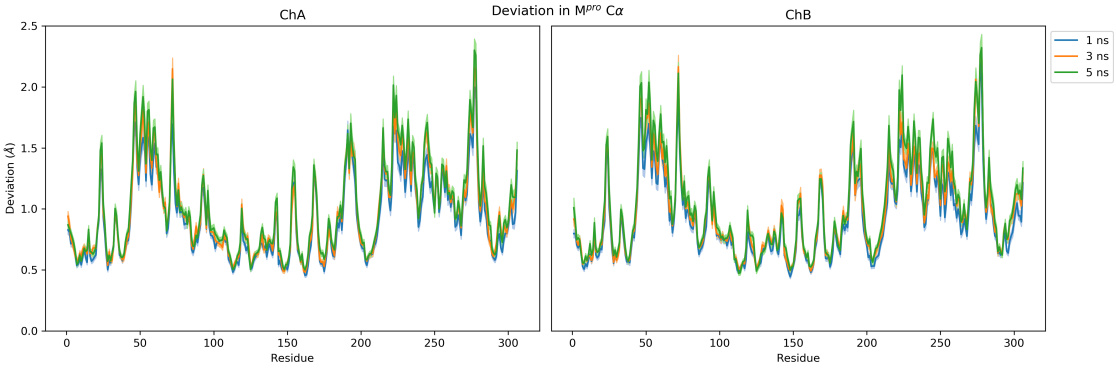


Figure S4.3.4: Cα deviations up to 5 ns, with ± SEM shown as shaded regions.

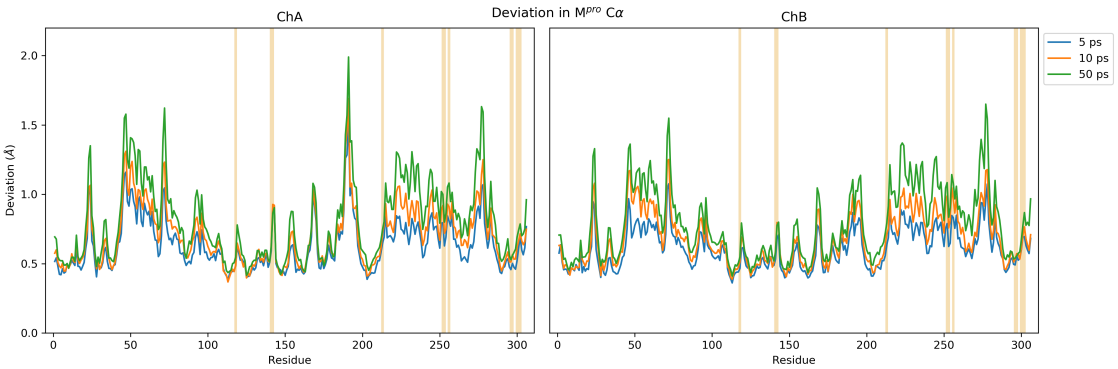


Figure S4.3.5: Cα deviations, with residues constituting the pelitinib binding allosteric site in both chains A and B (see **Figure 4.2** in main text) highlighted in light orange.

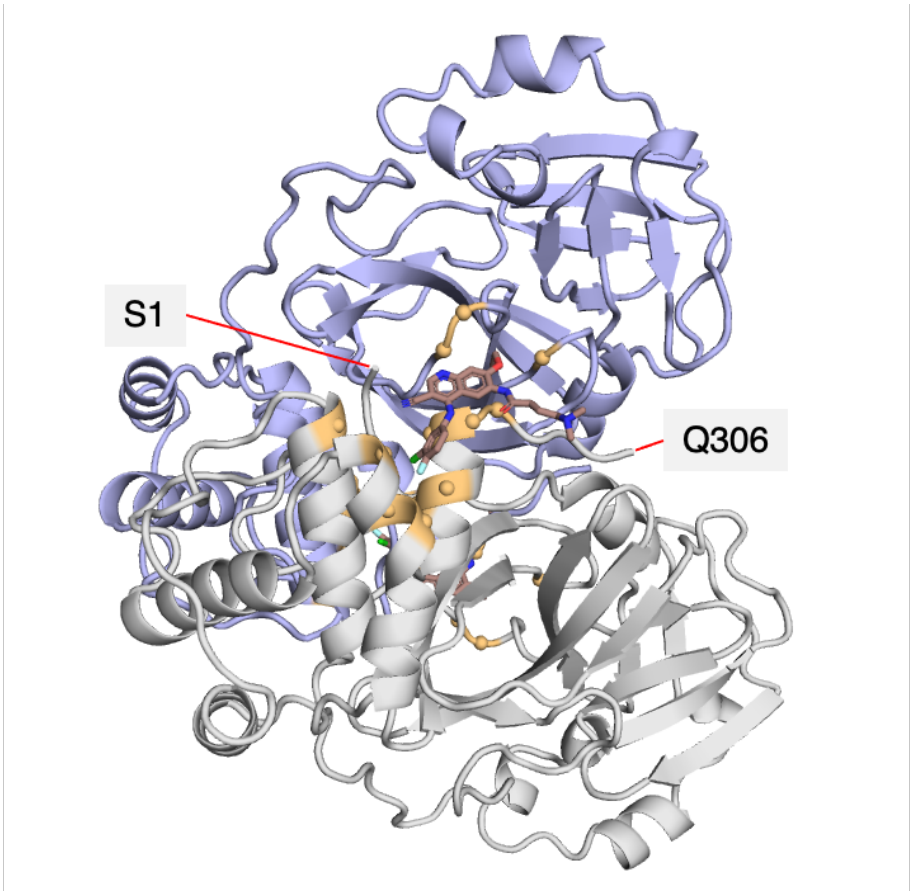


Figure S4.3.6: View of the M^{pro} dimer (ChA and ChB shown in blue and grey respectively, with the N- and C-terminal residues of ChB labelled) complexed with two molecules of the allosterically binding inhibitor pelitinib (shown as brown sticks) (PDB 7AXM).[7] Residues constituting the allosteric sites (within 4 Å of pelitinib; see **Figure 4.2** in main text) are shown in orange.

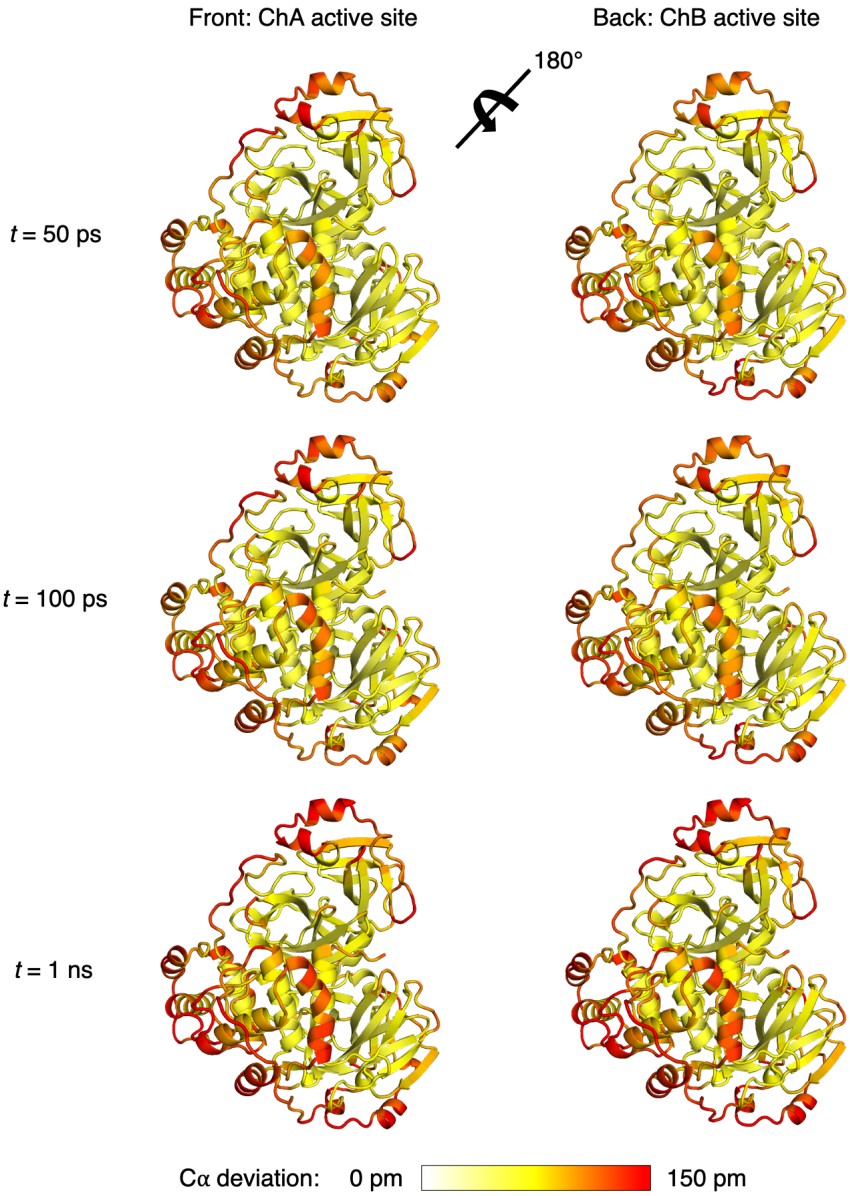


Figure S4.3.7: 180° rotated views of the M^{pro}-s05 complex prior to MD simulations, and the response to s05 removal in terms of average Cα deviations at t = 50 ps, 100 ps, and 1 ns, represented on a white-yellow-red scale. All figures were generated with PyMOL.[8]

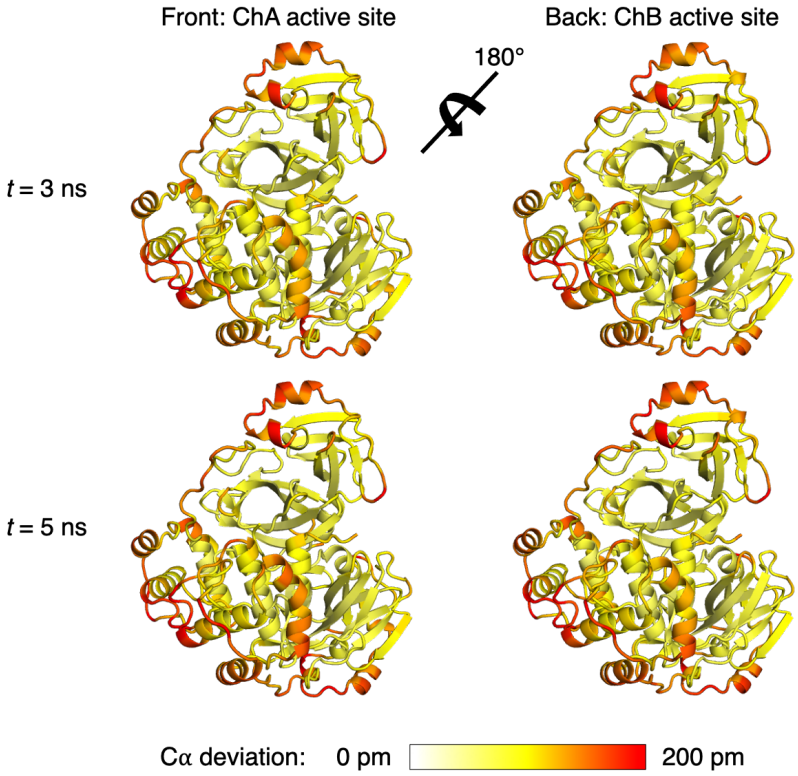


Figure S4.3.8: 180° rotated views of the M^{pro}-s05 complex prior to MD simulations, and the response to s05 removal in terms of average Cα deviations beyond 1 ns, represented on a white-yellow-red scale. Note the different scale used here compared to the previous figure to illustrate more clearly the intensified responses. All figures were generated with PyMOL.[8]

Section S4.4: Coherent displacements in residues upon substrate removal

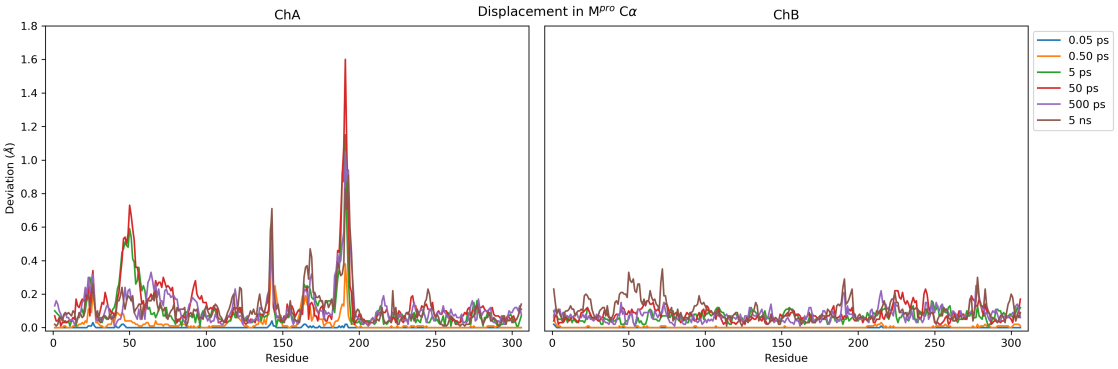


Figure S4.4.1: Magnitudes of average Cα displacement vectors over the full range of sampled timescales.

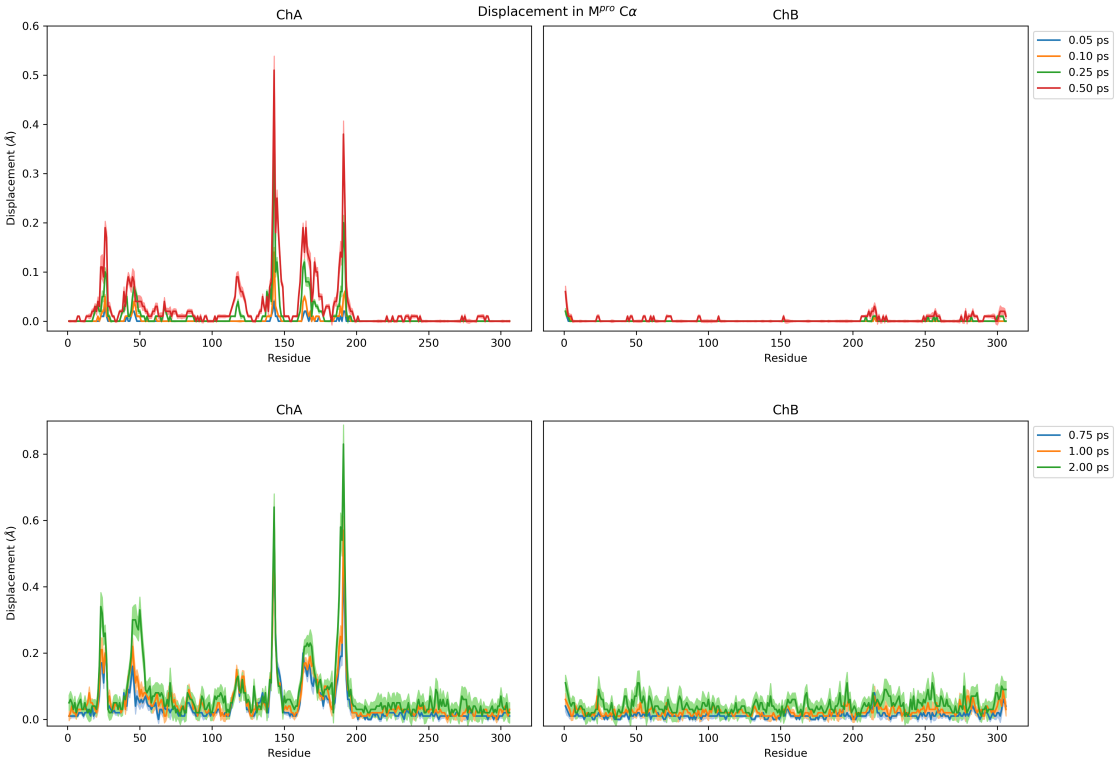


Figure S4.4.2: Magnitudes of average Cα displacement vectors at different time points up to 2 ps, with ± SEM shown as shaded regions. Note that the two sets of plots are shown at different scales for clarity.

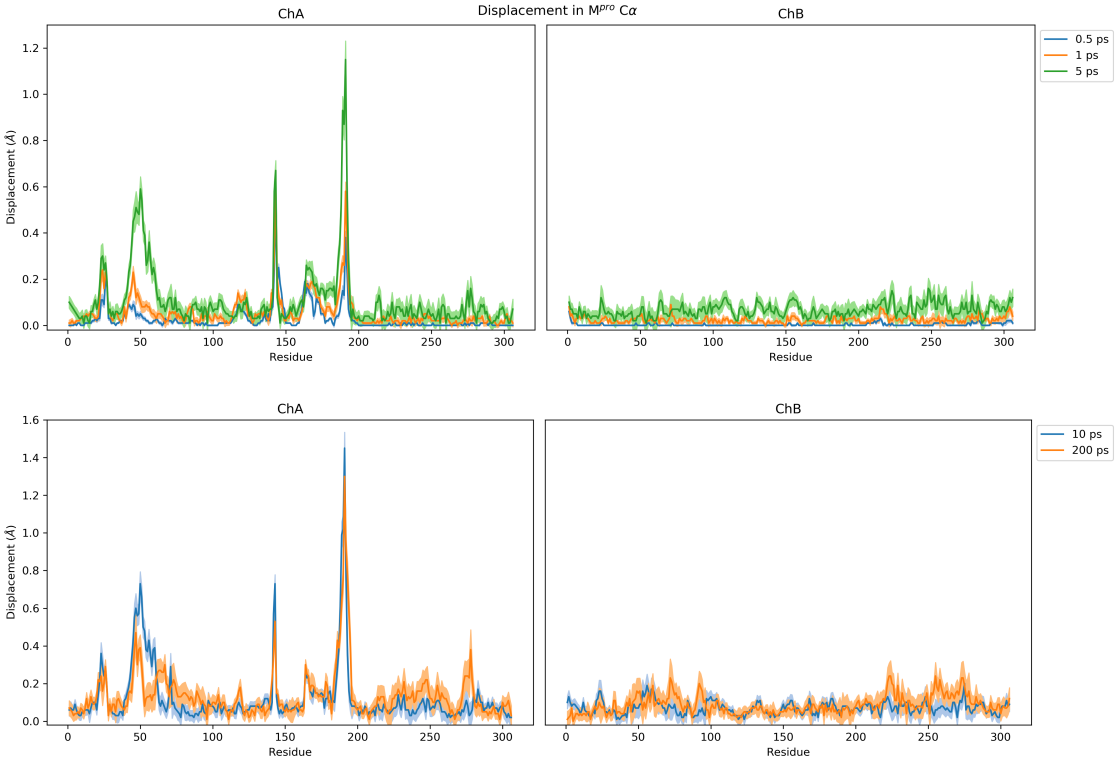


Figure S4.4.3: Magnitudes of average Cα displacement vectors at different time points up to 200 ps, with ± SEM shown as shaded regions. Note that the two sets of plots are shown at different scales for clarity.

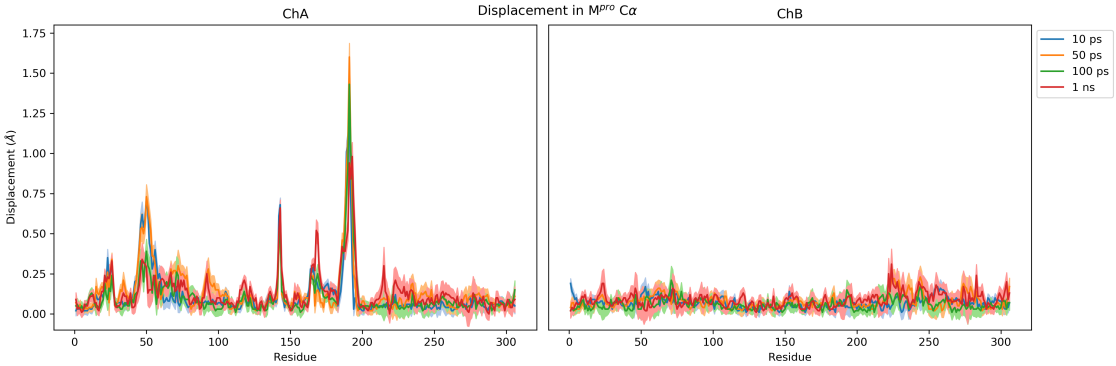


Figure S4.4.4: Magnitudes of average Cα displacement vectors up to 1 ns, with ± SEM shown as shaded regions.

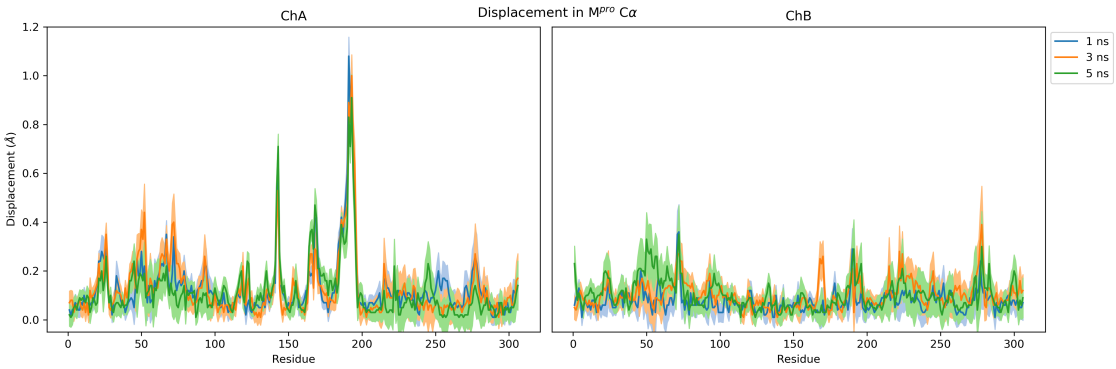


Figure S4.4.5: Magnitudes of average Cα displacement vectors up to 5 ns, with ± SEM shown as shaded regions.

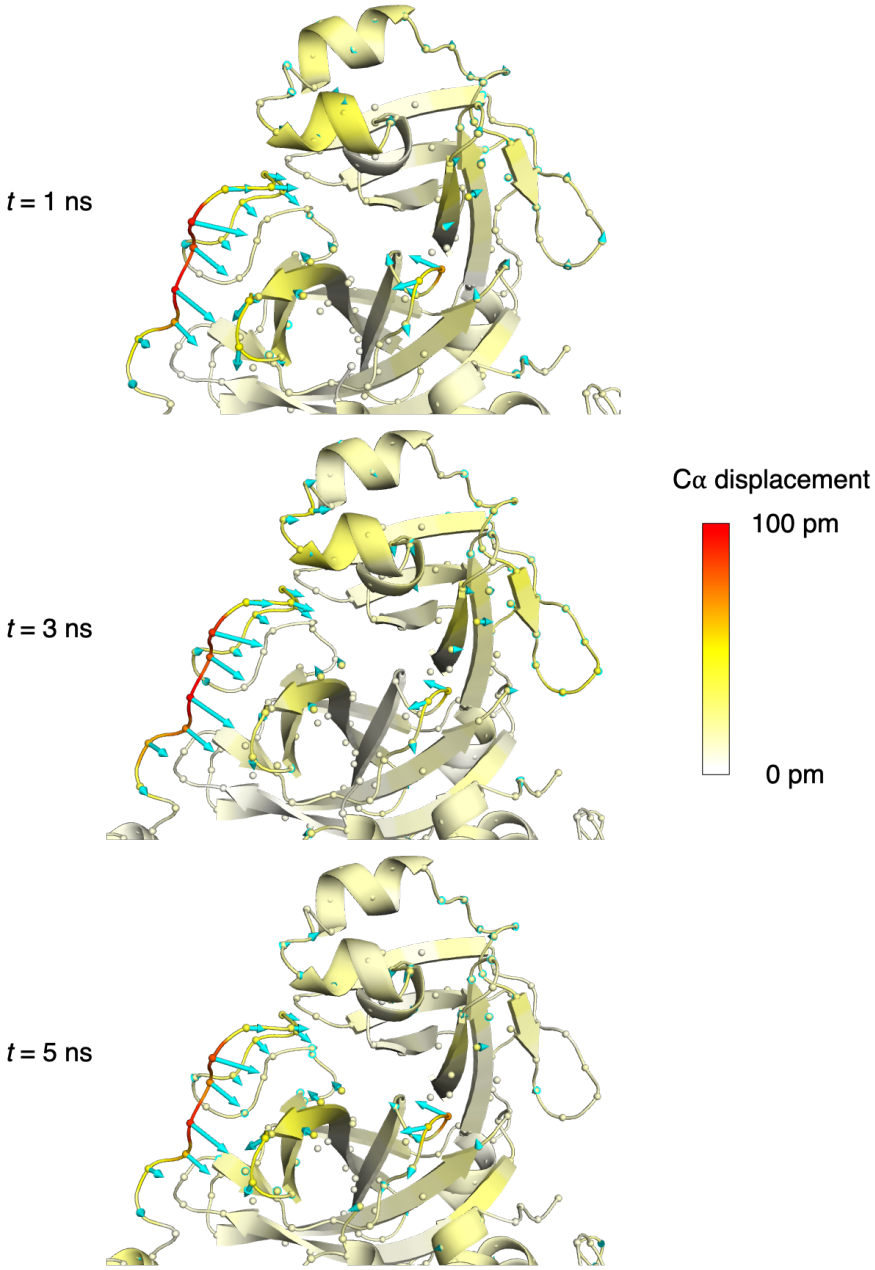


Figure S4.4.6: Views of the M^{pro}-s05 complex prior to MD simulations with a focus around the original substrate binding site, and the M^{pro} response to s05 removal beyond 1 ns from averaging Cα displacement vectors. Displacement magnitudes are represented on a white-yellow-red scale.[8] Vectors with length ≥15 pm are displayed as cyan arrows with a scale-up factor of 5.[9]

Section S4.5: Methods

Table S4.5.1: Protonation states of the histidine residues of SARS-CoV-2 M^{pro} (PDB 6YB7)[1] in this study. The setup is identical to that reported in a previous study.[2] Following standard AMBER force field nomenclature,[10] neutral histidine residues that are Nδ-protonated and Nε-protonated are referred to as "HID" and "HIE" respectively.

His	State
41	HID
64	HIE
80	HID
163	HIE
164	HIE
172	HIE
246	HIE

References

1. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
2. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
3. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.*, et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015,** *43*, D364.
4. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577.
5. Ciccotti, G.; Ferrario, M. Non-equilibrium by molecular dynamics: a dynamical approach. *Mol. Simul.* **2016,** *42*, 1385.
6. Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur. Phys. J. B* **2021,** *94*, 144.
7. Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.*, et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021,** *372*, 642.
8. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
9. Law, S. M. modevectors.py. https://pymolwiki.org/index.php/Modevectors (accessed 2022-06-30).
10. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.

Appendix for Chapter 5

(Supplementary Information to Manuscript III)

Section S5.1: Equilibrium simulations of M^{pro}-peptide complexes

The peptides considered here include the substrate peptides s01 (nsp4/5) and s02 (nsp5/6), the inhibitor peptide p12,[1] and the P1 Gln-to-Ala mutated peptides s01-QP1A and s05-QP1A. The equilibrium molecular dynamics (MD) simulations for s05 (nsp8/9) have been previously reported (see **Manuscript II** in **Chapter 4** and **Appendix for Chapter 4**).[2] The analyses are presented similarly as previously reported.[2]

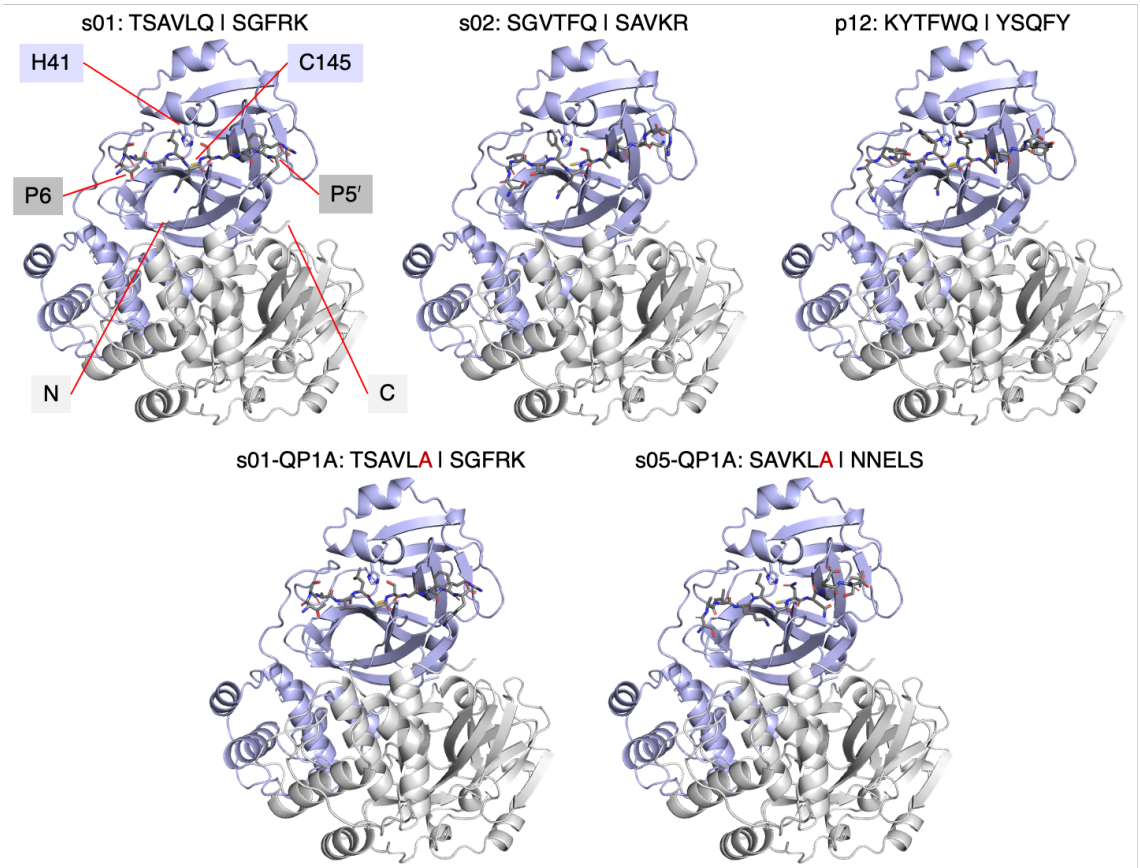


Figure S5.1.1: View of the starting non-covalent complexes of dimeric M^{pro} (chains A and B shown as blue and grey cartoons, respectively, with the active site His41 and Cys145 shown as sticks and labelled in the s01-bound complex; PDB 6YB7)[3] and the comparatively modelled s01, s02, p12, s01-QP1A, and s05-QP1A peptides (all shown as dark grey sticks).[1] Hydrogens are omitted for clarity. Chains A and B are subsequently referred to as ChA and ChB, respectively. The N- and C-terminal residues of ChB are labelled in the s01-bound complex.

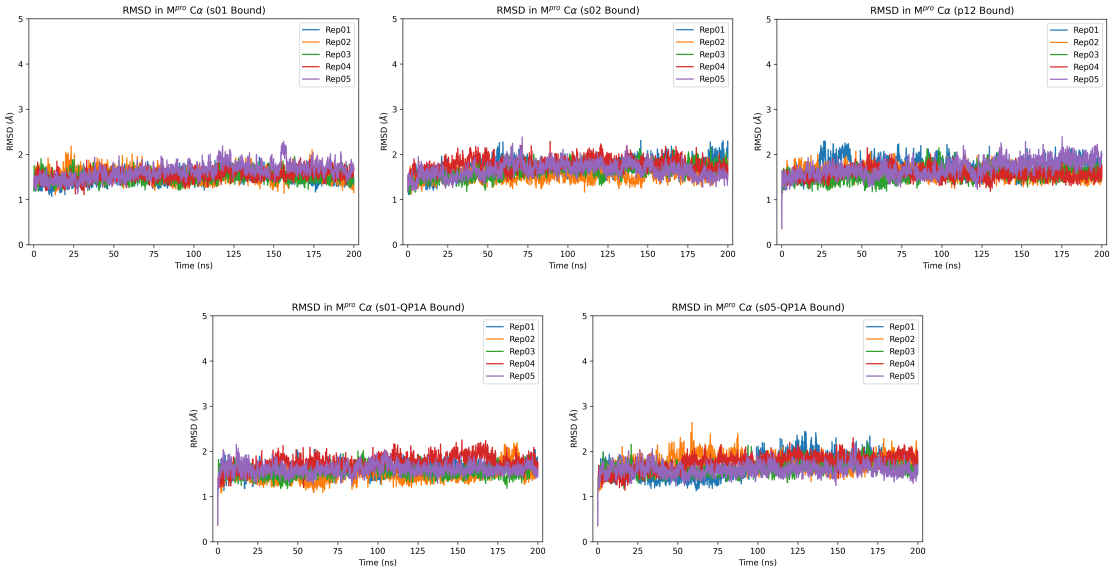


Figure S5.1.2: Time evolution of M^{pro} Cα root mean square deviation (RMSD) in each of the five 200 ns production molecular dynamics (MD) simulations relative to the protein crystal structure for each M^{pro}-peptide complex. Each replica was subjected to 200 ps NVT equilibration, during which the protein and peptide non-hydrogen atoms were restrained. This was followed by 200 ps NPT equilibration, during which the restraints were released for the native substrate (s01, s02) bound M^{pro} systems, and backbone restraints were retained in the other cases (p12, s01-QP1A, s05-QP1A bound M^{pro}).

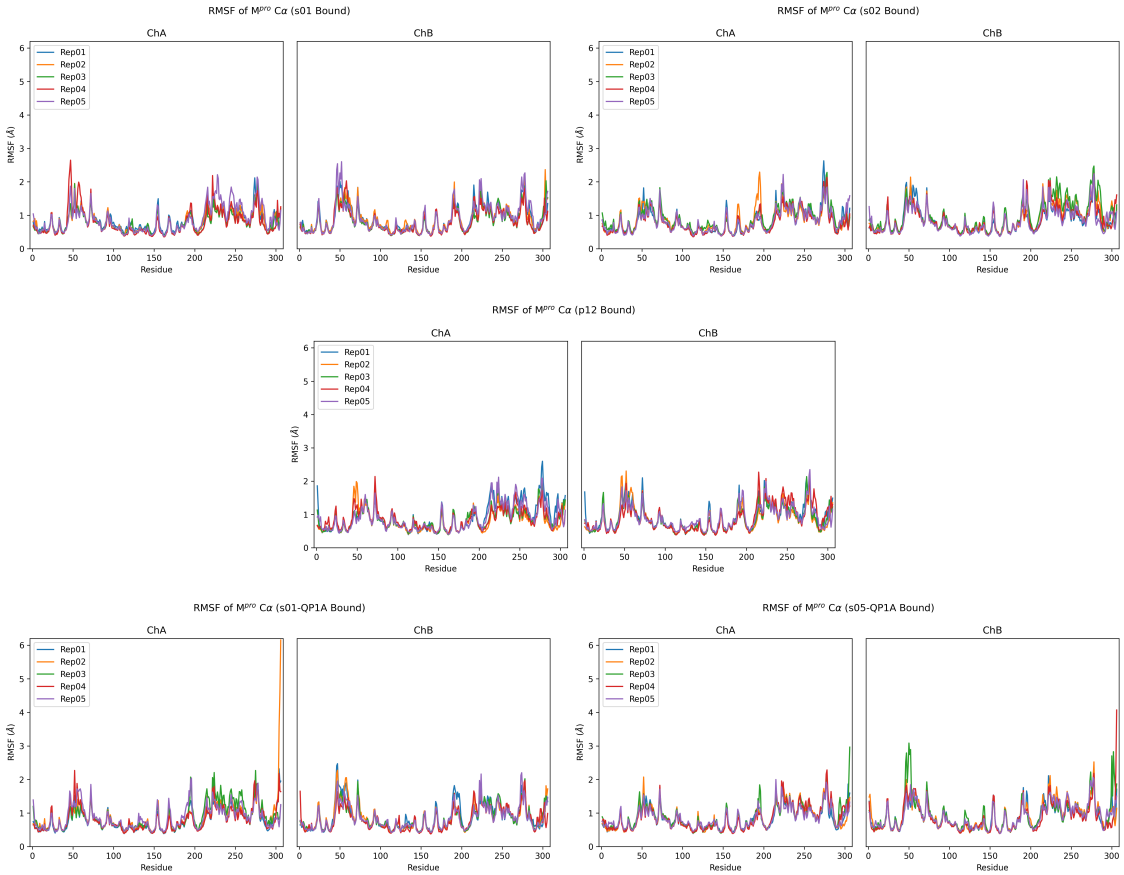


Figure S5.1.3: M^{pro} Cα root mean square fluctuation (RMSF) in each of the five 150 ns equilibrium MD simulations, for each M^{pro}-peptide complex. The first 50 ns of each 200 ns production MD simulation was discarded as equilibration.

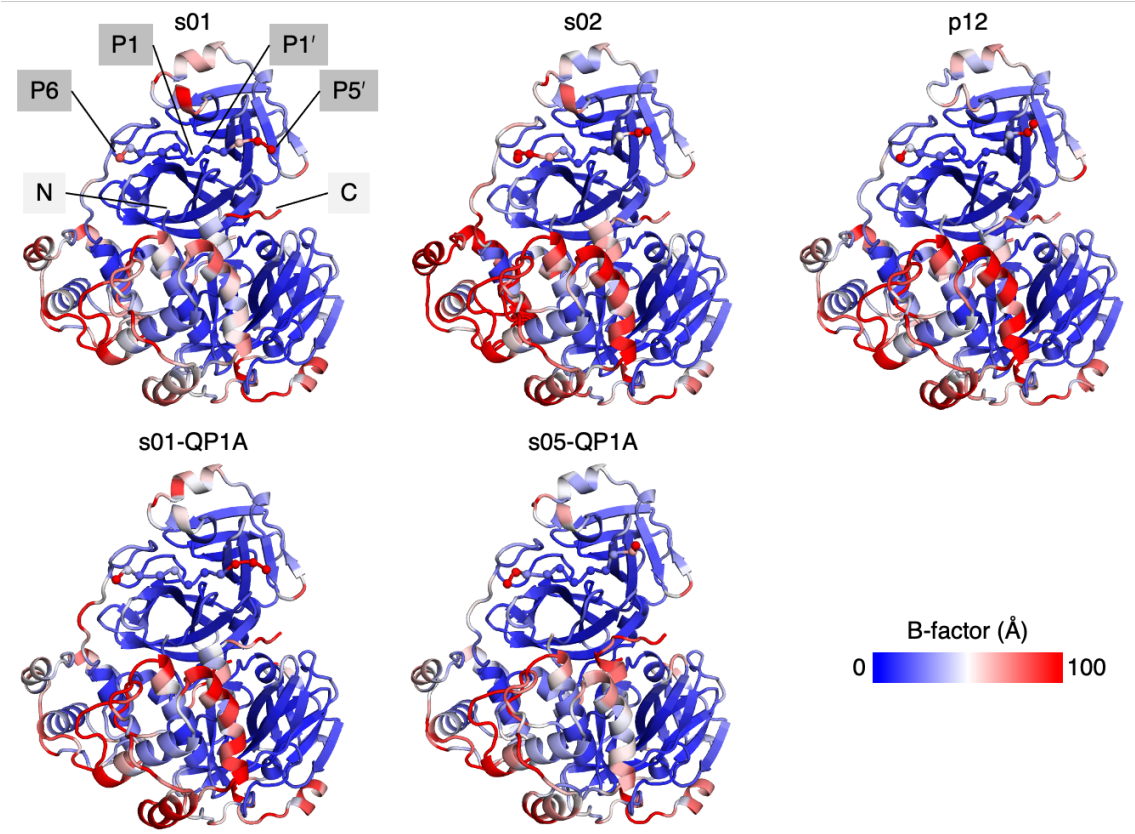


Figure S5.1.4: Views of the global average structure of each M^{pro}-peptide complex over the combined 5 × 150 ns equilibrium MD simulations. The colouring is based on a 0-100 Å^2 blue-white-red scale of the MD-derived B-factors, with rigid regions in blue and flexible regions in red. The N- and C-terminal residues of ChB are labelled in the s01-bound complex. The peptide Cα atoms are shown as spheres, with the terminal (P6 and P5′) and the scissile amide (between P1 and P1′) positions labelled for s01.

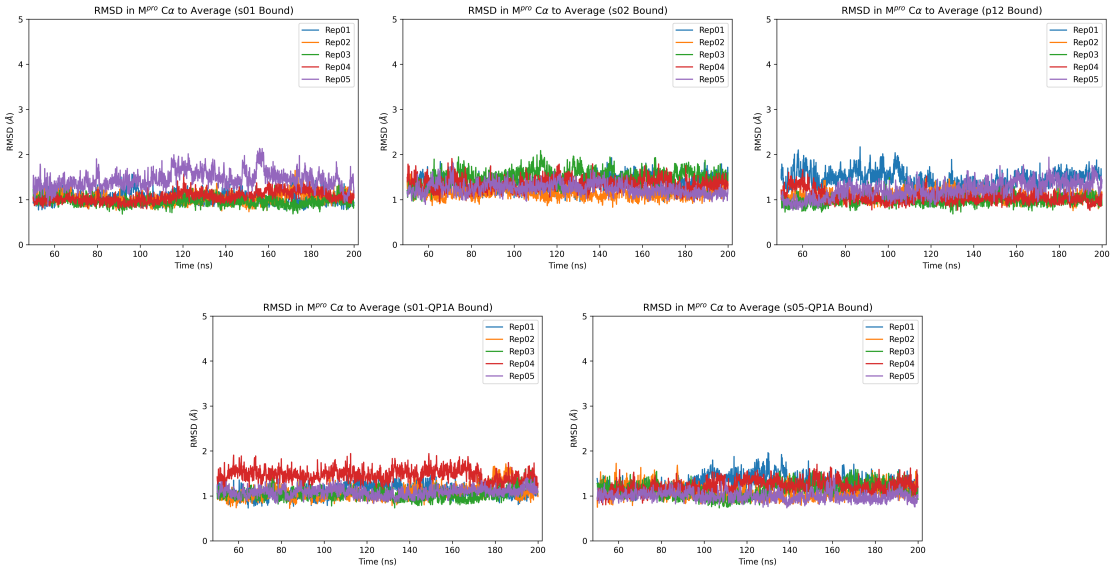


Figure S5.1.5: Time evolution of M^{pro} Cα RMSD in each of the five 150 ns equilibrium MD simulations (starting from t = 50 ns), relative to the equilibrium MD-derived global average structure of the complex, for each M^{pro}-peptide complex.

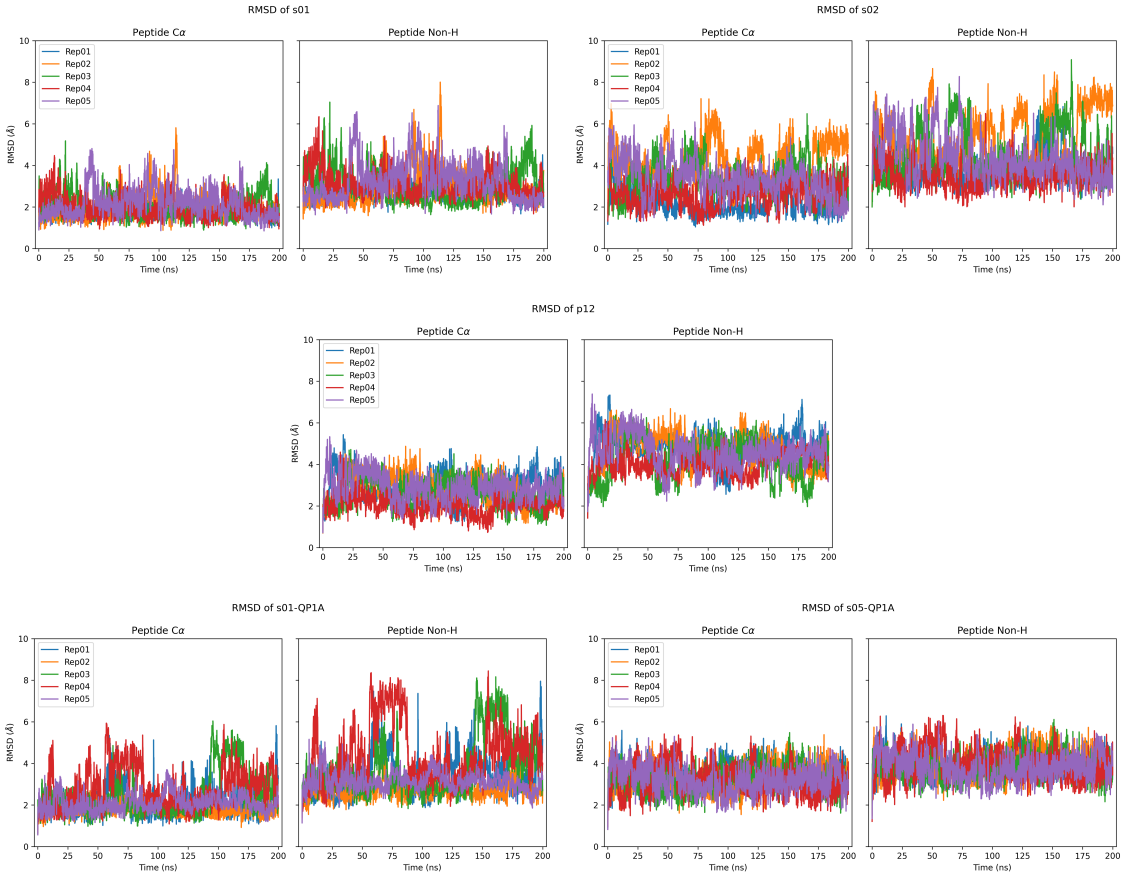


Figure S5.1.6: Time evolution of the peptide Cα and non-hydrogen atom RMSDs over the 5 × 200 ns production MD simulations relative to the starting conformation, with trajectories fitted based on M^{pro} Cα atoms, for each M^{pro}-peptide complex.

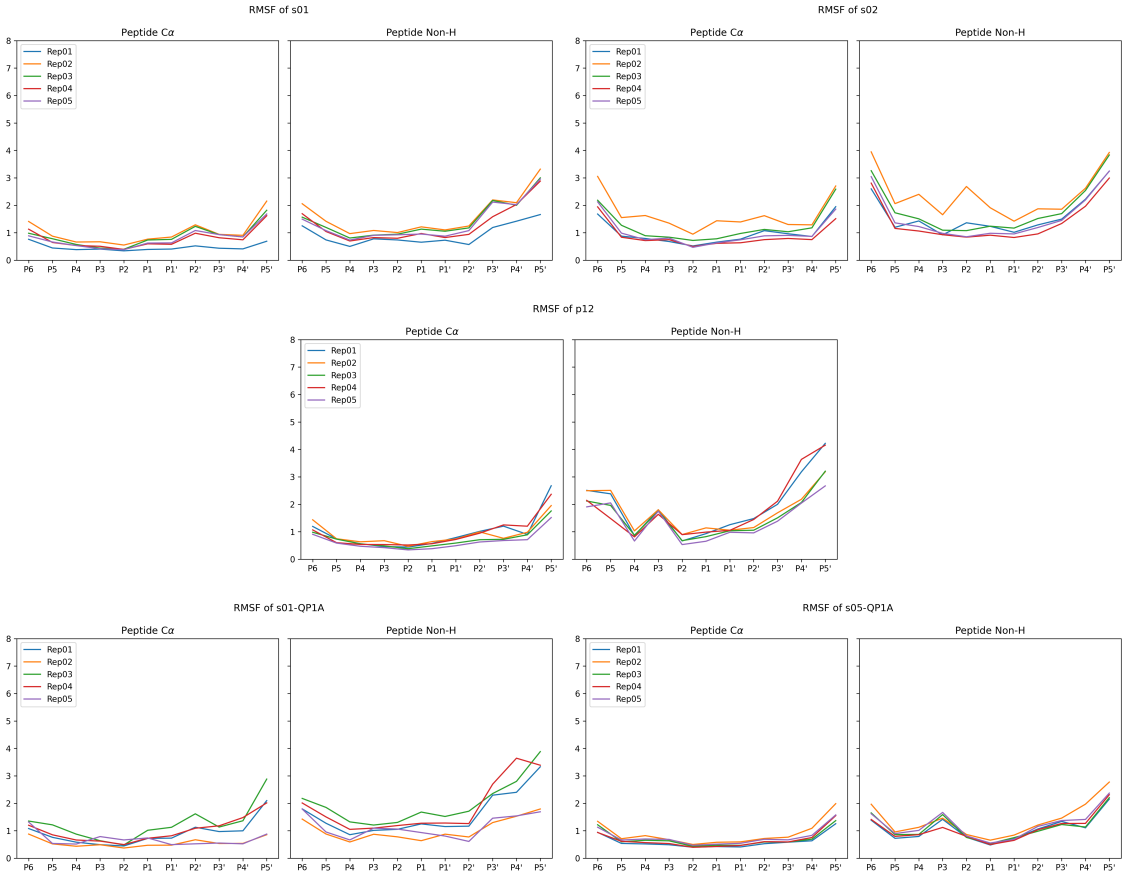


Figure S5.1.7: Per-residue RMSF of the peptide (left) Cα and (right) non-hydrogen atoms in the 5 × 150 ns equilibrium MD simulations, for each M^{pro}-peptide complex.

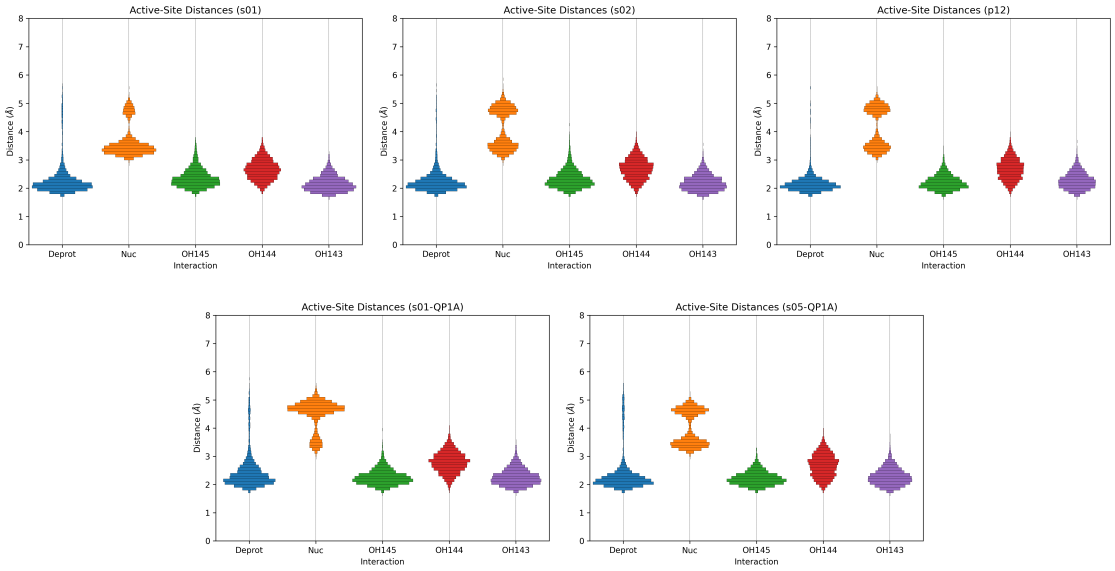


Figure S5.1.8: Distributions of distances in the active site that relate to the initiation of reaction between M^{pro} Cys145 and the scissile amide bond, observed over the combined 5 × 150 ns equilibrium MD simulations, for each M^{pro}-peptide complex. These include the His41-Nε2 – Cys145-Hγ distance (d_{Deprot}, blue); the Cys145-Sγ – peptide P1-C distance (d_{Nuc}, orange); and the distances from the peptide scissile amide carbonyl P1-O to the backbone amide H atoms of Cys145 (d_{OH145}, green), Ser144 (d_{OH144}, red), and Gly143 (d_{OH143}, purple), which together form the oxyanion hole (see **main text Figure 5.1**).

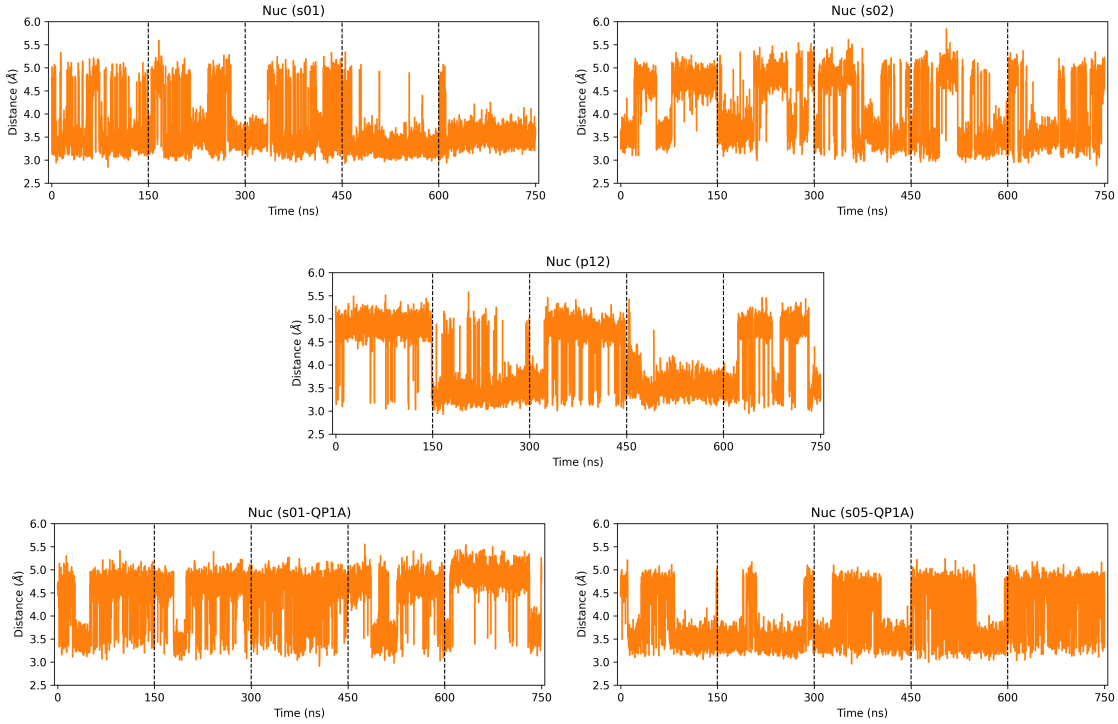


Figure S5.1.9: Time evolution of d_{Nuc} over the combined 5 × 150 ns equilibrium MD simulations.

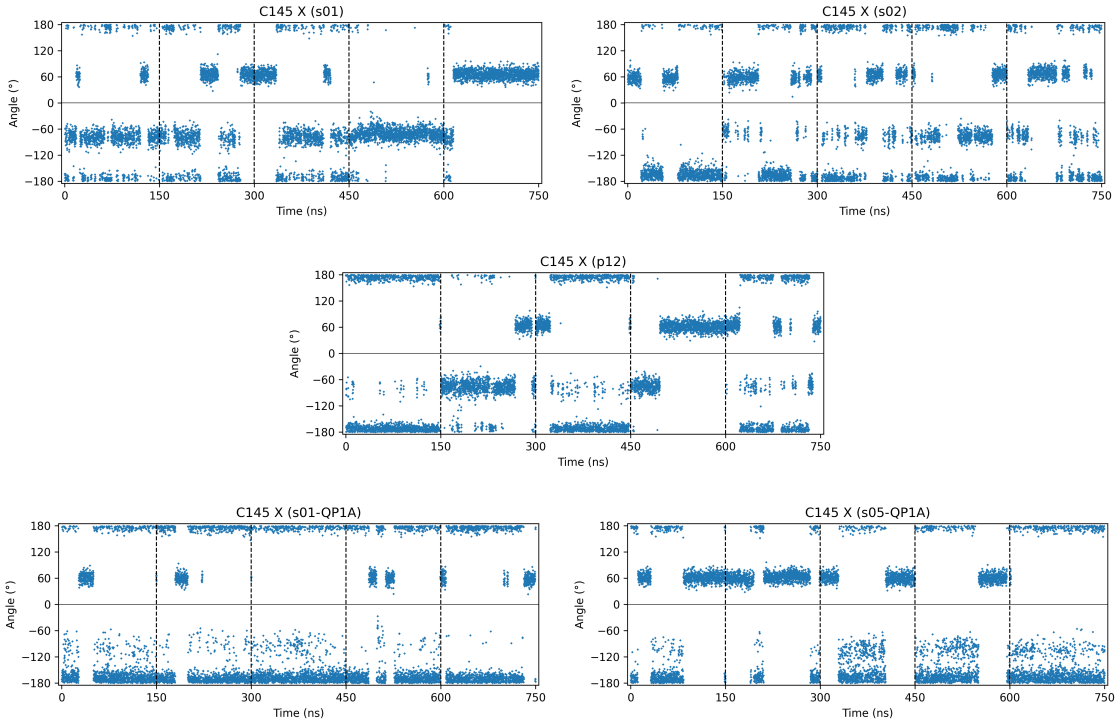


Figure S5.1.10: Time evolution of the Cys145 χ dihedral (N-Cα-Cβ-Sγ) over the combined 5 × 150 ns equilibrium MD simulations. By comparison to the time evolution of d_{Nuc} (**Figure S5.1.9**), d_{Nuc} is short (~3.5 Å) when Cys145 adopts a *gauche* conformation (χ centred around ±60°), and long (~5.0 Å) when Cys145 adopts a *trans* conformation (χ centred around ±180°).

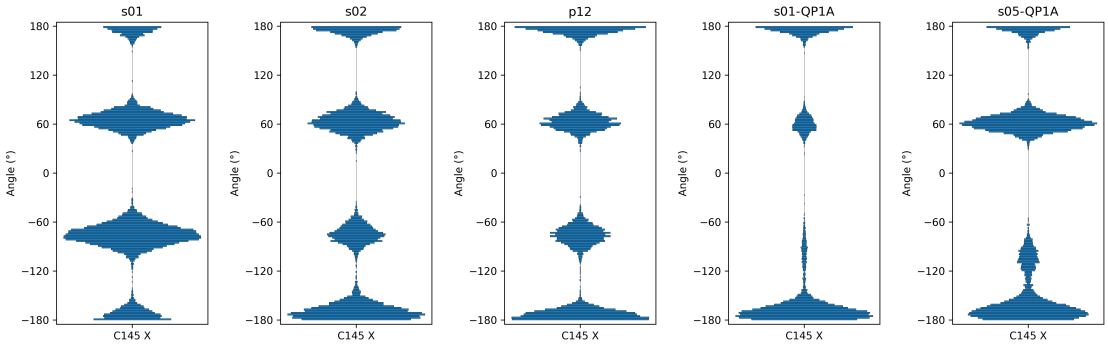


Figure S5.1.11: Distributions of the Cys145 χ dihedral (N-Cα-Cβ-Sγ) over the combined 5 × 150 ns equilibrium MD simulations. By defining the *gauche* conformer to be when -120° < χ < 120° and the *trans* conformer otherwise, the *gauche*:*trans* percentage population ratio observed when M^{pro} is complexed with s01, s02, p12, s01-QP1A, and s05-QP1A are 82:18, 53:47, 52:48, 19:81, 52:48, respectively. For comparison, the *gauche*:*trans* ratio in the previously reported equilibrium MD simulations of the M^{pro}-s05 complex[2] was 68:32.

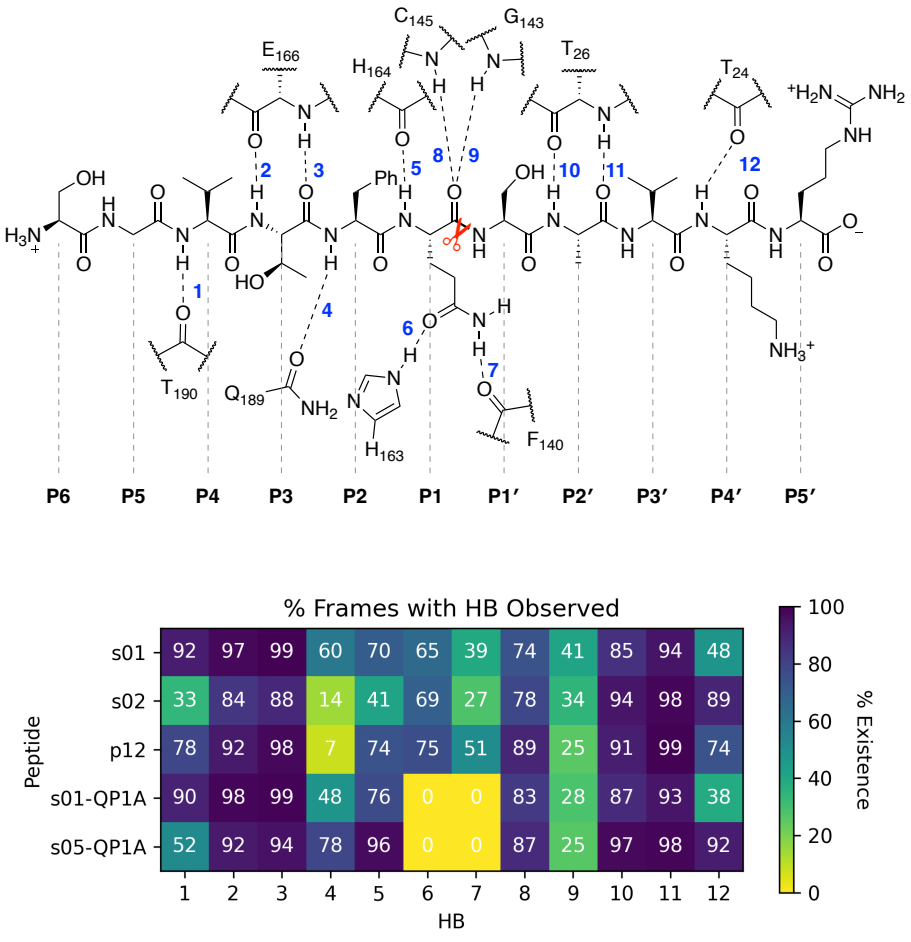


Figure S5.1.12: (top) Illustration of the 12 conserved M^{pro}-substrate hydrogen bonds (HBs)[1] exemplified by s02; (bottom) a heatmap showing the percentage of frames (analysed every ns) where the HB is observed, over the combined 5 × 150 ns equilibrium MD simulations, for each M^{pro}-peptide complex. Note that HBs 6 and 7 are dependent on the P1-Gln sidechain and thus are not formed in the cases of s01-QP1A and s05-QP1A.

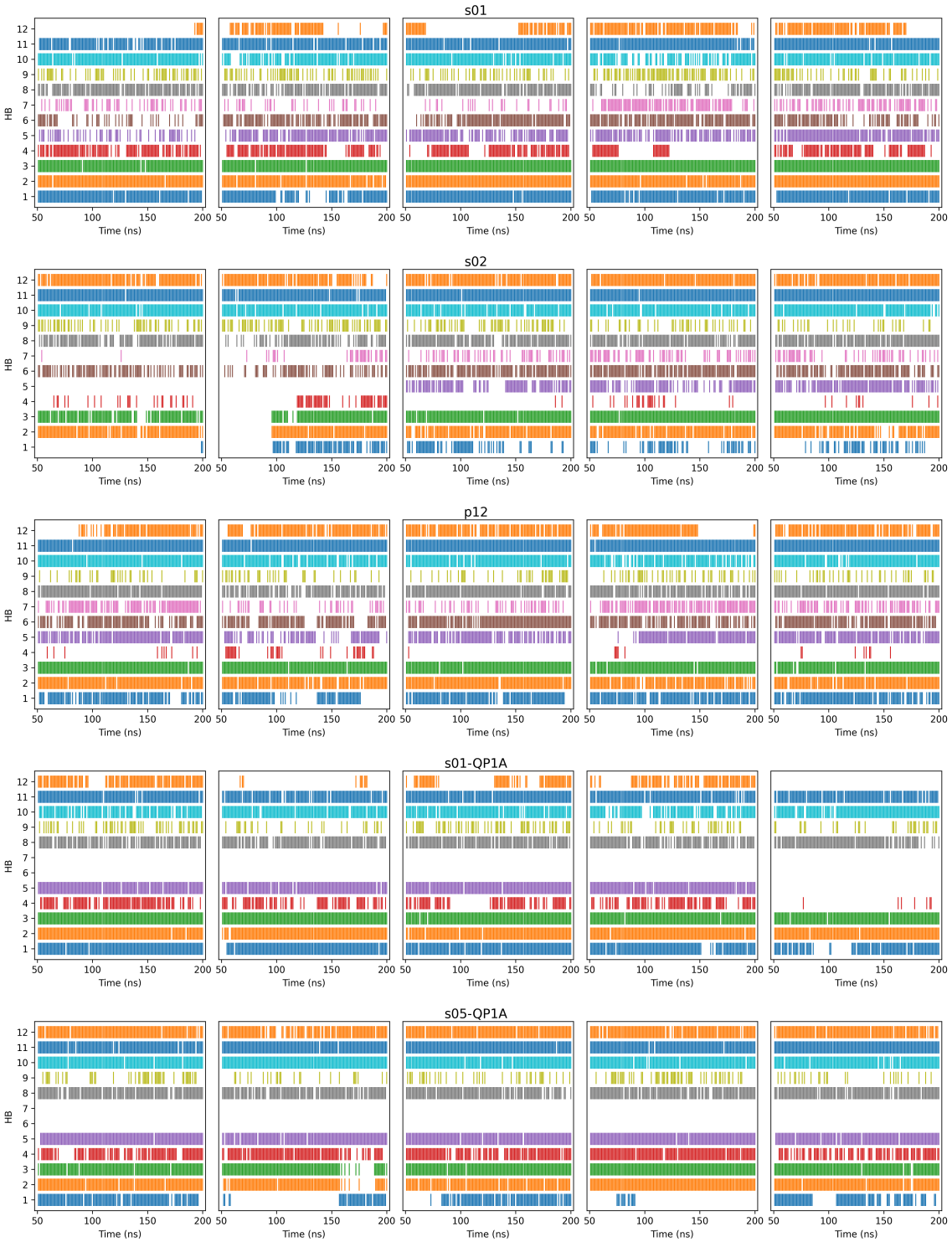


Figure S5.1.13: Event plots showing the existence of the 12 conserved HBs (frames analysed every ns) in each of the five 150 ns equilibrium MD simulations for each M^{pro}-peptide complex.

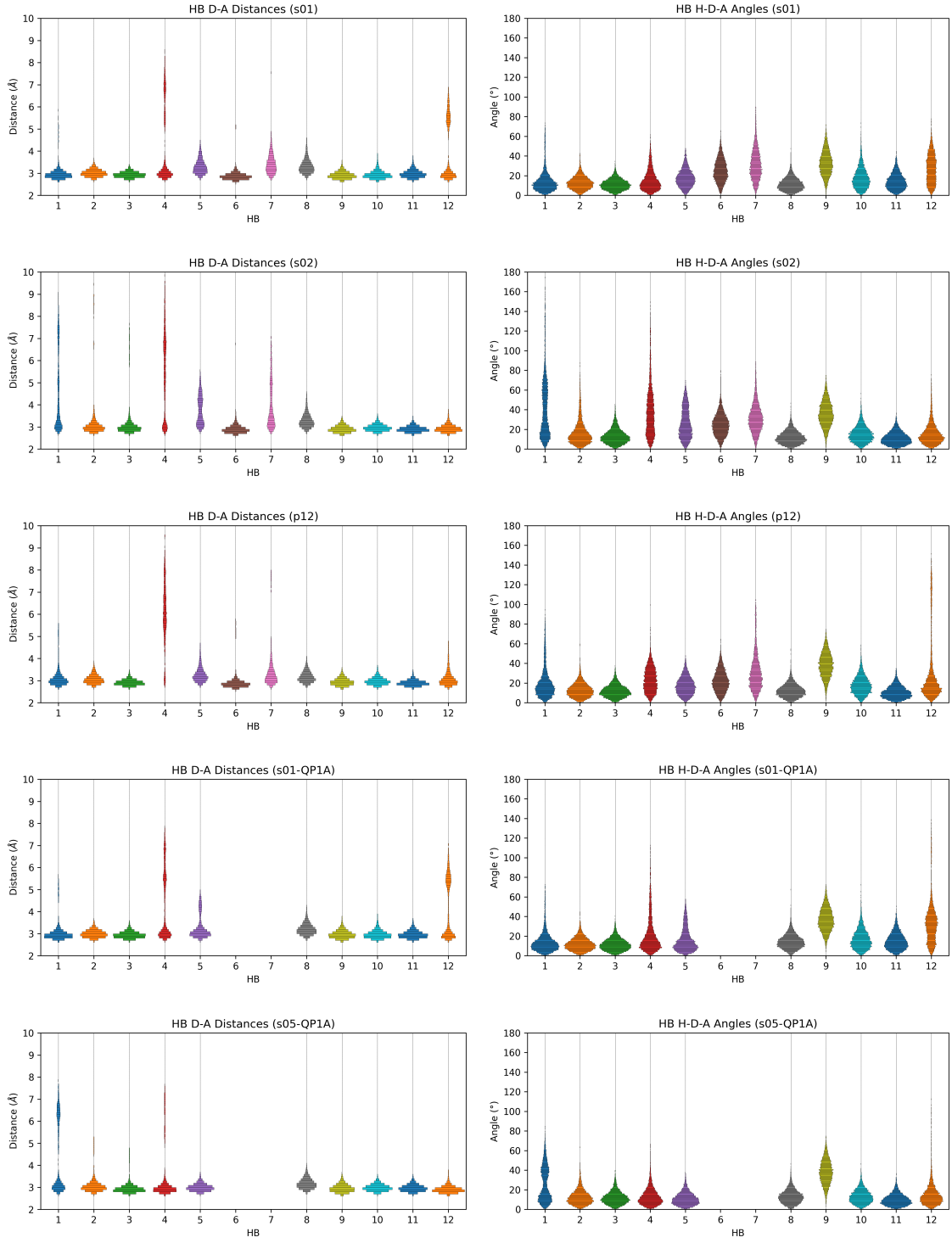


Figure S5.1.14: Distributions of (left) the donor-acceptor distances ($d_{\text{D-A}}$) and (right) the hydrogen-donor-acceptor angles ($\angle_{\text{H-D-A}}$) for each of the 12 conserved HBs, observed over the combined 5 × 150 ns equilibrium MD simulations, for each M^{pro}-peptide complex. HBs are defined based on a combined distance ($d_{\text{D-A}} \leq 3.5$ Å) and angle ($\angle_{\text{H-D-A}} \leq 30°$) criteria.

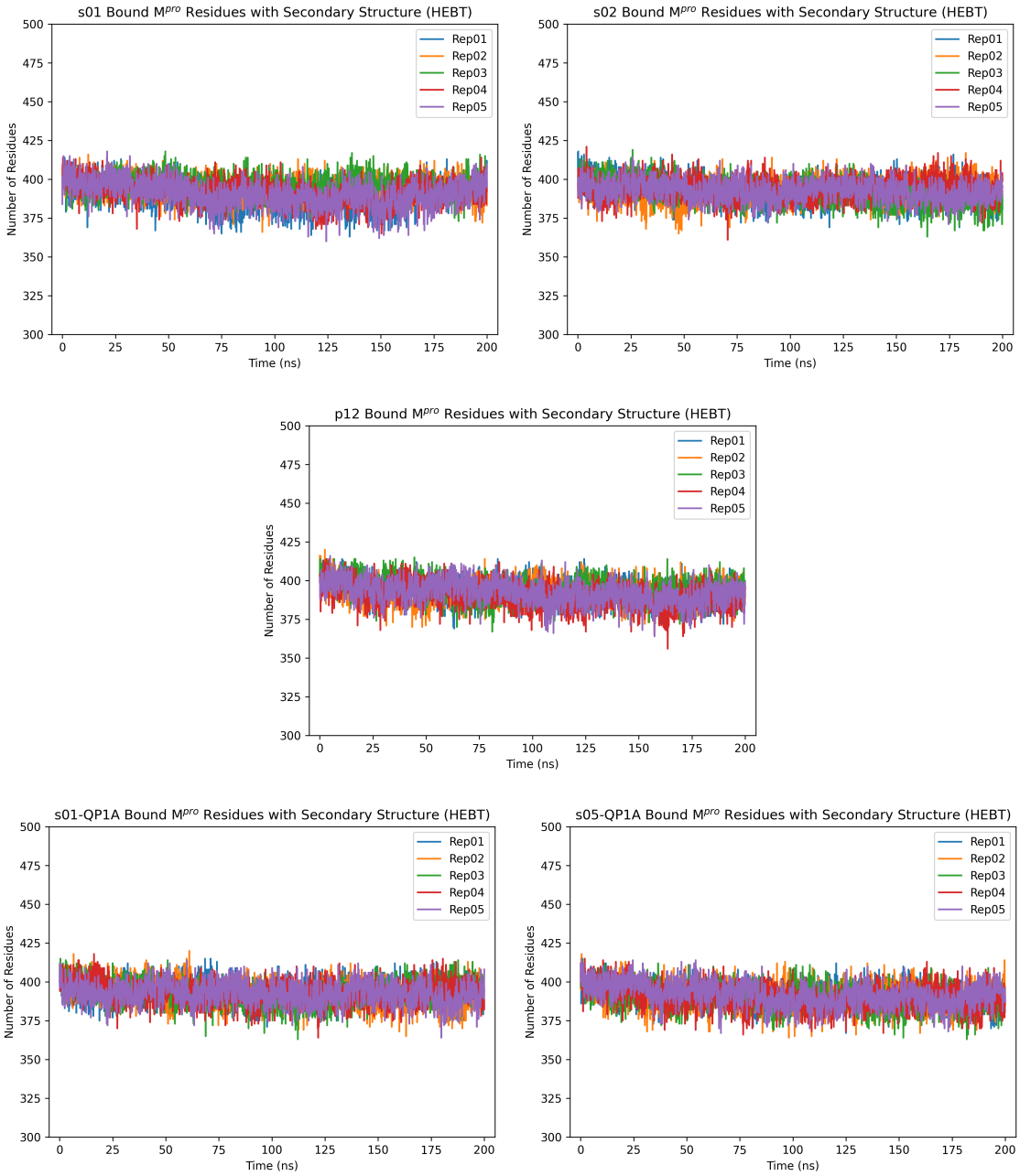


Figure S5.1.15: Time evolution of the number of M^{pro} residues possessing secondary structure (H = α-helix; E = extended strand; B = β-bridge; T = hydrogen bonded turn) in each of the five 200 ns production MD simulations for each M^{pro}-peptide complex, as defined by DSSP (version 2.0.4).[4, 5]

Section S5.2: Effect of P1 Gln-to-Ala mutation

Table S5.2.1: The average distance differences in the five active-site distances ($\Delta d = d_{neq} - d_{eqm}$) following the P1 mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

Active Site	s01: P1 Q->A				s05: P1 Q->A		
Deprot	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	-0.01	0.00		0.1	0.00	0.00
	0.25	-0.04	0.01		0.25	-0.03	0.01
	0.5	-0.01	0.01		0.5	-0.01	0.01
	1	-0.06	0.04		1	-0.05	0.04
	2	-0.10	0.04		2	-0.06	0.05
	3	-0.02	0.06		3	-0.14	0.06
	5	-0.08	0.05		5	-0.07	0.04
	10	-0.01	0.06		10	-0.06	0.06
	20	-0.05	0.04		20	-0.04	0.06
	30	-0.07	0.06		30	-0.02	0.05
	50	-0.08	0.05		50	-0.10	0.06
	100	-0.06	0.06		100	-0.17	0.07
	300	-0.10	0.08		300	-0.05	0.05
	500	-0.04	0.06		500	-0.05	0.06
	1000	-0.03	0.06		1000	-0.04	0.07
Nuc	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.02	0.00
	0.1	0.03	0.00		0.1	0.03	0.00
	0.25	-0.01	0.01		0.25	-0.04	0.01
	0.5	0.01	0.01		0.5	0.01	0.01
	1	0.00	0.02		1	-0.01	0.01
	2	0.04	0.03		2	-0.03	0.02
	3	0.00	0.03		3	-0.02	0.02
	5	0.02	0.03		5	-0.02	0.03
	10	0.04	0.04		10	0.07	0.04
	20	0.23	0.05		20	0.10	0.05
	30	0.22	0.05		30	0.05	0.04
	50	0.30	0.05		50	0.12	0.05
	100	0.37	0.06		100	0.19	0.05
	300	0.32	0.06		300	0.13	0.04
	500	0.44	0.06		500	0.12	0.05
	1000	0.52	0.06		1000	0.10	0.05
OH145	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.02	0.00		0.1	0.03	0.00
	0.25	-0.01	0.01		0.25	-0.02	0.01
	0.5	-0.02	0.02		0.5	0.01	0.01
	1	-0.02	0.02		1	-0.04	0.02
	2	-0.02	0.03		2	-0.02	0.02
	3	0.00	0.03		3	-0.03	0.02
	5	-0.03	0.03		5	-0.01	0.02
	10	-0.03	0.03		10	-0.02	0.03
	20	-0.04	0.03		20	-0.03	0.03
	30	-0.03	0.03		30	0.02	0.03
	50	-0.07	0.03		50	-0.01	0.02
	100	-0.08	0.04		100	0.00	0.03
	300	-0.03	0.03		300	-0.03	0.03
	500	-0.08	0.03		500	0.01	0.03
	1000	-0.03	0.03		1000	0.01	0.03
OH144	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	-0.01	0.00		0.05	0.00	0.00
	0.1	-0.03	0.00		0.1	-0.02	0.00
	0.25	-0.05	0.01		0.25	-0.05	0.01
	0.5	-0.04	0.02		0.5	-0.01	0.01
	1	-0.04	0.03		1	-0.08	0.03
	2	-0.08	0.03		2	-0.07	0.03
	3	-0.06	0.03		3	-0.07	0.03
	5	-0.02	0.04		5	-0.04	0.03
	10	0.05	0.04		10	-0.01	0.03
	20	-0.11	0.04		20	0.00	0.04
	30	0.02	0.04		30	0.00	0.03
	50	0.00	0.04		50	0.00	0.04
	100	0.04	0.04		100	-0.05	0.03
	300	0.08	0.04		300	-0.04	0.03
	500	0.06	0.04		500	-0.01	0.04
	1000	0.06	0.04		1000	-0.01	0.04
OH143	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.00	0.00
	0.1	-0.01	0.00		0.1	-0.03	0.00
	0.25	0.04	0.01		0.25	0.03	0.01
	0.5	0.05	0.02		0.5	0.03	0.01
	1	0.04	0.02		1	0.04	0.03
	2	0.06	0.03		2	0.04	0.03
	3	0.03	0.03		3	0.06	0.03
	5	0.00	0.03		5	-0.02	0.03
	10	0.05	0.03		10	0.04	0.03
	20	0.05	0.03		20	0.03	0.03
	30	0.01	0.03		30	-0.01	0.03
	50	0.04	0.03		50	0.04	0.03
	100	0.10	0.03		100	-0.04	0.03
	300	0.02	0.03		300	-0.02	0.03
	500	0.06	0.03		500	0.02	0.03
	1000	0.07	0.03		1000	-0.01	0.03

Table S5.2.2: The average distance differences in the donor-acceptor distances of HBs 1-5 following the P1 mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB1-5	s01: P1 Q->A				s05: P1 Q->A		
HB1	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.00	0.00		0.25	-0.01	0.00
	0.5	0.00	0.00		0.5	-0.01	0.00
	1	-0.02	0.01		1	0.00	0.02
	2	-0.01	0.02		2	-0.02	0.02
	3	0.02	0.02		3	-0.05	0.03
	5	0.01	0.02		5	-0.01	0.03
	10	-0.01	0.02		10	0.00	0.03
	20	0.06	0.03		20	0.04	0.03
	30	0.01	0.04		30	0.03	0.03
	50	0.01	0.03		50	-0.03	0.03
	100	-0.01	0.03		100	-0.05	0.04
	300	0.03	0.04		300	0.03	0.04
	500	-0.06	0.03		500	-0.07	0.04
	1000	-0.02	0.04		1000	-0.06	0.04
HB2	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	-0.01	0.00		0.1	-0.01	0.00
	0.25	0.01	0.00		0.25	0.00	0.00
	0.5	0.01	0.01		0.5	-0.01	0.01
	1	0.00	0.01		1	0.00	0.02
	2	-0.01	0.02		2	0.02	0.02
	3	0.00	0.01		3	-0.07	0.02
	5	-0.02	0.02		5	-0.03	0.02
	10	0.01	0.02		10	-0.03	0.02
	20	0.01	0.02		20	-0.01	0.02
	30	0.00	0.02		30	-0.03	0.02
	50	0.00	0.02		50	0.01	0.02
	100	0.02	0.02		100	0.01	0.02
	300	0.02	0.02		300	-0.01	0.03
	500	0.02	0.02		500	0.04	0.03
	1000	0.00	0.02		1000	0.05	0.03
HB3	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	-0.01	0.00		0.05	-0.01	0.00
	0.1	-0.01	0.00		0.1	-0.01	0.00
	0.25	0.00	0.00		0.25	0.00	0.00
	0.5	-0.01	0.01		0.5	-0.01	0.01
	1	-0.01	0.01		1	-0.01	0.01
	2	-0.04	0.01		2	-0.01	0.02
	3	-0.01	0.01		3	-0.03	0.02
	5	-0.04	0.02		5	0.01	0.02
	10	-0.02	0.01		10	-0.01	0.02
	20	-0.02	0.02		20	-0.02	0.02
	30	-0.02	0.01		30	-0.02	0.01
	50	0.00	0.01		50	-0.01	0.01
	100	0.01	0.01		100	0.01	0.01
	300	-0.01	0.01		300	-0.02	0.02
	500	0.01	0.01		500	0.03	0.02
	1000	-0.02	0.02		1000	0.00	0.02
HB4	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	-0.02	0.00		0.1	-0.02	0.00
	0.25	-0.01	0.00		0.25	0.00	0.00
	0.5	-0.02	0.01		0.5	0.00	0.01
	1	0.02	0.03		1	0.02	0.02
	2	0.01	0.03		2	0.00	0.03
	3	0.03	0.03		3	0.01	0.03
	5	-0.01	0.03		5	-0.06	0.04
	10	0.01	0.03		10	-0.04	0.04
	20	0.06	0.04		20	0.01	0.05
	30	0.00	0.06		30	-0.03	0.06
	50	-0.02	0.07		50	-0.03	0.08
	100	0.11	0.08		100	0.04	0.07
	300	-0.03	0.11		300	0.01	0.10
	500	-0.04	0.13		500	0.13	0.12
	1000	0.09	0.14		1000	0.26	0.16
HB5	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.01	0.00
	0.1	-0.02	0.00		0.1	-0.03	0.00
	0.25	-0.09	0.01		0.25	-0.11	0.01
	0.5	-0.08	0.01		0.5	-0.06	0.01
	1	-0.10	0.02		1	-0.08	0.02
	2	-0.11	0.03		2	-0.09	0.03
	3	-0.15	0.03		3	-0.04	0.03
	5	-0.12	0.03		5	-0.09	0.03
	10	-0.13	0.03		10	-0.13	0.02
	20	-0.18	0.04		20	-0.14	0.02
	30	-0.17	0.03		30	-0.15	0.03
	50	-0.18	0.03		50	-0.20	0.02
	100	-0.23	0.03		100	-0.16	0.02
	300	-0.25	0.03		300	-0.18	0.03
	500	-0.26	0.04		500	-0.13	0.03
	1000	-0.24	0.03		1000	-0.19	0.02

Table S5.2.3: The average distance differences in the donor-acceptor distances of HBs 8-12 following the P1 mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB8-12	s01: P1 Q->A				s05: P1 Q->A		
HB8	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.02	0.00		0.1	0.03	0.00
	0.25	-0.01	0.01		0.25	-0.02	0.01
	0.5	-0.02	0.02		0.5	0.01	0.01
	1	-0.02	0.02		1	-0.04	0.02
	2	-0.02	0.03		2	-0.03	0.02
	3	-0.01	0.03		3	-0.03	0.02
	5	-0.04	0.03		5	-0.01	0.02
	10	-0.04	0.03		10	-0.03	0.02
	20	-0.05	0.03		20	-0.03	0.03
	30	-0.04	0.03		30	0.01	0.03
	50	-0.08	0.03		50	-0.02	0.02
	100	-0.08	0.04		100	0.00	0.03
	300	-0.04	0.03		300	-0.03	0.03
	500	-0.10	0.03		500	0.00	0.03
	1000	-0.04	0.03		1000	0.00	0.03
HB9	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.00	0.00
	0.1	-0.02	0.00		0.1	-0.03	0.00
	0.25	0.02	0.01		0.25	0.01	0.01
	0.5	0.02	0.01		0.5	0.02	0.01
	1	0.02	0.02		1	0.01	0.02
	2	0.03	0.02		2	0.00	0.02
	3	0.01	0.02		3	0.03	0.02
	5	0.00	0.02		5	0.00	0.02
	10	0.04	0.02		10	0.03	0.02
	20	0.03	0.02		20	0.01	0.02
	30	0.00	0.02		30	-0.01	0.02
	50	0.02	0.02		50	0.03	0.02
	100	0.05	0.02		100	-0.01	0.02
	300	0.02	0.02		300	0.00	0.02
	500	0.03	0.02		500	0.03	0.02
	1000	0.03	0.02		1000	0.00	0.02
HB10	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.04	0.00		0.25	0.03	0.00
	0.5	0.03	0.01		0.5	0.01	0.01
	1	0.03	0.01		1	0.02	0.01
	2	0.02	0.02		2	0.01	0.02
	3	0.01	0.02		3	0.00	0.02
	5	-0.01	0.02		5	0.02	0.02
	10	0.00	0.02		10	0.02	0.02
	20	0.01	0.02		20	-0.01	0.02
	30	0.03	0.02		30	0.01	0.02
	50	0.08	0.02		50	0.02	0.02
	100	0.06	0.02		100	0.04	0.02
	300	0.07	0.02		300	0.03	0.02
	500	-0.01	0.02		500	0.05	0.02
	1000	0.04	0.02		1000	0.04	0.02
HB11	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.01	0.00		0.25	0.02	0.00
	0.5	0.00	0.01		0.5	0.00	0.00
	1	-0.02	0.01		1	0.03	0.01
	2	-0.01	0.02		2	0.00	0.02
	3	0.02	0.02		3	-0.02	0.02
	5	-0.01	0.02		5	0.01	0.02
	10	0.01	0.02		10	0.01	0.02
	20	0.01	0.02		20	-0.01	0.02
	30	0.01	0.02		30	0.00	0.02
	50	0.02	0.02		50	0.00	0.02
	100	0.01	0.02		100	0.04	0.02
	300	-0.03	0.02		300	0.00	0.02
	500	-0.02	0.02		500	0.03	0.02
	1000	0.02	0.02		1000	0.02	0.02
HB12	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.00	0.00		0.25	0.00	0.00
	0.5	0.00	0.00		0.5	0.00	0.00
	1	-0.01	0.02		1	0.00	0.02
	2	-0.01	0.03		2	-0.02	0.02
	3	0.03	0.03		3	0.03	0.02
	5	-0.01	0.03		5	-0.01	0.02
	10	-0.04	0.03		10	-0.01	0.03
	20	-0.01	0.03		20	-0.02	0.02
	30	-0.03	0.03		30	-0.02	0.03
	50	0.01	0.03		50	-0.04	0.03
	100	-0.02	0.03		100	0.07	0.03
	300	0.01	0.05		300	0.01	0.03
	500	0.01	0.04		500	-0.01	0.03
	1000	-0.09	0.05		1000	0.01	0.04

Table S5.2.4: The ten most perturbed Cα atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P1 mutation in s01 and s05. Only displacements ≥0.01 Å are shown.

s01 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	6ALA_CA	C	0.03	0.00
	2	140PHE_CA	A	0.01	0.00
	3	144SER_CA	A	0.01	0.00
	4	163HIE_CA	A	0.01	0.00
	5	165MET_CA	A	0.01	0.00
	6	166GLU_CA	A	0.01	0.00
	7	1SER_CA	B	0.01	0.00
	8	2GLY_CA	B	0.01	0.00
	9	5LEU_CA	C	0.01	0.00
	10	7SER_CA	C	0.01	0.00
0.5	1	6ALA_CA	C	0.17	0.02
	2	140PHE_CA	A	0.12	0.01
	3	163HIE_CA	A	0.12	0.01
	4	141LEU_CA	A	0.11	0.01
	5	162MET_CA	A	0.11	0.01
	6	7SER_CA	C	0.08	0.01
	7	142ASN_CA	A	0.06	0.01
	8	8GLY_CA	C	0.06	0.01
	9	139SER_CA	A	0.05	0.01
	10	161TYR_CA	A	0.05	0.00
5	1	6ALA_CA	C	0.31	0.03
	2	73VAL_CA	A	0.20	0.05
	3	72ASN_CA	A	0.18	0.06
	4	170GLY_CA	A	0.15	0.04
	5	196THR_CA	B	0.15	0.05
	6	24THR_CA	A	0.13	0.04
	7	74GLN_CA	A	0.13	0.04
	8	244GLN_CA	A	0.13	0.04
	9	245ASP_CA	A	0.13	0.04
	10	24THR_CA	B	0.13	0.04
50	1	6ALA_CA	C	0.43	0.03
	2	72ASN_CA	A	0.33	0.08
	3	47GLU_CA	A	0.30	0.07
	4	11LYS_CA	C	0.30	0.09
	5	8GLY_CA	C	0.29	0.04
	6	23GLY_CA	A	0.28	0.06
	7	72ASN_CA	B	0.25	0.08
	8	278GLY_CA	B	0.24	0.07
	9	24THR_CA	A	0.23	0.06
	10	48ASP_CA	A	0.23	0.05
500	1	6ALA_CA	C	0.38	0.04
	2	11LYS_CA	C	0.30	0.11
	3	47GLU_CA	A	0.28	0.09
	4	215GLY_CA	A	0.27	0.10
	5	1THR_CA	C	0.27	0.13
	6	53ASN_CA	A	0.25	0.08
	7	277ASN_CA	A	0.25	0.10
	8	54TYR_CA	A	0.24	0.06
	9	2SER_CA	C	0.23	0.08
	10	9PHE_CA	C	0.22	0.06
1000	1	6ALA_CA	C	0.47	0.04
	2	9PHE_CA	C	0.37	0.06
	3	10ARG_CA	C	0.37	0.08
	4	11LYS_CA	C	0.37	0.13
	5	71GLY_CA	A	0.35	0.08
	6	73VAL_CA	A	0.34	0.09
	7	24THR_CA	A	0.32	0.07
	8	56ASP_CA	A	0.32	0.09
	9	59ILE_CA	A	0.32	0.08
	10	277ASN_CA	B	0.31	0.10

s05 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	6ALA_CA	C	0.02	0.00
	2	144SER_CA	A	0.01	0.00
	3	164HIE_CA	A	0.01	0.00
	4	166GLU_CA	A	0.01	0.00
	5	1SER_CA	B	0.01	0.00
	6	5LEU_CA	C	0.01	0.00
	7				
	8				
	9				
	10				
0.5	1	6ALA_CA	C	0.20	0.01
	2	163HIE_CA	A	0.16	0.01
	3	162MET_CA	A	0.15	0.01
	4	140PHE_CA	A	0.08	0.01
	5	141LEU_CA	A	0.08	0.01
	6	172HIE_CA	A	0.08	0.00
	7	161TYR_CA	A	0.07	0.00
	8	165MET_CA	A	0.07	0.01
	9	166GLU_CA	A	0.07	0.01
	10	164HIE_CA	A	0.06	0.01
5	1	6ALA_CA	C	0.30	0.03
	2	222ARG_CA	B	0.21	0.05
	3	223PHE_CA	B	0.20	0.04
	4	274ASN_CA	B	0.19	0.05
	5	72ASN_CA	A	0.18	0.06
	6	275GLY_CA	B	0.16	0.04
	7	11SER_CA	C	0.16	0.08
	8	224THR_CA	B	0.15	0.04
	9	5LEU_CA	C	0.15	0.02
	10	163HIE_CA	A	0.14	0.02
50	1	6ALA_CA	C	0.40	0.03
	2	72ASN_CA	B	0.24	0.08
	3	47GLU_CA	A	0.21	0.06
	4	226THR_CA	A	0.19	0.06
	5	73VAL_CA	B	0.19	0.06
	6	46SER_CA	A	0.18	0.06
	7	225THR_CA	A	0.17	0.06
	8	278GLY_CA	A	0.17	0.09
	9	23GLY_CA	B	0.17	0.07
	10	93THR_CA	A	0.16	0.05
500	1	6ALA_CA	C	0.42	0.03
	2	1SER_CA	C	0.38	0.09
	3	191ALA_CA	B	0.33	0.09
	4	190THR_CA	B	0.29	0.08
	5	274ASN_CA	B	0.23	0.09
	6	275GLY_CA	B	0.22	0.08
	7	52PRO_CA	A	0.21	0.08
	8	141LEU_CA	A	0.19	0.04
	9	215GLY_CA	A	0.19	0.10
	10	51ASN_CA	A	0.18	0.07
1000	1	6ALA_CA	C	0.48	0.03
	2	72ASN_CA	A	0.41	0.11
	3	1SER_CA	C	0.39	0.10
	4	47GLU_CA	A	0.36	0.09
	5	71GLY_CA	A	0.34	0.09
	6	48ASP_CA	A	0.32	0.06
	7	2ALA_CA	C	0.31	0.06
	8	45THR_CA	A	0.30	0.06
	9	46SER_CA	A	0.30	0.08
	10	52PRO_CA	A	0.29	0.09

Table S5.2.5: The ten most perturbed non-hydrogen atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P1 mutation in s01 and s05.

s01 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	6ALA_CB	C	0.09	0.01
	2	140PHE_O	A	0.05	0.00
	3	163HIE_CE1	A	0.04	0.00
	4	166GLU_OE1	A	0.04	0.00
	5	163HIE_NE2	A	0.04	0.00
	6	144SER_OG	A	0.03	0.00
	7	163HIE_ND1	A	0.03	0.00
	8	166GLU_OE2	A	0.03	0.00
	9	166GLU_CB	A	0.03	0.00
	10	6ALA_CA	C	0.03	0.00
0.5	1	163HIE_CE1	A	0.46	0.02
	2	163HIE_NE2	A	0.39	0.02
	3	6ALA_CB	C	0.36	0.03
	4	163HIE_ND1	A	0.35	0.02
	5	140PHE_O	A	0.23	0.02
	6	163HIE_CD2	A	0.20	0.01
	7	6ALA_CA	C	0.17	0.02
	8	140PHE_C	A	0.17	0.01
	9	163HIE_N	A	0.15	0.01
	10	162MET_SD	A	0.15	0.01
5	1	6ALA_CB	C	0.59	0.04
	2	6ALA_CA	C	0.32	0.03
	3	73VAL_CG1	A	0.29	0.07
	4	72ASN_OD1	A	0.28	0.10
	5	142ASN_OD1	A	0.28	0.06
	6	196THR_CG2	B	0.27	0.08
	7	142ASN_ND2	A	0.27	0.06
	8	47GLU_OE2	B	0.26	0.09
	9	244GLN_OE1	A	0.26	0.08
	10	163HIE_CE1	A	0.26	0.03
50	1	6ALA_CB	C	0.80	0.04
	2	72ASN_OD1	A	0.65	0.15
	3	47GLU_OE2	A	0.65	0.16
	4	72ASN_ND2	A	0.55	0.14
	5	145CYS_SG	A	0.54	0.06
	6	72ASN_CG	A	0.53	0.12
	7	47GLU_CD	A	0.48	0.13
	8	72ASN_ND2	B	0.48	0.13
	9	142ASN_ND2	A	0.46	0.08
	10	47GLU_OE1	A	0.45	0.16
500	1	6ALA_CB	C	0.75	0.05
	2	145CYS_SG	A	0.70	0.07
	3	11LYS_NZ	C	0.66	0.29
	4	142ASN_OD1	B	0.53	0.15
	5	11LYS_CE	C	0.52	0.24
	6	306GLN_NE2	A	0.52	0.20
	7	11LYS_CD	C	0.49	0.21
	8	236LYS_NZ	B	0.49	0.25
	9	47GLU_OE2	A	0.48	0.23
	10	277ASN_OD1	A	0.47	0.20
1000	1	6ALA_CB	C	0.87	0.04
	2	145CYS_SG	A	0.76	0.07
	3	10ARG_NH2	C	0.55	0.23
	4	244GLN_OE1	A	0.54	0.23
	5	244GLN_NE2	A	0.54	0.25
	6	11LYS_OC2	C	0.49	0.16
	7	279ARG_NH1	A	0.48	0.21
	8	235MET_SD	B	0.48	0.18
	9	6ALA_CA	C	0.47	0.04
	10	10ARG_CZ	C	0.47	0.19

s05 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	6ALA_CB	C	0.12	0.01
	2	163HIE_CE1	A	0.04	0.00
	3	163HIE_NE2	A	0.04	0.00
	4	140PHE_O	A	0.04	0.00
	5	163HIE_ND1	A	0.04	0.00
	6	166GLU_OE1	A	0.03	0.00
	7	144SER_OG	A	0.03	0.00
	8	166GLU_CB	A	0.03	0.00
	9	165MET_O	A	0.03	0.00
	10	172HIE_ND1	A	0.02	0.00
0.5	1	163HIE_CE1	A	0.50	0.02
	2	6ALA_CB	C	0.45	0.02
	3	163HIE_NE2	A	0.41	0.02
	4	163HIE_ND1	A	0.40	0.02
	5	163HIE_N	A	0.21	0.01
	6	162MET_SD	A	0.20	0.01
	7	6ALA_CA	C	0.20	0.01
	8	163HIE_CD2	A	0.19	0.01
	9	140PHE_O	A	0.19	0.02
	10	163HIE_CB	A	0.19	0.01
5	1	6ALA_CB	C	0.60	0.03
	2	163HIE_CE1	A	0.39	0.03
	3	163HIE_ND1	A	0.33	0.03
	4	163HIE_NE2	A	0.31	0.03
	5	6ALA_CA	C	0.31	0.03
	6	142ASN_ND2	A	0.29	0.06
	7	72ASN_OD1	A	0.29	0.10
	8	72ASN_ND2	A	0.28	0.10
	9	72ASN_CG	A	0.28	0.09
	10	72ASN_CB	A	0.26	0.08
50	1	6ALA_CB	C	0.79	0.04
	2	236LYS_NZ	B	0.48	0.18
	3	72ASN_OD1	B	0.44	0.14
	4	74GLN_NE2	B	0.43	0.12
	5	236LYS_NZ	A	0.42	0.18
	6	236LYS_CE	B	0.41	0.15
	7	76ARG_NH1	A	0.40	0.15
	8	6ALA_CA	C	0.40	0.03
	9	47GLU_OE2	A	0.39	0.17
	10	76ARG_NH2	A	0.38	0.15
500	1	6ALA_CB	C	0.80	0.03
	2	236LYS_NZ	B	0.75	0.26
	3	236LYS_CE	B	0.67	0.22
	4	222ARG_NH1	B	0.67	0.22
	5	222ARG_NH2	B	0.59	0.28
	6	222ARG_CZ	B	0.54	0.22
	7	60ARG_NH1	B	0.51	0.18
	8	47GLU_OE1	A	0.50	0.22
	9	1SER_OG	C	0.49	0.14
	10	107GLN_OE1	B	0.46	0.15
1000	1	67LEU_CD1	A	0.86	0.13
	2	6ALA_CB	C	0.86	0.04
	3	222ARG_NH1	A	0.82	0.29
	4	222ARG_NH2	A	0.78	0.31
	5	72ASN_OD1	A	0.73	0.21
	6	107GLN_OE1	B	0.72	0.15
	7	47GLU_OE2	A	0.70	0.19
	8	222ARG_CZ	A	0.69	0.26
	9	47GLU_CG	A	0.65	0.14
	10	142ASN_ND2	A	0.65	0.10

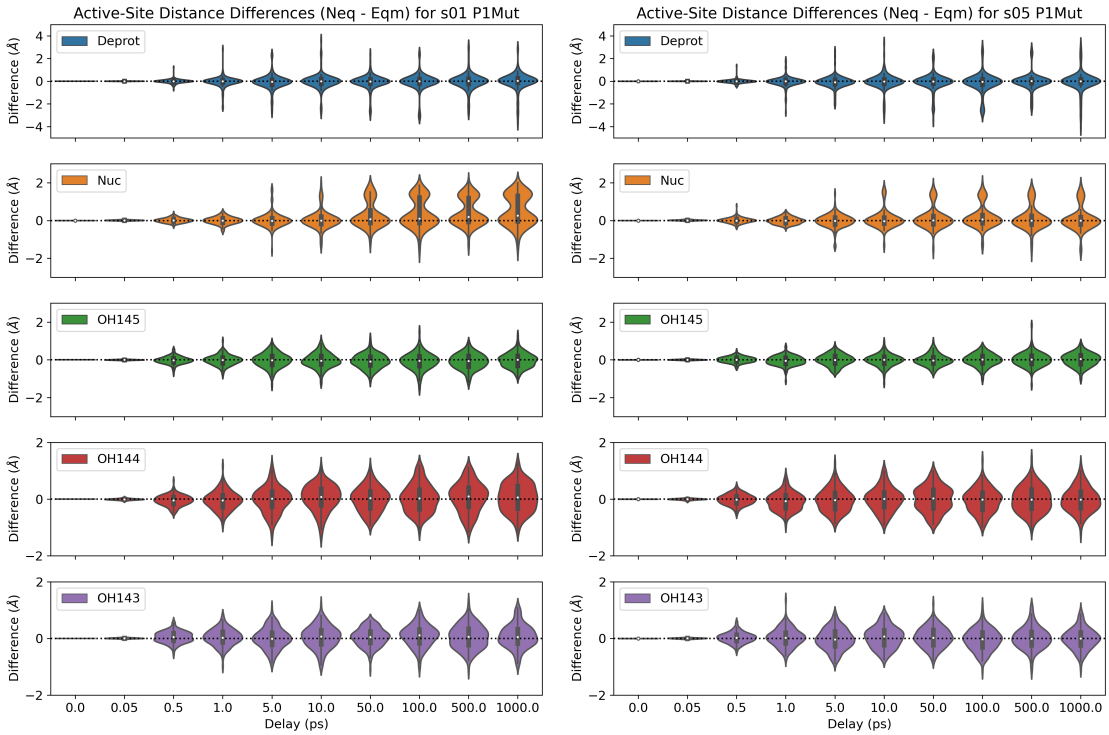


Figure S5.2.1: Distributions of the distance differences in the five active-site distances ($\Delta d = d_{\mathrm{neq}} - d_{\mathrm{eqm}}$) following the P1 mutation in s01 and s05.

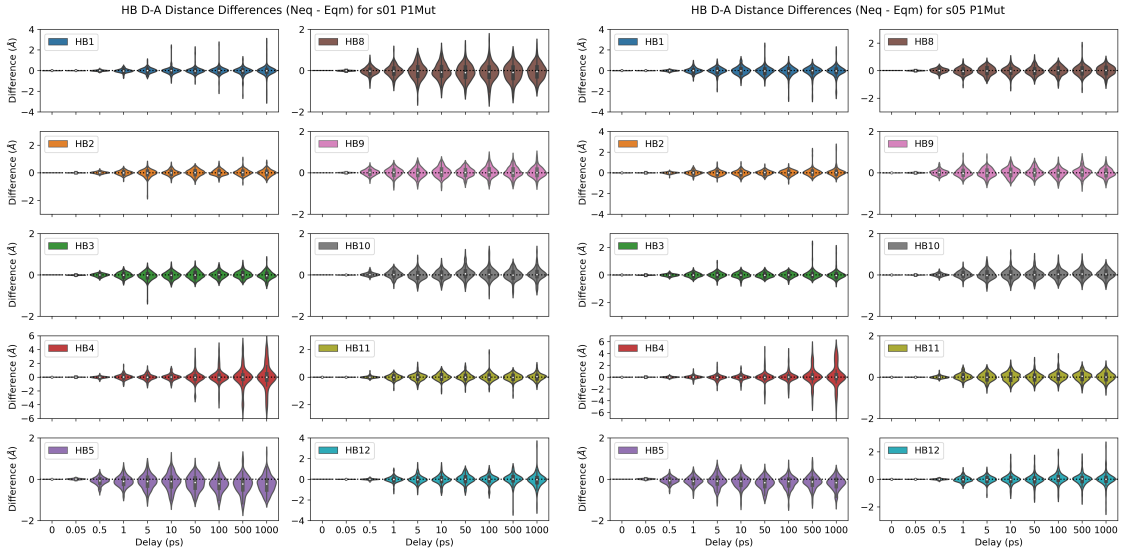


Figure S5.2.2: Distributions of the distance differences in the ten donor-acceptor distances of HBs 1-5 and 8-12 following the P1 mutation in s01 and s05.

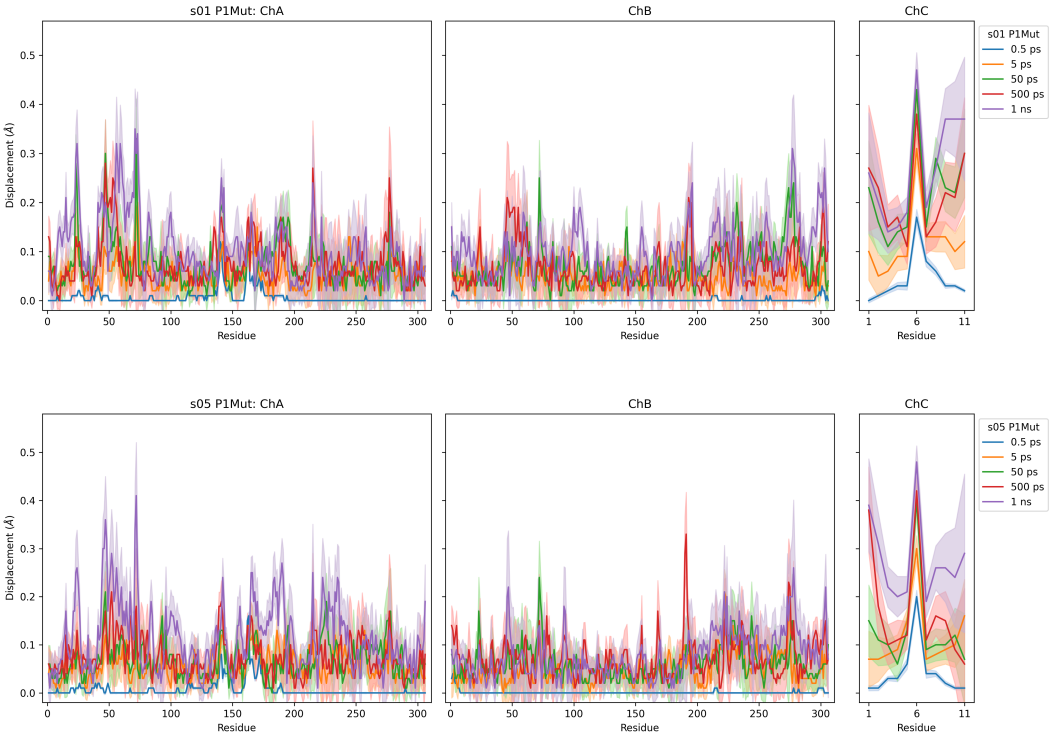


Figure S5.2.3: Magnitudes of the average M^{pro} Cα displacement vectors, with ± standard error of the mean (SEM) shown as shaded regions, following the P1 mutation in s01 and s05.

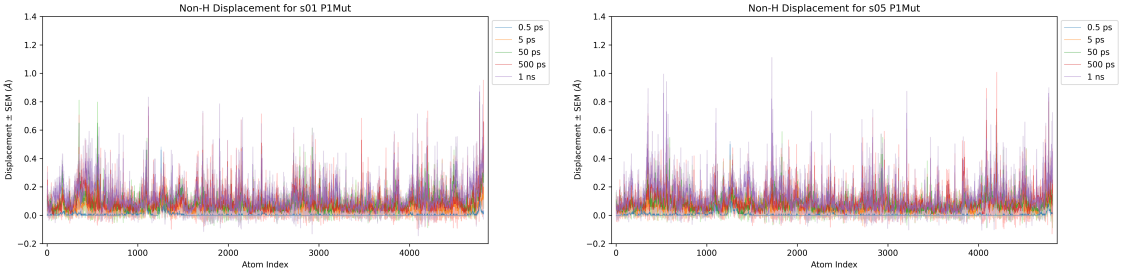


Figure S5.2.4: Magnitudes of the average non-hydrogen atom displacement vectors in the M^{pro}-peptide complex, with ± SEM shown as shaded regions, following the P1 mutation in s01 and s05.

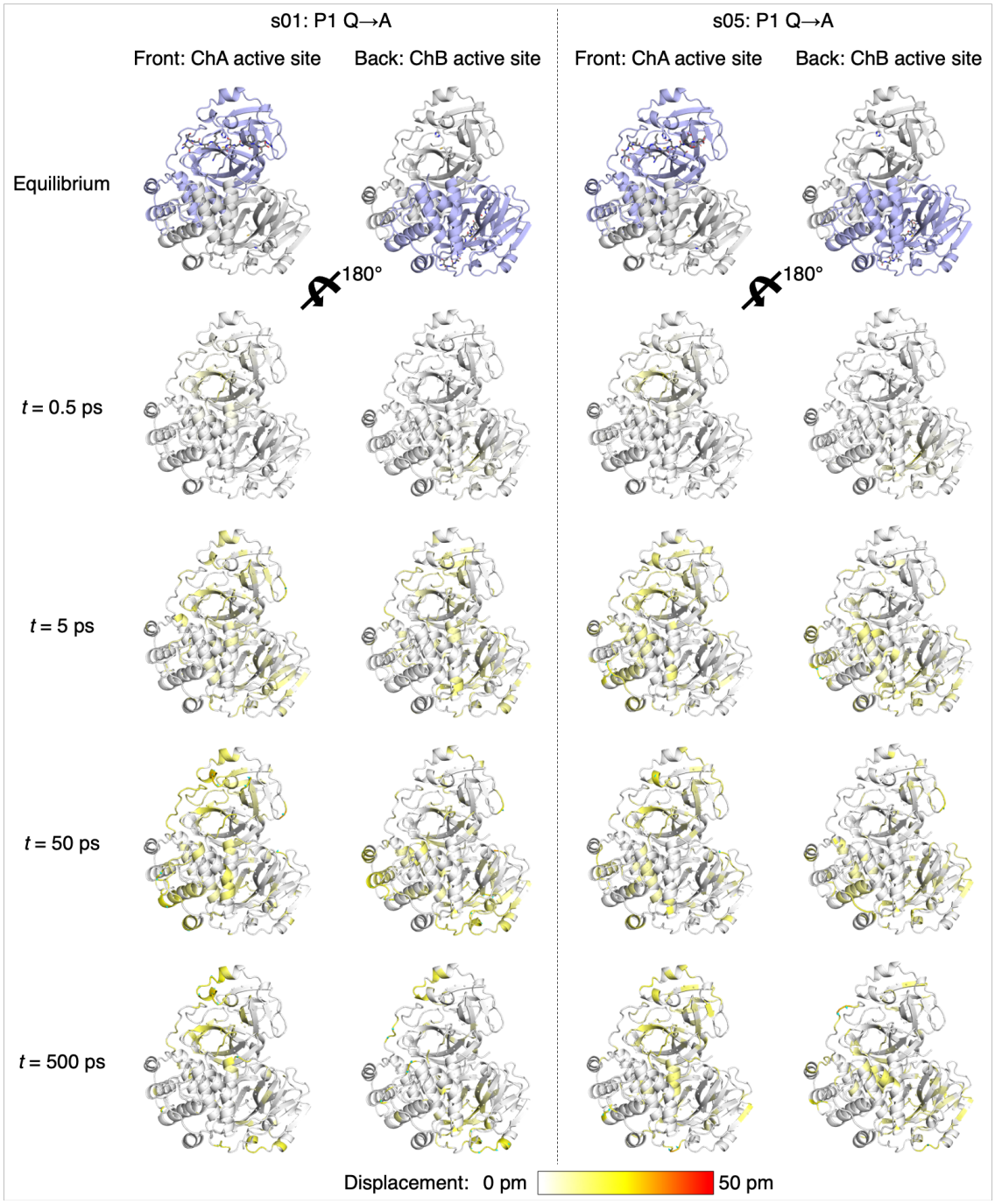


Figure S5.2.5: Views of the M^{pro} response to the P1 mutation in s01 and s05 from averaging Cα displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as cyan arrows with a scale-up factor of 5.[7]

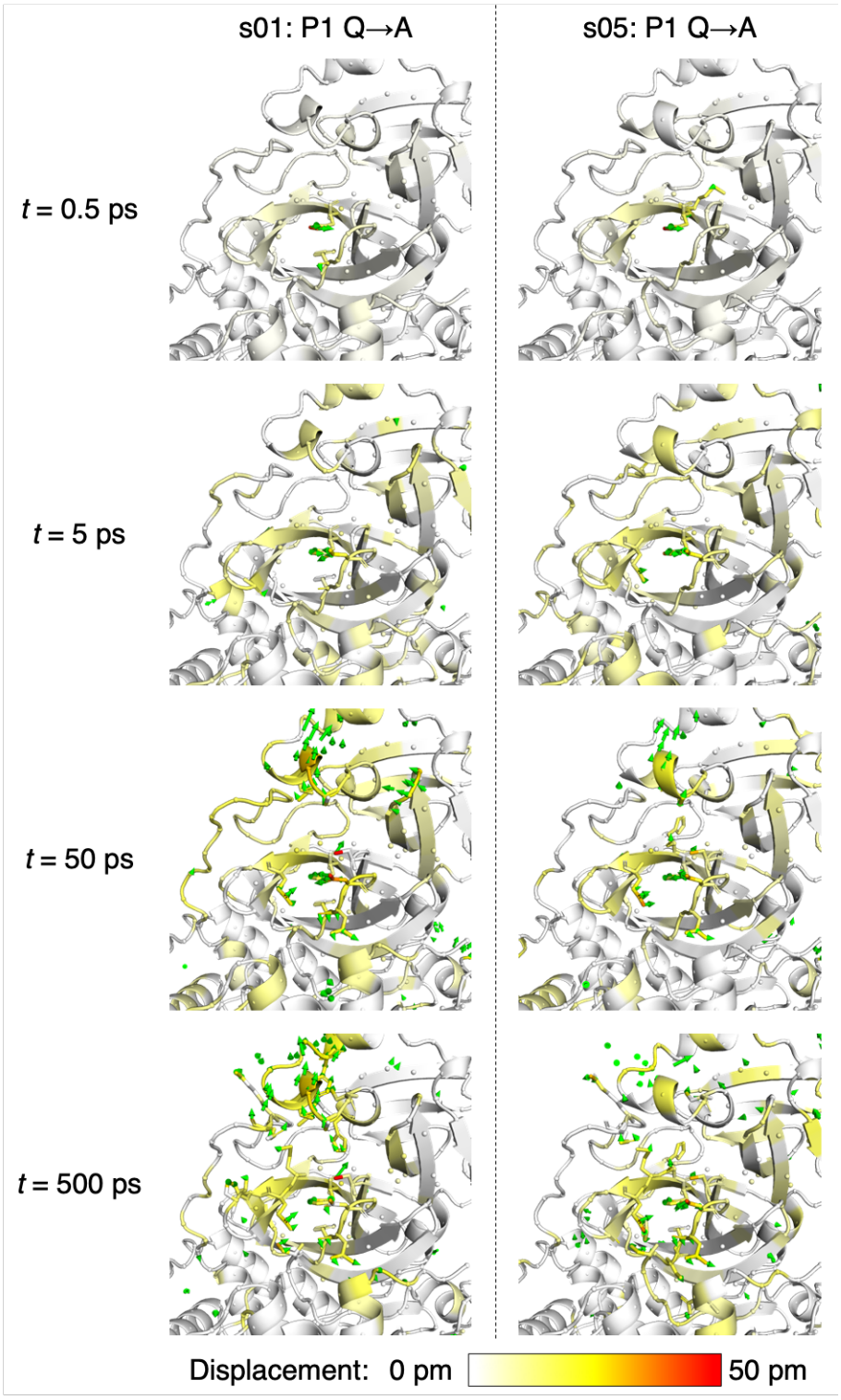


Figure S5.2.6: Views of the M^{pro} response to the P1 mutation in s01 and s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations, with a focus on the S1 subsite. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7] Residues that show such significant displacements and that are within 10 Å of the substrate P1 residue are shown as sticks.

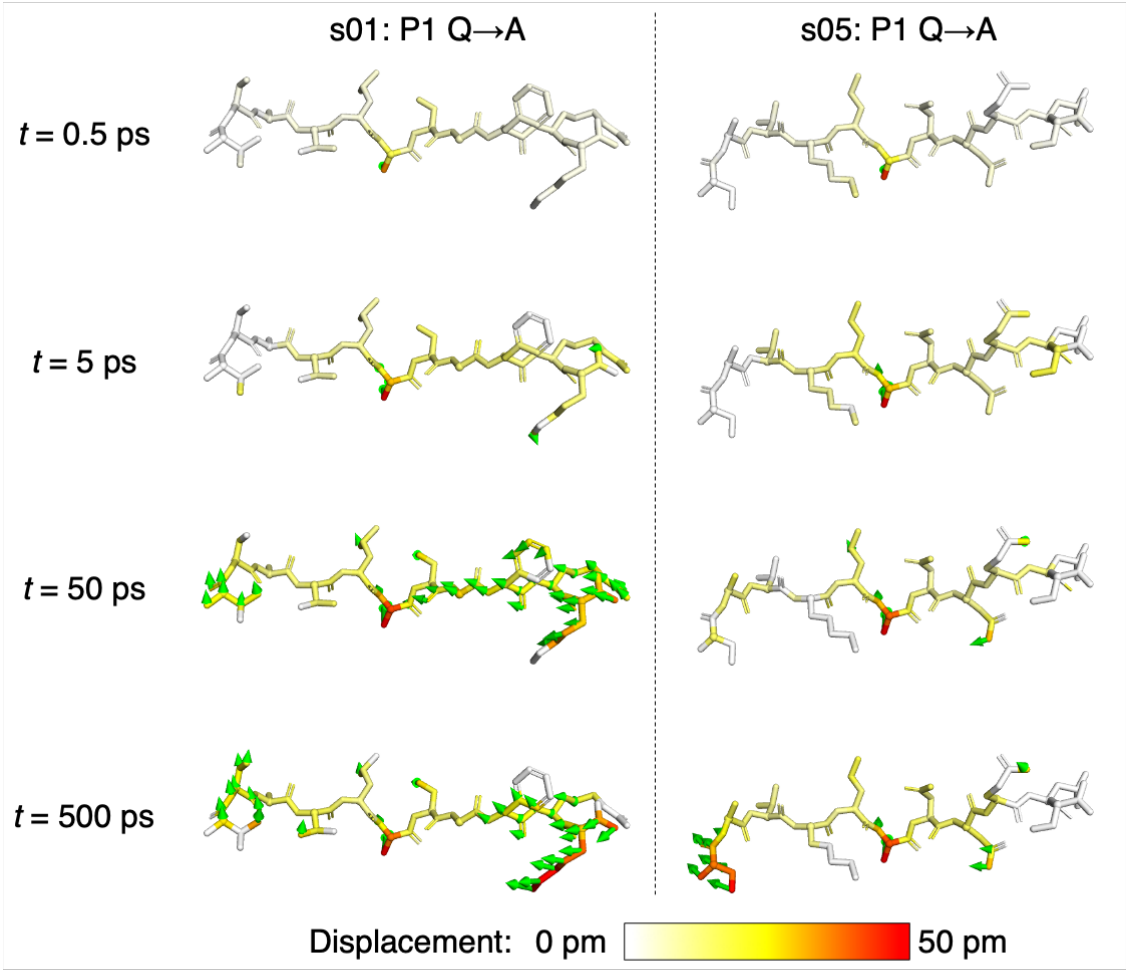


Figure S5.2.7: Views of the peptide response to the P1 mutation in s01 and s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7]

Section S5.3: Effect of P2 mutation

Table S5.3.1: The average distance differences in the five active-site distances ($\Delta d = d_{neq} - d_{eqm}$) following the P2 mutation in s01, s05, s02, and p12. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

Active	s01: P2 L->A			s05: P2 L->A			s02: P2 F->A			p12: P2 W->A		
Deprot	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.01	0.00
	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.00	0.00	0.1	0.01	0.00
	0.25	0.02	0.01	0.25	0.01	0.01	0.25	0.01	0.01	0.25	0.03	0.01
	0.5	0.02	0.01	0.5	0.00	0.01	0.5	-0.01	0.01	0.5	-0.01	0.02
	1	-0.04	0.03	1	-0.02	0.03	1	-0.02	0.04	1	-0.01	0.02
	2	-0.03	0.04	2	-0.06	0.05	2	0.00	0.04	2	-0.06	0.03
	3	0.06	0.04	3	-0.07	0.06	3	0.01	0.05	3	-0.04	0.03
	5	-0.03	0.06	5	-0.11	0.05	5	0.00	0.05	5	-0.03	0.03
	10	0.06	0.06	10	-0.03	0.06	10	-0.04	0.04	10	-0.07	0.04
	20	0.06	0.05	20	-0.04	0.06	20	-0.04	0.04	20	-0.05	0.03
	30	0.02	0.07	30	0.00	0.07	30	0.01	0.05	30	-0.03	0.04
	50	0.06	0.07	50	-0.06	0.07	50	0.07	0.05	50	-0.02	0.04
	100	0.07	0.07	100	0.04	0.08	100	0.02	0.05	100	-0.05	0.03
	300	-0.04	0.08	300	0.03	0.07	300	-0.08	0.04	300	0.00	0.04
	500	0.03	0.06	500	0.10	0.06	500	0.12	0.06	500	-0.02	0.05
	100	0.01	0.08	100	0.05	0.06	100	0.15	0.06	100	-0.02	0.04
Nuc	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00
	0.25	0.00	0.00	0.25	0.01	0.00	0.25	0.00	0.00	0.25	0.01	0.00
	0.5	0.00	0.01	0.5	0.01	0.01	0.5	-0.02	0.01	0.5	-0.01	0.01
	1	0.01	0.01	1	0.00	0.01	1	-0.04	0.01	1	-0.02	0.01
	2	0.00	0.02	2	-0.02	0.02	2	-0.03	0.02	2	0.00	0.02
	3	-0.05	0.02	3	0.02	0.02	3	-0.04	0.02	3	-0.02	0.02
	5	0.00	0.03	5	-0.04	0.02	5	0.00	0.03	5	-0.05	0.02
	10	-0.02	0.03	10	0.00	0.03	10	-0.03	0.02	10	-0.01	0.02
	20	0.03	0.03	20	0.00	0.03	20	-0.04	0.02	20	-0.02	0.03
	30	-0.02	0.03	30	0.03	0.03	30	-0.04	0.03	30	-0.01	0.03
	50	-0.03	0.04	50	0.02	0.04	50	-0.05	0.03	50	-0.03	0.03
	100	-0.04	0.04	100	0.07	0.04	100	-0.04	0.04	100	-0.05	0.04
	300	0.03	0.05	300	0.09	0.05	300	-0.02	0.05	300	-0.03	0.05
	500	0.12	0.05	500	0.05	0.05	500	-0.04	0.06	500	-0.05	0.06
	100	0.15	0.05	100	0.00	0.05	100	-0.12	0.06	100	-0.04	0.06
OH145	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.01	0.00	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.01	0.00
	0.25	0.01	0.01	0.25	-0.01	0.00	0.25	0.04	0.01	0.25	0.02	0.01
	0.5	0.02	0.01	0.5	-0.01	0.01	0.5	0.04	0.01	0.5	0.00	0.01
	1	0.06	0.02	1	-0.02	0.02	1	0.05	0.02	1	0.03	0.02
	2	0.04	0.03	2	-0.02	0.02	2	0.09	0.03	2	0.05	0.02
	3	0.09	0.03	3	-0.04	0.02	3	-0.01	0.03	3	0.07	0.02
	5	0.10	0.03	5	-0.04	0.02	5	0.03	0.03	5	0.06	0.02
	10	0.06	0.03	10	0.00	0.02	10	0.05	0.03	10	0.05	0.02
	20	0.03	0.03	20	0.01	0.03	20	0.02	0.04	20	0.03	0.02
	30	0.05	0.03	30	-0.02	0.03	30	0.02	0.03	30	0.03	0.03
	50	0.00	0.03	50	0.01	0.03	50	0.01	0.03	50	0.04	0.03
	100	0.05	0.04	100	-0.06	0.03	100	-0.02	0.03	100	0.05	0.03
	300	0.06	0.04	300	0.00	0.03	300	0.00	0.04	300	0.01	0.03
	500	0.00	0.04	500	0.01	0.03	500	0.08	0.04	500	0.08	0.03
	100	0.04	0.03	100	0.02	0.03	100	0.01	0.04	100	0.07	0.03
OH144	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.01	0.00	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.02	0.00
	0.25	0.01	0.00	0.25	-0.01	0.00	0.25	0.02	0.01	0.25	-0.01	0.00
	0.5	0.02	0.01	0.5	0.00	0.01	0.5	0.06	0.01	0.5	0.02	0.01
	1	0.04	0.02	1	-0.02	0.02	1	0.06	0.02	1	0.03	0.02
	2	0.03	0.03	2	-0.03	0.03	2	0.10	0.04	2	0.04	0.03
	3	0.02	0.04	3	-0.03	0.03	3	-0.01	0.03	3	0.05	0.04
	5	0.08	0.03	5	-0.02	0.03	5	0.03	0.03	5	0.04	0.03
	10	0.13	0.03	10	0.03	0.03	10	0.03	0.04	10	0.03	0.03
	20	-0.03	0.04	20	-0.01	0.03	20	0.00	0.04	20	0.03	0.03
	30	0.06	0.04	30	0.02	0.03	30	0.03	0.04	30	0.01	0.04
	50	0.02	0.04	50	0.01	0.04	50	-0.03	0.04	50	0.01	0.04
	100	0.03	0.04	100	-0.05	0.03	100	-0.02	0.04	100	0.03	0.03
	300	0.08	0.04	300	-0.01	0.04	300	0.06	0.04	300	0.04	0.04
	500	-0.03	0.04	500	-0.03	0.04	500	0.04	0.04	500	0.08	0.04
	100	0.03	0.04	100	-0.07	0.04	100	0.00	0.04	100	-0.03	0.04
OH143	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.01	0.00	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.01	0.00
	0.25	0.00	0.00	0.25	-0.01	0.00	0.25	0.00	0.00	0.25	-0.02	0.00
	0.5	0.02	0.01	0.5	0.00	0.01	0.5	0.00	0.01	0.5	0.01	0.01
	1	-0.01	0.02	1	0.00	0.02	1	0.02	0.02	1	-0.03	0.02
	2	0.01	0.03	2	0.00	0.03	2	-0.05	0.03	2	-0.05	0.03
	3	-0.01	0.03	3	0.01	0.03	3	-0.02	0.03	3	-0.05	0.03
	5	-0.02	0.03	5	0.00	0.03	5	0.00	0.03	5	-0.05	0.03
	10	-0.01	0.03	10	-0.01	0.03	10	0.01	0.03	10	-0.08	0.03
	20	0.02	0.03	20	-0.05	0.03	20	0.00	0.03	20	0.02	0.03
	30	0.01	0.03	30	0.02	0.03	30	0.05	0.03	30	0.00	0.03
	50	0.02	0.03	50	-0.02	0.03	50	0.02	0.03	50	-0.04	0.03
	100	0.03	0.03	100	0.03	0.03	100	0.04	0.03	100	-0.06	0.03
	300	-0.03	0.03	300	-0.02	0.03	300	0.04	0.03	300	0.02	0.03
	500	-0.02	0.03	500	-0.01	0.03	500	-0.02	0.03	500	-0.04	0.03
	100	0.03	0.03	100	-0.01	0.03	100	0.02	0.03	100	-0.05	0.03

Table S5.3.2: The average distance differences in the donor-acceptor distances of HBs 1-5 following the P2 mutation in s01, s05, s02, and p12. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB1-	s01: P2 L->A			s05: P2 L->A			s02: P2 F->A			p12: P2 W->A		
HB1	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	-0.01	0.00	0.1	0.00	0.00
	0.25	0.00	0.00	0.25	-0.01	0.00	0.25	-0.03	0.00	0.25	-0.01	0.00
	0.5	-0.01	0.00	0.5	0.00	0.01	0.5	-0.06	0.01	0.5	-0.02	0.01
	1	-0.04	0.01	1	0.00	0.02	1	-0.04	0.02	1	0.00	0.02
	2	0.00	0.02	2	-0.03	0.02	2	-0.06	0.04	2	-0.04	0.02
	3	0.03	0.02	3	0.00	0.03	3	-0.07	0.04	3	-0.05	0.02
	5	0.01	0.02	5	0.01	0.03	5	-0.02	0.04	5	-0.09	0.02
	10	0.00	0.02	10	0.03	0.03	10	-0.08	0.04	10	-0.05	0.03
	20	0.02	0.02	20	0.00	0.03	20	-0.02	0.04	20	-0.05	0.03
	30	-0.04	0.03	30	0.02	0.03	30	-0.06	0.05	30	-0.07	0.03
	50	-0.01	0.03	50	-0.01	0.03	50	-0.04	0.05	50	-0.05	0.03
	100	-0.01	0.03	100	0.01	0.03	100	-0.16	0.07	100	-0.04	0.04
	300	-0.01	0.04	300	0.01	0.04	300	-0.16	0.09	300	0.00	0.04
	500	-0.04	0.03	500	0.02	0.04	500	-0.29	0.09	500	-0.02	0.04
	1000	-0.08	0.04	1000	0.04	0.05	1000	-0.40	0.10	1000	0.01	0.04
HB2	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.00	0.00
	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.01	0.00	0.25	0.00	0.01
	0.5	0.01	0.01	0.5	-0.01	0.01	0.5	0.00	0.01	0.5	-0.01	0.01
	1	0.00	0.01	1	0.00	0.01	1	0.01	0.01	1	0.00	0.02
	2	-0.01	0.02	2	-0.01	0.02	2	0.01	0.02	2	0.00	0.02
	3	0.00	0.02	3	-0.01	0.02	3	0.04	0.02	3	-0.04	0.02
	5	-0.03	0.02	5	-0.02	0.02	5	0.03	0.02	5	0.03	0.02
	10	0.00	0.02	10	-0.01	0.02	10	0.03	0.03	10	-0.04	0.02
	20	0.02	0.02	20	-0.02	0.02	20	0.00	0.02	20	-0.02	0.02
	30	0.00	0.02	30	0.00	0.02	30	-0.01	0.03	30	-0.01	0.02
	50	0.00	0.02	50	0.04	0.02	50	0.01	0.02	50	0.02	0.02
	100	0.01	0.02	100	-0.01	0.02	100	0.00	0.04	100	0.04	0.03
	300	0.00	0.02	300	0.01	0.02	300	0.02	0.03	300	0.02	0.03
	500	0.02	0.02	500	0.03	0.02	500	-0.01	0.04	500	0.10	0.03
	1000	0.00	0.02	1000	0.01	0.02	1000	0.06	0.06	1000	0.04	0.03
HB3	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.01	0.00	0.05	0.01	0.00	0.05	0.01	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.02	0.00	0.1	0.02	0.00
	0.25	0.00	0.01	0.25	-0.01	0.00	0.25	0.01	0.01	0.25	0.02	0.01
	0.5	0.00	0.01	0.5	-0.01	0.01	0.5	0.02	0.01	0.5	0.00	0.01
	1	-0.01	0.01	1	-0.02	0.01	1	0.04	0.01	1	0.00	0.01
	2	0.00	0.01	2	-0.02	0.01	2	0.00	0.02	2	0.01	0.01
	3	0.00	0.01	3	-0.01	0.01	3	0.02	0.02	3	0.01	0.01
	5	-0.03	0.01	5	-0.03	0.02	5	0.02	0.02	5	0.05	0.02
	10	0.01	0.01	10	0.00	0.01	10	0.03	0.02	10	0.00	0.01
	20	-0.03	0.02	20	-0.04	0.01	20	-0.01	0.02	20	0.01	0.02
	30	0.00	0.01	30	-0.01	0.01	30	0.01	0.02	30	0.00	0.01
	50	0.02	0.01	50	0.00	0.01	50	0.02	0.02	50	0.02	0.02
	100	0.03	0.01	100	0.00	0.02	100	0.01	0.02	100	0.02	0.02
	300	0.01	0.01	300	0.01	0.02	300	0.01	0.03	300	0.03	0.02
	500	0.00	0.01	500	0.01	0.02	500	0.01	0.03	500	0.06	0.02
	1000	0.01	0.02	1000	-0.01	0.02	1000	0.06	0.04	1000	0.02	0.02
HB4	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	-0.01	0.00	0.05	-0.02	0.00	0.05	-0.02	0.00
	0.1	0.01	0.00	0.1	0.01	0.00	0.1	-0.03	0.01	0.1	-0.03	0.01
	0.25	0.02	0.01	0.25	0.00	0.01	0.25	-0.08	0.02	0.25	-0.11	0.02
	0.5	0.01	0.01	0.5	-0.05	0.01	0.5	-0.04	0.03	0.5	-0.10	0.03
	1	0.07	0.03	1	0.00	0.02	1	-0.07	0.04	1	-0.12	0.04
	2	0.08	0.03	2	0.00	0.03	2	-0.09	0.05	2	-0.15	0.06
	3	0.02	0.03	3	0.00	0.03	3	-0.06	0.05	3	-0.26	0.06
	5	0.00	0.04	5	-0.02	0.04	5	-0.12	0.06	5	-0.23	0.07
	10	-0.01	0.04	10	-0.05	0.04	10	-0.18	0.08	10	-0.36	0.07
	20	0.12	0.05	20	-0.01	0.06	20	-0.12	0.09	20	-0.27	0.08
	30	0.06	0.05	30	0.01	0.05	30	-0.20	0.09	30	-0.25	0.10
	50	0.04	0.08	50	-0.08	0.07	50	-0.14	0.10	50	-0.27	0.10
	100	0.08	0.10	100	-0.16	0.08	100	-0.37	0.12	100	-0.23	0.11
	300	0.11	0.13	300	-0.12	0.10	300	-0.13	0.14	300	-0.23	0.12
	500	-0.07	0.15	500	-0.08	0.10	500	-0.18	0.15	500	-0.10	0.12
	1000	-0.07	0.16	1000	0.07	0.14	1000	-0.33	0.16	1000	0.03	0.15
HB5	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.01	0.00	0.05	-0.01	0.00	0.05	0.00	0.00
	0.1	0.00	0.01	0.1	0.01	0.00	0.1	-0.03	0.00	0.1	-0.01	0.00
	0.25	0.00	0.01	0.25	0.01	0.01	0.25	-0.04	0.01	0.25	-0.02	0.01
	0.5	-0.03	0.01	0.5	0.02	0.01	0.5	-0.03	0.01	0.5	-0.01	0.01
	1	-0.04	0.02	1	-0.03	0.01	1	-0.04	0.02	1	-0.02	0.02
	2	-0.05	0.03	2	0.03	0.02	2	-0.05	0.02	2	-0.03	0.03
	3	-0.03	0.03	3	0.02	0.03	3	-0.01	0.03	3	-0.01	0.03
	5	0.00	0.03	5	0.03	0.03	5	-0.03	0.03	5	-0.08	0.03
	10	0.03	0.03	10	0.01	0.03	10	0.00	0.03	10	-0.07	0.03
	20	-0.07	0.03	20	-0.02	0.03	20	-0.01	0.03	20	-0.05	0.03
	30	0.04	0.03	30	0.02	0.03	30	0.01	0.03	30	-0.03	0.03
	50	-0.01	0.03	50	0.01	0.03	50	0.02	0.03	50	-0.03	0.03
	100	0.01	0.03	100	0.01	0.03	100	-0.03	0.03	100	-0.09	0.03
	300	-0.01	0.04	300	0.07	0.03	300	-0.02	0.04	300	-0.07	0.04
	500	0.00	0.04	500	0.00	0.03	500	0.06	0.04	500	-0.01	0.04
	1000	0.07	0.04	1000	0.03	0.03	1000	0.09	0.04	1000	-0.02	0.04

Table S5.3.3: The average distance differences in the donor-acceptor distances of HBs 8-12 following the P2 mutation in s01, s05, s02, and p12. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB8-	s01: P2 L->A			s05: P2 L->A			s02: P2 F->A			p12: P2 W->A		
HB8	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.01	0.00	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.01	0.00
	0.25	0.01	0.01	0.25	-0.01	0.00	0.25	0.03	0.01	0.25	0.02	0.01
	0.5	0.02	0.01	0.5	-0.01	0.01	0.5	0.04	0.01	0.5	0.00	0.01
	1	0.06	0.02	1	-0.01	0.01	1	0.05	0.02	1	0.03	0.02
	2	0.04	0.03	2	-0.01	0.02	2	0.08	0.03	2	0.04	0.02
	3	0.09	0.03	3	-0.03	0.02	3	-0.01	0.03	3	0.07	0.02
	5	0.10	0.03	5	-0.04	0.02	5	0.02	0.03	5	0.06	0.02
	10	0.06	0.03	10	0.00	0.02	10	0.04	0.03	10	0.04	0.02
	20	0.03	0.03	20	0.01	0.03	20	0.02	0.03	20	0.03	0.02
	30	0.05	0.03	30	-0.02	0.03	30	0.02	0.03	30	0.03	0.03
	50	0.00	0.03	50	0.01	0.03	50	0.01	0.03	50	0.04	0.03
	100	0.05	0.04	100	-0.06	0.03	100	-0.02	0.03	100	0.05	0.03
	300	0.06	0.03	300	0.00	0.03	300	0.00	0.04	300	0.01	0.03
	500	0.00	0.04	500	-0.01	0.03	500	0.08	0.04	500	0.09	0.03
	1000	0.04	0.03	1000	0.02	0.03	1000	0.01	0.04	1000	0.06	0.03
HB9	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.01	0.00	0.1	0.00	0.00	0.1	0.01	0.00	0.1	0.01	0.00
	0.25	0.00	0.00	0.25	-0.01	0.00	0.25	0.01	0.00	0.25	-0.01	0.00
	0.5	0.02	0.01	0.5	0.00	0.01	0.5	0.01	0.01	0.5	0.01	0.01
	1	0.00	0.01	1	0.00	0.01	1	0.01	0.01	1	-0.01	0.01
	2	0.02	0.02	2	0.01	0.02	2	-0.03	0.02	2	-0.02	0.02
	3	0.00	0.02	3	0.00	0.02	3	-0.01	0.02	3	-0.02	0.02
	5	0.00	0.02	5	0.00	0.02	5	0.01	0.02	5	-0.02	0.02
	10	0.01	0.02	10	0.00	0.01	10	0.00	0.02	10	-0.05	0.02
	20	0.02	0.02	20	-0.01	0.02	20	0.00	0.02	20	0.00	0.01
	30	0.01	0.02	30	-0.01	0.02	30	0.02	0.02	30	0.01	0.02
	50	-0.01	0.02	50	-0.02	0.02	50	0.00	0.02	50	-0.02	0.02
	100	0.02	0.02	100	0.02	0.01	100	0.03	0.02	100	-0.02	0.02
	300	0.00	0.02	300	-0.02	0.02	300	0.03	0.02	300	0.00	0.02
	500	0.00	0.02	500	-0.02	0.02	500	-0.01	0.02	500	-0.03	0.02
	1000	0.02	0.02	1000	0.00	0.02	1000	0.00	0.02	1000	-0.05	0.02
HB10	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00
	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.02	0.00	0.25	0.02	0.00
	0.5	0.00	0.00	0.5	0.00	0.00	0.5	0.01	0.00	0.5	-0.01	0.00
	1	0.01	0.01	1	0.01	0.01	1	0.02	0.01	1	0.01	0.01
	2	0.00	0.02	2	-0.01	0.02	2	0.03	0.02	2	0.03	0.02
	3	0.00	0.02	3	0.00	0.02	3	0.01	0.01	3	0.01	0.02
	5	-0.02	0.02	5	0.02	0.02	5	0.04	0.02	5	0.03	0.02
	10	-0.01	0.02	10	0.00	0.02	10	-0.01	0.02	10	-0.02	0.02
	20	-0.03	0.02	20	-0.03	0.02	20	-0.01	0.02	20	-0.01	0.02
	30	-0.03	0.02	30	0.00	0.02	30	0.01	0.02	30	0.00	0.02
	50	0.02	0.02	50	0.00	0.02	50	0.01	0.02	50	0.01	0.02
	100	-0.01	0.02	100	0.01	0.02	100	0.04	0.02	100	0.02	0.02
	300	0.02	0.02	300	0.04	0.02	300	0.00	0.02	300	-0.03	0.02
	500	-0.01	0.02	500	-0.02	0.02	500	0.00	0.02	500	0.03	0.02
	1000	-0.01	0.02	1000	0.03	0.02	1000	0.01	0.02	1000	0.06	0.02
HB11	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00
	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.00	0.00
	0.5	0.01	0.00	0.5	0.00	0.00	0.5	0.00	0.00	0.5	0.00	0.00
	1	-0.01	0.01	1	0.01	0.01	1	-0.01	0.01	1	-0.01	0.01
	2	-0.02	0.02	2	-0.02	0.01	2	-0.01	0.02	2	-0.01	0.01
	3	0.01	0.02	3	0.00	0.02	3	0.01	0.01	3	0.00	0.01
	5	-0.02	0.02	5	0.00	0.02	5	-0.01	0.02	5	0.01	0.01
	10	-0.02	0.02	10	0.02	0.01	10	-0.01	0.01	10	-0.01	0.01
	20	-0.01	0.03	20	0.03	0.02	20	-0.01	0.01	20	-0.01	0.01
	30	0.00	0.02	30	0.01	0.02	30	0.01	0.02	30	-0.03	0.01
	50	-0.02	0.02	50	-0.01	0.02	50	0.01	0.02	50	0.01	0.01
	100	-0.03	0.02	100	0.01	0.02	100	0.00	0.01	100	0.01	0.02
	300	-0.01	0.02	300	0.03	0.02	300	0.00	0.02	300	0.02	0.01
	500	-0.02	0.02	500	0.00	0.02	500	-0.03	0.02	500	0.02	0.01
	1000	0.03	0.02	1000	0.03	0.02	1000	-0.02	0.02	1000	0.04	0.01
HB12	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM	t	Av Diff	SEM
	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00
	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00	0.1	0.00	0.00
	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.00	0.00
	0.5	0.00	0.00	0.5	0.00	0.00	0.5	0.00	0.00	0.5	0.00	0.00
	1	-0.02	0.02	1	0.00	0.01	1	0.01	0.01	1	0.00	0.02
	2	-0.01	0.03	2	-0.03	0.02	2	-0.01	0.02	2	0.00	0.02
	3	0.00	0.03	3	0.00	0.02	3	-0.01	0.02	3	0.05	0.02
	5	-0.01	0.03	5	-0.03	0.02	5	0.02	0.02	5	-0.01	0.03
	10	-0.05	0.03	10	-0.02	0.03	10	0.00	0.02	10	-0.01	0.03
	20	0.04	0.03	20	0.02	0.03	20	0.00	0.02	20	-0.06	0.02
	30	0.01	0.03	30	0.01	0.03	30	0.04	0.02	30	0.00	0.03
	50	-0.09	0.04	50	-0.03	0.03	50	0.00	0.03	50	-0.02	0.02
	100	-0.01	0.04	100	0.03	0.03	100	-0.04	0.02	100	0.01	0.03
	300	-0.09	0.05	300	0.00	0.03	300	0.01	0.02	300	0.00	0.04
	500	0.02	0.05	500	-0.04	0.04	500	0.00	0.03	500	-0.03	0.04
	1000	-0.09	0.05	1000	0.01	0.04	1000	-0.01	0.03	1000	0.01	0.04

Table S5.3.4: The ten most perturbed Cα atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P2 mutation in s01, s05, s02, and p12. Only displacements ≥0.01 Å are shown.

s01 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	5ALA_CA	C	0.03	0.00
	2	165MET_CA	A	0.01	0.00
	3	187ASP_CA	A	0.01	0.00
	4	4VAL_CA	C	0.01	0.00
	5	6GLN_CA	C	0.01	0.00
	6				
	7				
	8				
	9				
	10				
0.5	1	187ASP_CA	A	0.10	0.01
	2	5ALA_CA	C	0.10	0.02
	3	186VAL_CA	A	0.08	0.01
	4	6GLN_CA	C	0.08	0.01
	5	165MET_CA	A	0.06	0.01
	6	49MET_CA	A	0.05	0.01
	7	4VAL_CA	C	0.05	0.01
	8	41HID_CA	A	0.04	0.00
	9	46SER_CA	A	0.04	0.00
	10	47GLU_CA	A	0.04	0.00
5	1	244GLN_CA	B	0.16	0.05
	2	97LYS_CA	B	0.15	0.03
	3	46SER_CA	A	0.13	0.04
	4	301SER_CA	B	0.13	0.03
	5	302GLY_CA	B	0.13	0.04
	6	306GLN_CA	B	0.13	0.03
	7	187ASP_CA	A	0.12	0.03
	8	229ASP_CA	A	0.12	0.04
	9	96PRO_CA	B	0.12	0.04
	10	98THR_CA	B	0.12	0.03
50	1	222ARG_CA	B	0.34	0.07
	2	223PHE_CA	B	0.30	0.07
	3	224THR_CA	B	0.29	0.07
	4	225THR_CA	B	0.25	0.06
	5	191ALA_CA	B	0.24	0.07
	6	221ASN_CA	B	0.22	0.05
	7	5ALA_CA	C	0.22	0.03
	8	277ASN_CA	A	0.20	0.09
	9	226THR_CA	B	0.20	0.06
	10	245ASP_CA	B	0.20	0.06
500	1	11LYS_CA	C	0.26	0.12
	2	186VAL_CA	A	0.25	0.04
	3	53ASN_CA	A	0.22	0.07
	4	302GLY_CA	B	0.22	0.05
	5	187ASP_CA	A	0.21	0.04
	6	46SER_CA	A	0.20	0.07
	7	93THR_CA	B	0.20	0.08
	8	1THR_CA	C	0.20	0.11
	9	46SER_CA	B	0.19	0.10
	10	2SER_CA	C	0.19	0.07
1000	1	11LYS_CA	C	0.37	0.15
	2	47GLU_CA	B	0.27	0.10
	3	306GLN_CA	B	0.25	0.07
	4	45THR_CA	A	0.23	0.07
	5	46SER_CA	B	0.23	0.09
	6	277ASN_CA	B	0.22	0.10
	7	46SER_CA	A	0.21	0.08
	8	186VAL_CA	A	0.21	0.05
	9	193ALA_CA	A	0.21	0.07
	10	248ASP_CA	A	0.21	0.08

s05 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	165MET_CA	A	0.01	0.00
	2	4LYS_CA	C	0.01	0.00
	3	5ALA_CA	C	0.01	0.00
	4	6GLN_CA	C	0.01	0.00
	5				
	6				
	7				
	8				
	9				
	10				
0.5	1	165MET_CA	A	0.06	0.00
	2	49MET_CA	A	0.05	0.01
	3	4LYS_CA	C	0.05	0.01
	4	44CYS_CA	A	0.04	0.00
	5	47GLU_CA	A	0.04	0.00
	6	48ASP_CA	A	0.04	0.00
	7	143GLY_CA	A	0.04	0.01
	8	173ALA_CA	A	0.04	0.00
	9	6GLN_CA	C	0.04	0.01
	10	41HID_CA	A	0.03	0.00
5	1	72ASN_CA	A	0.16	0.05
	2	277ASN_CA	B	0.16	0.05
	3	45THR_CA	A	0.13	0.03
	4	283GLY_CA	B	0.13	0.05
	5	44CYS_CA	A	0.12	0.03
	6	46SER_CA	A	0.12	0.04
	7	73VAL_CA	A	0.12	0.04
	8	194ALA_CA	A	0.12	0.05
	9	227LEU_CA	A	0.12	0.04
	10	228ASN_CA	A	0.12	0.04
50	1	47GLU_CA	A	0.20	0.06
	2	53ASN_CA	A	0.20	0.06
	3	46SER_CA	A	0.19	0.06
	4	72ASN_CA	B	0.19	0.09
	5	50LEU_CA	A	0.18	0.06
	6	51ASN_CA	A	0.18	0.06
	7	52PRO_CA	A	0.18	0.06
	8	279ARG_CA	B	0.18	0.06
	9	49MET_CA	A	0.17	0.05
	10	72ASN_CA	A	0.17	0.09
500	1	11SER_CA	C	0.44	0.19
	2	277ASN_CA	B	0.32	0.11
	3	278GLY_CA	B	0.29	0.09
	4	10LEU_CA	C	0.29	0.09
	5	277ASN_CA	A	0.28	0.10
	6	53ASN_CA	A	0.26	0.08
	7	72ASN_CA	A	0.25	0.10
	8	190THR_CA	B	0.25	0.08
	9	218TRP_CA	A	0.23	0.06
	10	191ALA_CA	B	0.23	0.10
1000	1	11SER_CA	C	0.36	0.17
	2	278GLY_CA	A	0.21	0.12
	3	52PRO_CA	B	0.21	0.10
	4	55GLU_CA	B	0.21	0.08
	5	56ASP_CA	B	0.21	0.11
	6	59ILE_CA	B	0.20	0.10
	7	60ARG_CA	B	0.20	0.09
	8	93THR_CA	B	0.20	0.08
	9	278GLY_CA	B	0.20	0.09
	10	10LEU_CA	C	0.20	0.09

s02 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	187ASP_CA	A	0.02	0.00
	2	5ALA_CA	C	0.02	0.00
	3	4THR_CA	C	0.01	0.00
	4				
	5				
	6				
	7				
	8				
	9				
	10				
0.5	1	188ARG_CA	A	0.13	0.01
	2	187ASP_CA	A	0.12	0.01
	3	5ALA_CA	C	0.11	0.01
	4	7SER_CA	C	0.09	0.01
	5	186VAL_CA	A	0.08	0.01
	6	6GLN_CA	C	0.08	0.01
	7	189GLN_CA	A	0.07	0.01
	8	4THR_CA	C	0.07	0.01
	9	49MET_CA	A	0.06	0.01
	10	165MET_CA	A	0.06	0.01
5	1	5ALA_CA	C	0.19	0.03
	2	277ASN_CA	B	0.17	0.06
	3	23GLY_CA	A	0.16	0.05
	4	256GLN_CA	B	0.16	0.05
	5	4THR_CA	C	0.16	0.03
	6	47GLU_CA	A	0.15	0.05
	7	277ASN_CA	A	0.15	0.06
	8	278GLY_CA	A	0.15	0.06
	9	46SER_CA	A	0.14	0.04
	10	50LEU_CA	A	0.14	0.04
50	1	50LEU_CA	A	0.36	0.05
	2	51ASN_CA	A	0.35	0.06
	3	56ASP_CA	A	0.33	0.07
	4	188ARG_CA	A	0.33	0.04
	5	189GLN_CA	A	0.33	0.04
	6	53ASN_CA	A	0.32	0.06
	7	47GLU_CA	A	0.31	0.07
	8	55GLU_CA	A	0.31	0.06
	9	52PRO_CA	A	0.28	0.06
	10	46SER_CA	A	0.27	0.07
500	1	1SER_CA	C	0.57	0.19
	2	189GLN_CA	A	0.36	0.05
	3	188ARG_CA	A	0.35	0.05
	4	47GLU_CA	A	0.29	0.09
	5	190THR_CA	A	0.29	0.06
	6	24THR_CA	A	0.28	0.07
	7	50LEU_CA	A	0.28	0.07
	8	187ASP_CA	A	0.28	0.04
	9	48ASP_CA	A	0.27	0.06
	10	51ASN_CA	A	0.26	0.07
1000	1	1SER_CA	C	0.67	0.19
	2	47GLU_CA	A	0.42	0.11
	3	189GLN_CA	A	0.37	0.06
	4	50LEU_CA	A	0.36	0.07
	5	188ARG_CA	A	0.36	0.05
	6	46SER_CA	A	0.35	0.10
	7	51ASN_CA	A	0.35	0.07
	8	190THR_CA	A	0.35	0.06
	9	48ASP_CA	A	0.32	0.07
	10	187ASP_CA	A	0.31	0.04

p12 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	5ALA_CA	C	0.03	0.00
	2	187ASP_CA	A	0.02	0.00
	3	49MET_CA	A	0.01	0.00
	4	164HIE_CA	A	0.01	0.00
	5	186VAL_CA	A	0.01	0.00
	6	188ARG_CA	A	0.01	0.00
	7	189GLN_CA	A	0.01	0.00
	8	4PHE_CA	C	0.01	0.00
	9	6GLN_CA	C	0.01	0.00
	10				
0.5	1	186VAL_CA	A	0.23	0.01
	2	187ASP_CA	A	0.23	0.02
	3	5ALA_CA	C	0.16	0.02
	4	164HIE_CA	A	0.12	0.01
	5	165MET_CA	A	0.12	0.01
	6	4PHE_CA	C	0.12	0.01
	7	49MET_CA	A	0.10	0.01
	8	166GLU_CA	A	0.10	0.01
	9	173ALA_CA	A	0.10	0.01
	10	50LEU_CA	A	0.09	0.01
5	1	186VAL_CA	A	0.49	0.04
	2	187ASP_CA	A	0.46	0.03
	3	185PHE_CA	A	0.34	0.03
	4	188ARG_CA	A	0.29	0.04
	5	189GLN_CA	A	0.23	0.03
	6	50LEU_CA	A	0.21	0.04
	7	51ASN_CA	A	0.21	0.04
	8	194ALA_CA	A	0.21	0.04
	9	190THR_CA	A	0.20	0.03
	10	5ALA_CA	C	0.19	0.03
50	1	187ASP_CA	A	0.44	0.04
	2	186VAL_CA	A	0.41	0.04
	3	55GLU_CA	A	0.35	0.06
	4	185PHE_CA	A	0.33	0.04
	5	277ASN_CA	A	0.33	0.09
	6	278GLY_CA	A	0.31	0.09
	7	54TYR_CA	A	0.30	0.04
	8	188ARG_CA	A	0.30	0.05
	9	56ASP_CA	A	0.29	0.06
	10	274ASN_CA	A	0.29	0.08
500	1	186VAL_CA	A	0.51	0.05
	2	187ASP_CA	A	0.47	0.04
	3	185PHE_CA	A	0.39	0.04
	4	274ASN_CA	A	0.38	0.09
	5	278GLY_CA	B	0.36	0.10
	6	4PHE_CA	C	0.36	0.05
	7	55GLU_CA	A	0.35	0.07
	8	56ASP_CA	A	0.32	0.08
	9	277ASN_CA	B	0.31	0.11
	10	5ALA_CA	C	0.29	0.04
1000	1	186VAL_CA	A	0.53	0.05
	2	187ASP_CA	A	0.51	0.05
	3	4PHE_CA	C	0.39	0.05
	4	185PHE_CA	A	0.37	0.05
	5	47GLU_CA	A	0.33	0.08
	6	59ILE_CA	A	0.31	0.10
	7	5ALA_CA	C	0.31	0.04
	8	46SER_CA	A	0.26	0.08
	9	60ARG_CA	A	0.26	0.11
	10	194ALA_CA	A	0.25	0.07

Table S5.3.5: The ten most perturbed non-hydrogen atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P2 mutation in s01, s05, s02, and p12.

s01 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	5ALA_N	C	0.05	0.00
	2	5ALA_CB	C	0.04	0.01
	3	165MET_CB	A	0.03	0.00
	4	5ALA_CA	C	0.03	0.00
	5	4VAL_O	C	0.03	0.00
	6	4VAL_C	C	0.02	0.00
	7	165MET_CG	A	0.02	0.00
	8	41HID_ND1	A	0.02	0.00
	9	5ALA_C	C	0.02	0.00
	10	49MET_CB	A	0.02	0.00
0.5	1	165MET_CE	A	0.20	0.02
	2	49MET_SD	A	0.20	0.03
	3	187ASP_O	A	0.19	0.02
	4	165MET_SD	A	0.19	0.02
	5	5ALA_CB	C	0.17	0.03
	6	165MET_CG	A	0.15	0.02
	7	165MET_CB	A	0.12	0.01
	8	187ASP_C	A	0.12	0.01
	9	5ALA_N	C	0.12	0.01
	10	49MET_CE	A	0.11	0.02
5	1	47GLU_OE2	A	0.38	0.10
	2	49MET_SD	A	0.34	0.06
	3	47GLU_OE2	B	0.26	0.09
	4	97LYS_CE	B	0.26	0.06
	5	236LYS_NZ	A	0.26	0.10
	6	107GLN_OE1	A	0.25	0.08
	7	244GLN_OE1	B	0.23	0.08
	8	245ASP_OD2	B	0.23	0.08
	9	97LYS_CD	B	0.23	0.05
	10	47GLU_CD	A	0.23	0.08
50	1	222ARG_NH1	B	0.49	0.18
	2	11LYS_NZ	C	0.49	0.17
	3	222ARG_NH1	A	0.48	0.18
	4	49MET_SD	A	0.48	0.07
	5	244GLN_NE2	B	0.45	0.15
	6	235MET_CE	B	0.44	0.14
	7	11LYS_CE	C	0.44	0.15
	8	222ARG_CG	B	0.43	0.11
	9	222ARG_NH2	B	0.43	0.17
	10	222ARG_CZ	B	0.42	0.16
500	1	165MET_CE	A	1.26	0.13
	2	165MET_SD	A	0.96	0.10
	3	49MET_CE	A	0.72	0.13
	4	306GLN_NE2	A	0.64	0.23
	5	165MET_CG	A	0.63	0.07
	6	72ASN_OD1	B	0.61	0.20
	7	222ARG_NH2	A	0.54	0.25
	8	49MET_SD	A	0.53	0.09
	9	279ARG_NH1	B	0.47	0.19
	10	248ASP_OD1	A	0.46	0.12
1000	1	165MET_CE	A	1.29	0.14
	2	49MET_CE	A	1.18	0.14
	3	165MET_SD	A	1.09	0.10
	4	279ARG_NH1	B	0.94	0.22
	5	11LYS_NZ	C	0.86	0.32
	6	165MET_CG	A	0.77	0.07
	7	11LYS_CE	C	0.76	0.28
	8	279ARG_CZ	B	0.72	0.17
	9	279ARG_NH2	B	0.61	0.18
	10	279ARG_NE	B	0.61	0.13

s05 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	5ALA_CB	C	0.04	0.01
	2	5ALA_N	C	0.04	0.00
	3	4LYS_O	C	0.02	0.00
	4	165MET_CB	A	0.02	0.00
	5	5ALA_O	C	0.02	0.00
	6	5ALA_C	C	0.02	0.00
	7	165MET_CG	A	0.02	0.00
	8	41HID_ND1	A	0.02	0.00
	9	4LYS_CB	C	0.02	0.00
	10	41HID_CG	A	0.02	0.00
0.5	1	49MET_SD	A	0.19	0.03
	2	165MET_SD	A	0.17	0.02
	3	165MET_CG	A	0.13	0.02
	4	165MET_CE	A	0.13	0.03
	5	49MET_CE	A	0.13	0.02
	6	165MET_CB	A	0.10	0.01
	7	49MET_CG	A	0.09	0.01
	8	49MET_CB	A	0.08	0.01
	9	165MET_O	A	0.07	0.01
	10	189GLN_CD	A	0.07	0.01
5	1	49MET_SD	A	0.31	0.05
	2	244GLN_NE2	B	0.27	0.08
	3	72ASN_ND2	A	0.26	0.10
	4	49MET_CE	A	0.25	0.05
	5	47GLU_OE1	B	0.25	0.11
	6	51ASN_ND2	A	0.23	0.08
	7	256GLN_NE2	A	0.22	0.06
	8	235MET_CE	A	0.21	0.10
	9	72ASN_CB	A	0.21	0.07
	10	228ASN_ND2	A	0.21	0.07
50	1	222ARG_NH1	A	0.46	0.16
	2	49MET_SD	A	0.46	0.06
	3	222ARG_NH2	A	0.46	0.15
	4	49MET_CE	A	0.42	0.08
	5	222ARG_CZ	A	0.40	0.13
	6	189GLN_NE2	A	0.39	0.08
	7	49MET_CG	A	0.37	0.06
	8	244GLN_NE2	A	0.36	0.15
	9	72ASN_ND2	A	0.35	0.17
	10	244GLN_NE2	B	0.34	0.15
500	1	49MET_CE	A	0.88	0.12
	2	165MET_SD	A	0.74	0.09
	3	49MET_SD	A	0.72	0.08
	4	256GLN_NE2	A	0.66	0.15
	5	165MET_CG	A	0.58	0.06
	6	11SER_OC2	C	0.56	0.23
	7	277ASN_OD1	A	0.55	0.19
	8	165MET_CE	A	0.53	0.10
	9	235MET_CE	B	0.51	0.20
	10	223PHE_CE2	B	0.49	0.16
1000	1	49MET_CE	A	1.13	0.13
	2	235MET_CE	A	0.87	0.22
	3	222ARG_NH2	B	0.83	0.32
	4	222ARG_NH1	B	0.82	0.29
	5	222ARG_CZ	B	0.76	0.27
	6	165MET_SD	A	0.71	0.09
	7	49MET_SD	A	0.67	0.08
	8	236LYS_NZ	A	0.62	0.21
	9	235MET_SD	A	0.61	0.17
	10	222ARG_NE	B	0.61	0.24

s02 t(ps)	Non-H Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	5ALA_CB	C	0.04	0.01
	2	5ALA_N	C	0.03	0.00
	3	49MET_CE	A	0.03	0.00
	4	5ALA_CA	C	0.02	0.00
	5	187ASP_CA	A	0.02	0.00
	6	5ALA_C	C	0.01	0.00
	7	49MET_SD	A	0.01	0.00
	8	188ARG_O	A	0.01	0.00
	9	4THR_C	C	0.01	0.00
	10	165MET_SD	A	0.01	0.00
0.5	1	165MET_SD	A	0.32	0.03
	2	165MET_CE	A	0.31	0.03
	3	49MET_CE	A	0.22	0.03
	4	165MET_CG	A	0.19	0.02
	5	188ARG_O	A	0.16	0.02
	6	49MET_SD	A	0.16	0.02
	7	5ALA_N	C	0.14	0.02
	8	188ARG_N	A	0.13	0.01
	9	188ARG_CB	A	0.13	0.01
	10	187ASP_O	A	0.13	0.01
5	1	165MET_CE	A	0.38	0.07
	2	165MET_SD	A	0.29	0.05
	3	4THR_CG2	C	0.25	0.05
	4	72ASN_ND2	B	0.23	0.09
	5	134PHE_CZ	A	0.23	0.07
	6	5ALA_N	C	0.22	0.03
	7	80HID_NE2	A	0.22	0.07
	8	238ASN_OD1	B	0.22	0.06
	9	277ASN_OD1	B	0.22	0.09
	10	107GLN_NE2	A	0.22	0.09
50	1	222ARG_NH2	B	0.55	0.20
	2	222ARG_NH1	B	0.53	0.21
	3	165MET_CE	A	0.52	0.09
	4	50LEU_CD2	A	0.51	0.09
	5	56ASP_OD1	A	0.48	0.09
	6	222ARG_CZ	B	0.45	0.18
	7	72ASN_OD1	A	0.44	0.17
	8	67LEU_CD2	B	0.43	0.10
	9	51ASN_OD1	A	0.42	0.10
	10	50LEU_CG	A	0.41	0.07
500	1	1SER_OG	C	0.75	0.28
	2	1SER_N	C	0.70	0.23
	3	165MET_CE	A	0.68	0.10
	4	1SER_CB	C	0.58	0.23
	5	1SER_CA	C	0.57	0.19
	6	11ARG_NH2	C	0.55	0.38
	7	235MET_CE	A	0.54	0.22
	8	165MET_SD	A	0.53	0.09
	9	235MET_SD	A	0.52	0.17
	10	60ARG_NH1	A	0.51	0.23
1000	1	1SER_OG	C	1.04	0.27
	2	1SER_CB	C	0.88	0.23
	3	294PHE_CE1	B	0.71	0.16
	4	294PHE_CZ	B	0.69	0.17
	5	1SER_N	C	0.68	0.24
	6	1SER_CA	C	0.67	0.19
	7	55GLU_OE2	A	0.67	0.19
	8	165MET_CE	A	0.63	0.12
	9	46SER_OG	A	0.58	0.15
	10	72ASN_ND2	B	0.54	0.21

p12 t(ps)	Non-H Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	5ALA_N	C	0.06	0.01
	2	5ALA_CB	C	0.04	0.01
	3	187ASP_O	A	0.04	0.00
	4	5ALA_O	C	0.03	0.00
	5	5ALA_CA	C	0.03	0.00
	6	5ALA_C	C	0.03	0.00
	7	187ASP_CA	A	0.02	0.00
	8	186VAL_O	A	0.02	0.00
	9	181PHE_CZ	A	0.02	0.00
	10	4PHE_C	C	0.02	0.00
0.5	1	165MET_SD	A	0.46	0.04
	2	165MET_CE	A	0.37	0.03
	3	187ASP_O	A	0.33	0.02
	4	186VAL_O	A	0.30	0.02
	5	165MET_CG	A	0.27	0.02
	6	186VAL_CG2	A	0.26	0.02
	7	186VAL_C	A	0.26	0.02
	8	186VAL_CB	A	0.25	0.01
	9	187ASP_N	A	0.24	0.02
	10	187ASP_CA	A	0.24	0.02
5	1	165MET_SD	A	0.64	0.07
	2	187ASP_O	A	0.51	0.04
	3	186VAL_O	A	0.50	0.04
	4	186VAL_C	A	0.50	0.04
	5	165MET_CE	A	0.50	0.07
	6	186VAL_CA	A	0.49	0.04
	7	187ASP_N	A	0.49	0.04
	8	186VAL_CG2	A	0.48	0.04
	9	186VAL_CB	A	0.47	0.04
	10	187ASP_CA	A	0.46	0.03
50	1	165MET_SD	A	1.00	0.09
	2	165MET_CE	A	0.77	0.08
	3	4PHE_CZ	C	0.60	0.10
	4	4PHE_CE2	C	0.57	0.10
	5	4PHE_CE1	C	0.55	0.10
	6	49MET_CE	A	0.52	0.10
	7	277ASN_OD1	A	0.52	0.15
	8	187ASP_O	A	0.51	0.05
	9	277ASN_ND2	A	0.50	0.14
	10	49MET_SD	A	0.49	0.09
500	1	165MET_SD	A	1.00	0.09
	2	4PHE_CE1	C	0.91	0.17
	3	165MET_CE	A	0.89	0.09
	4	4PHE_CZ	C	0.89	0.19
	5	4PHE_CD1	C	0.78	0.13
	6	4PHE_CE2	C	0.75	0.18
	7	11TYR_OH	C	0.74	0.26
	8	55GLU_OE1	A	0.63	0.18
	9	49MET_CE	A	0.62	0.13
	10	11TYR_CZ	C	0.62	0.22
1000	1	4PHE_CZ	C	1.11	0.18
	2	165MET_SD	A	1.04	0.09
	3	4PHE_CE2	C	1.01	0.18
	4	4PHE_CE1	C	1.00	0.17
	5	1LYS_NZ	C	0.89	0.24
	6	165MET_CE	A	0.88	0.10
	7	4PHE_CD1	C	0.80	0.13
	8	4PHE_CD2	C	0.79	0.14
	9	1LYS_CE	C	0.74	0.21
	10	49MET_CE	A	0.71	0.15

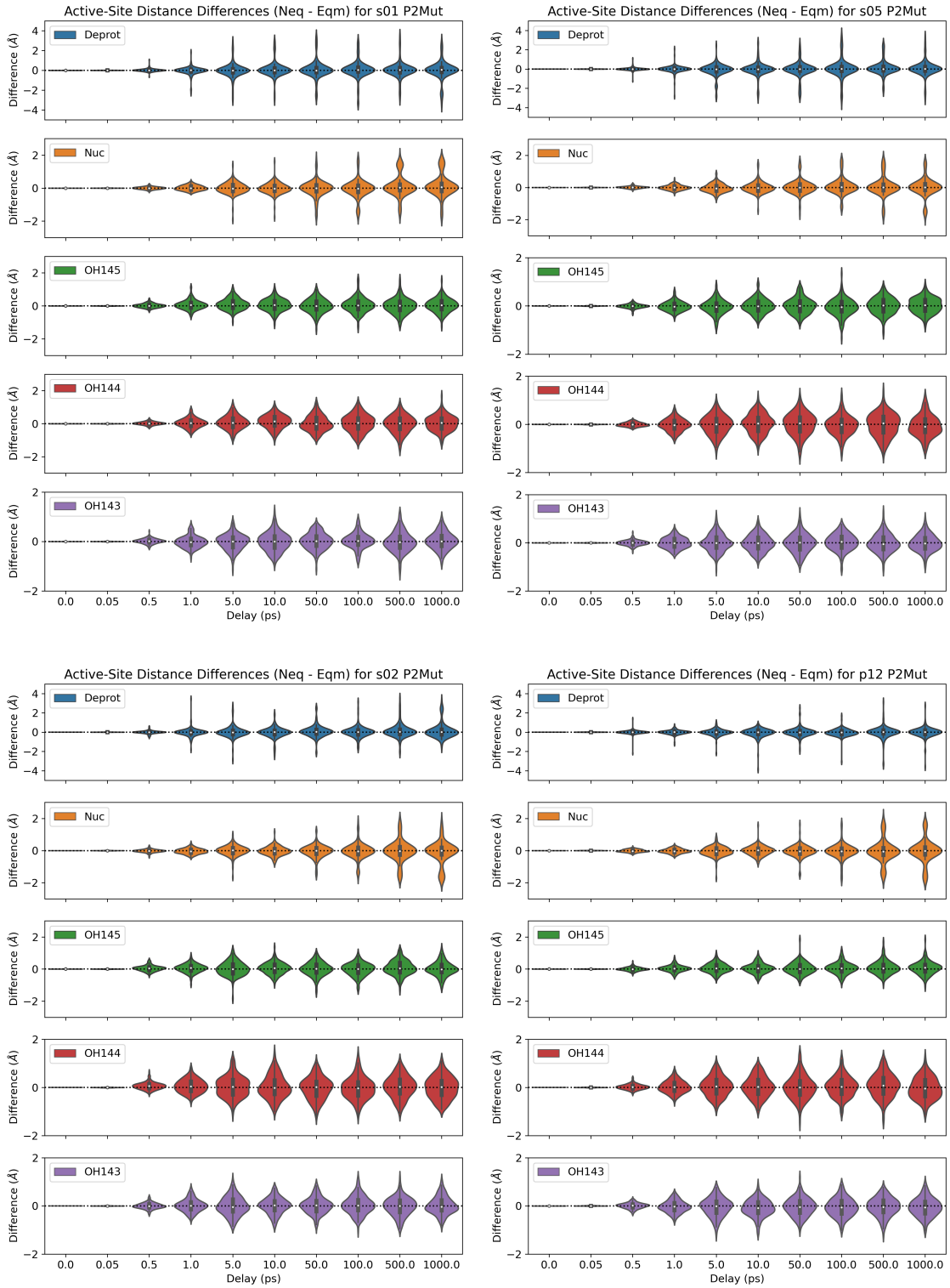


Figure S5.3.1: Distributions of the distance differences in the five active-site distances ($\Delta d = d_{\mathrm{neq}} - d_{\mathrm{eqm}}$) following the P2 mutation in s01, s05, s02, and p12.

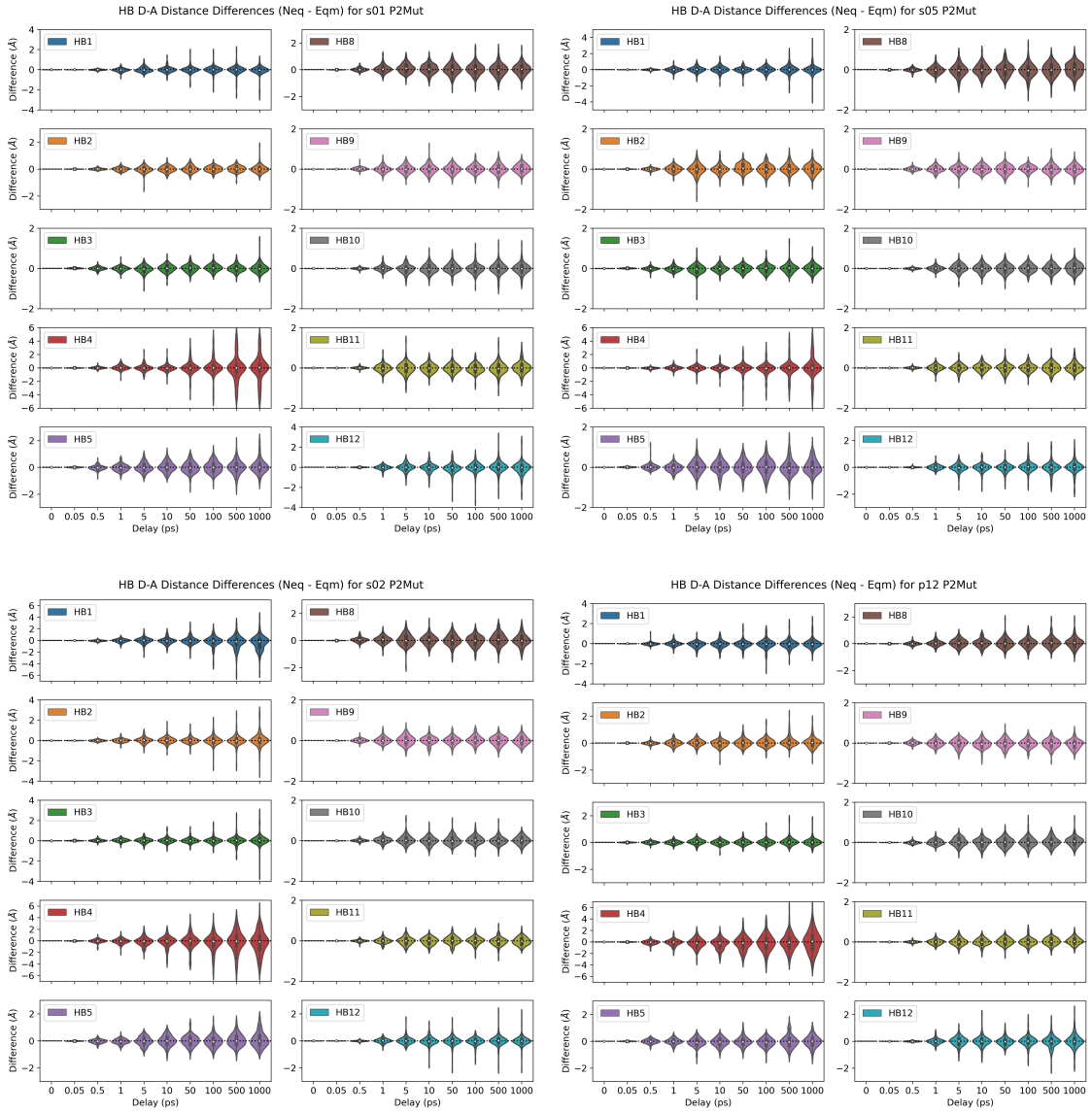


Figure S5.3.2: Distributions of the distance differences in the ten donor-acceptor distances of HBs 1-5 and 8-12 following the P2 mutation in s01, s05, s02, and p12.

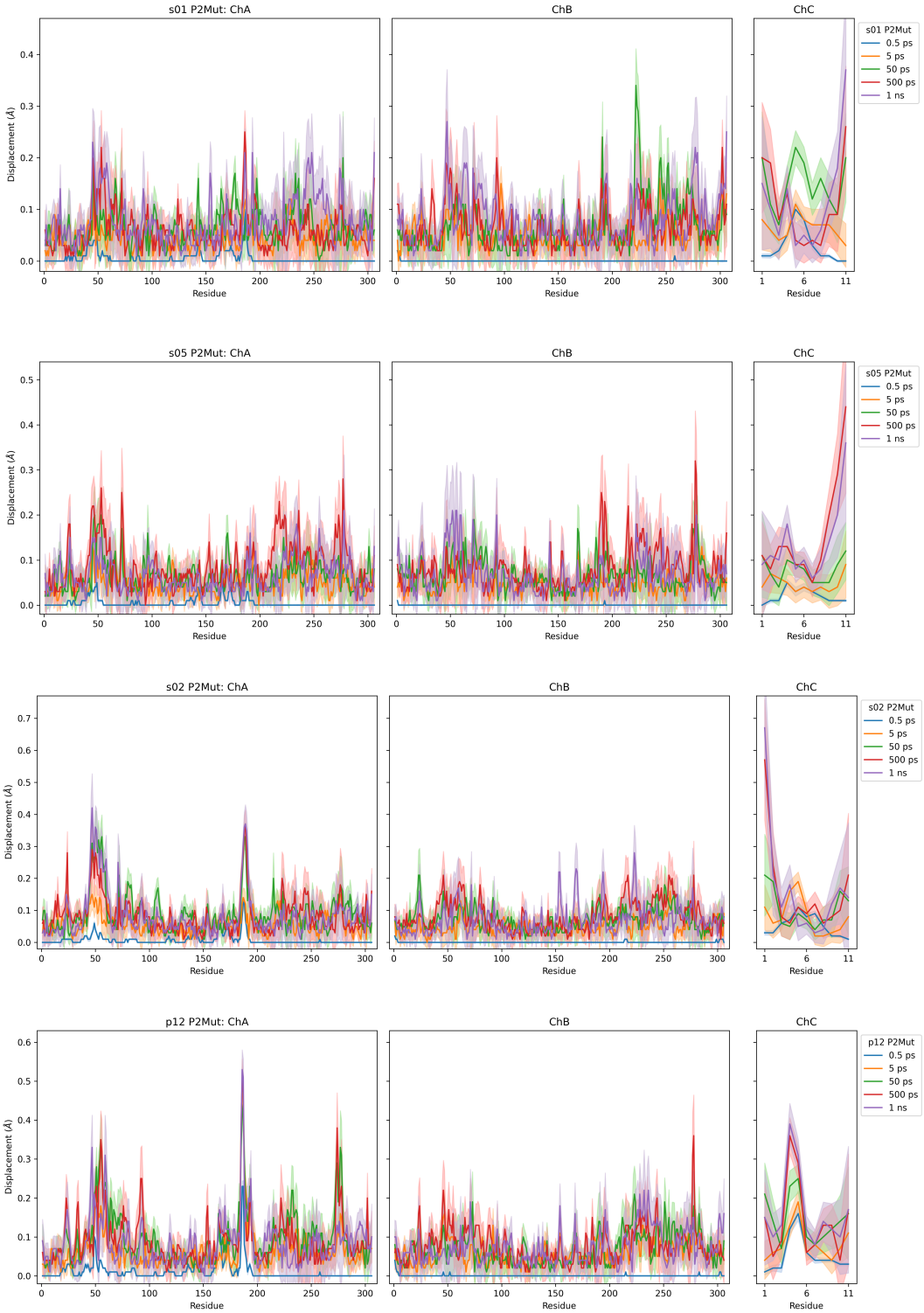


Figure S5.3.3: Magnitudes of the average M^{pro} Cα displacement vectors, with ± SEM shown as shaded regions, following the P2 mutation in s01, s05, s02, and p12.

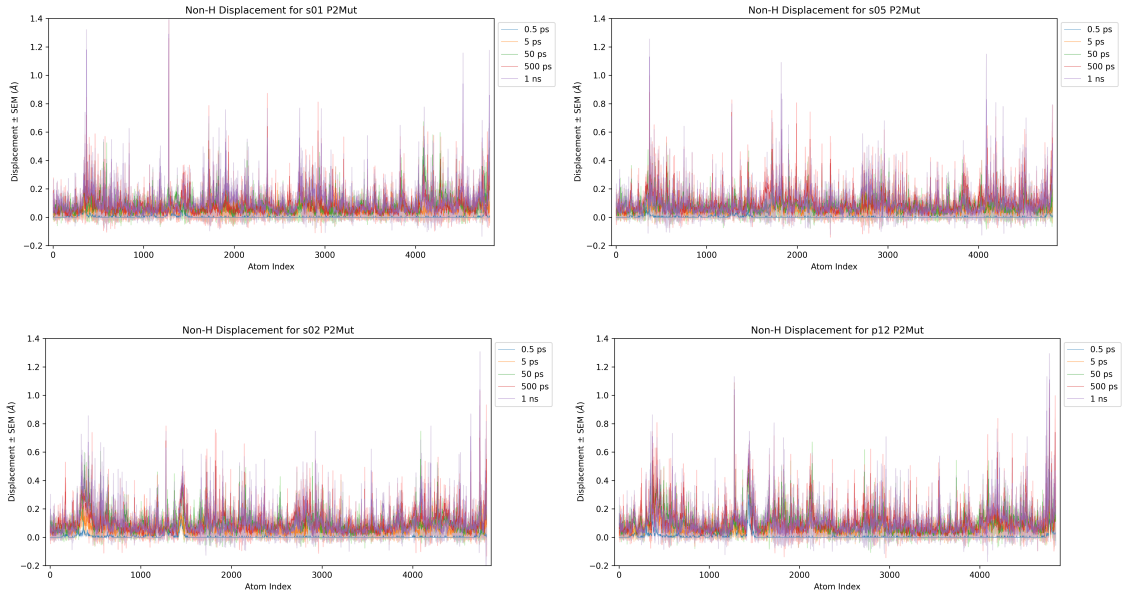
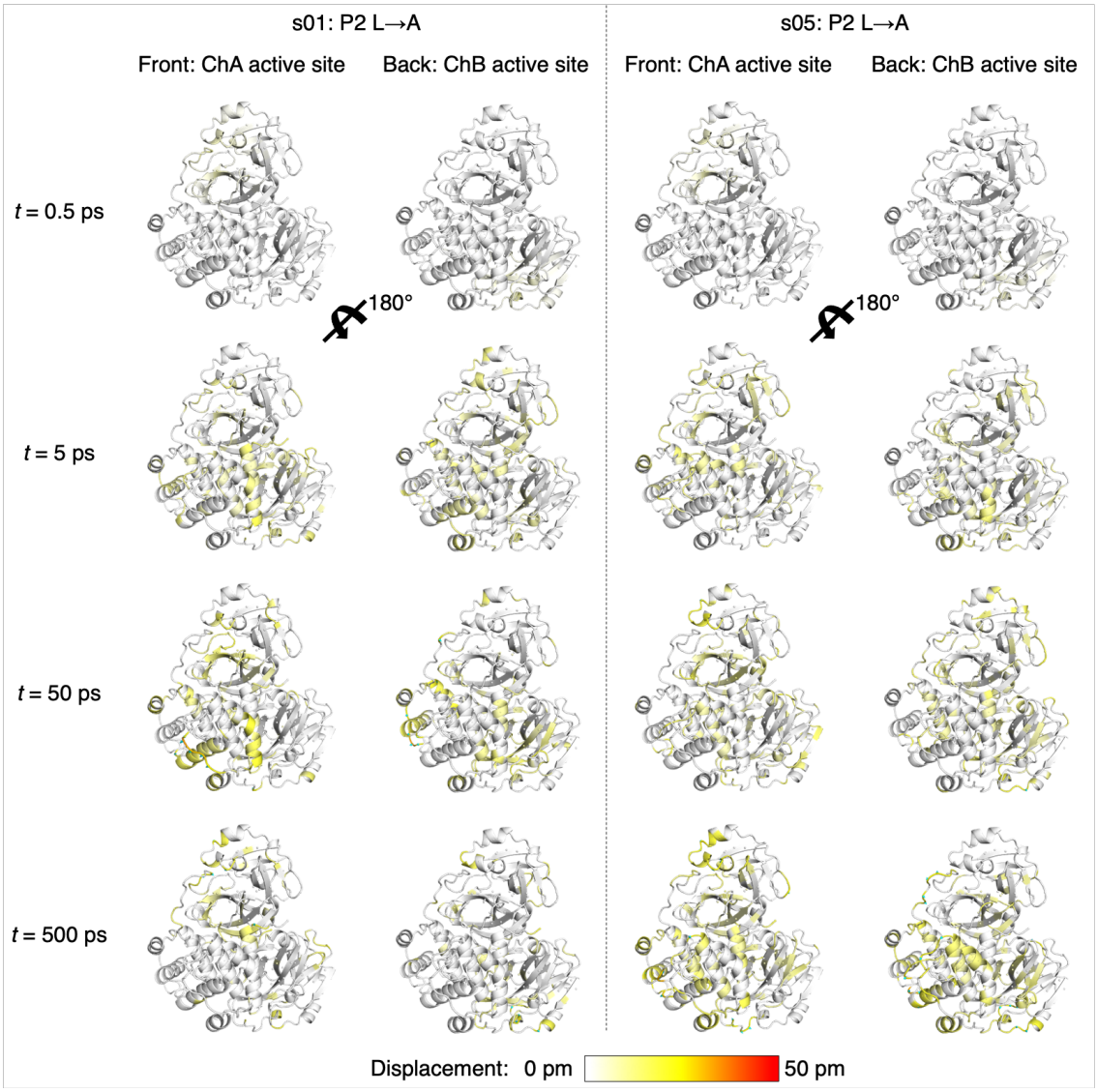


Figure S5.3.4: Magnitudes of the average non-hydrogen atom displacement vectors in the M^{pro}-peptide complex, with ± SEM shown as shaded regions, following the P2 mutation in s01, s05, s02, and p12.


s01: P2 L→A
s05: P2 L→A
Front: ChA active site
Back: ChB active site
Front: ChA active site
Back: ChB active site
t = 0.5 ps
180°
180°
t = 5 ps
t = 50 ps
t = 500 ps
Displacement: 0 pm
50 pm

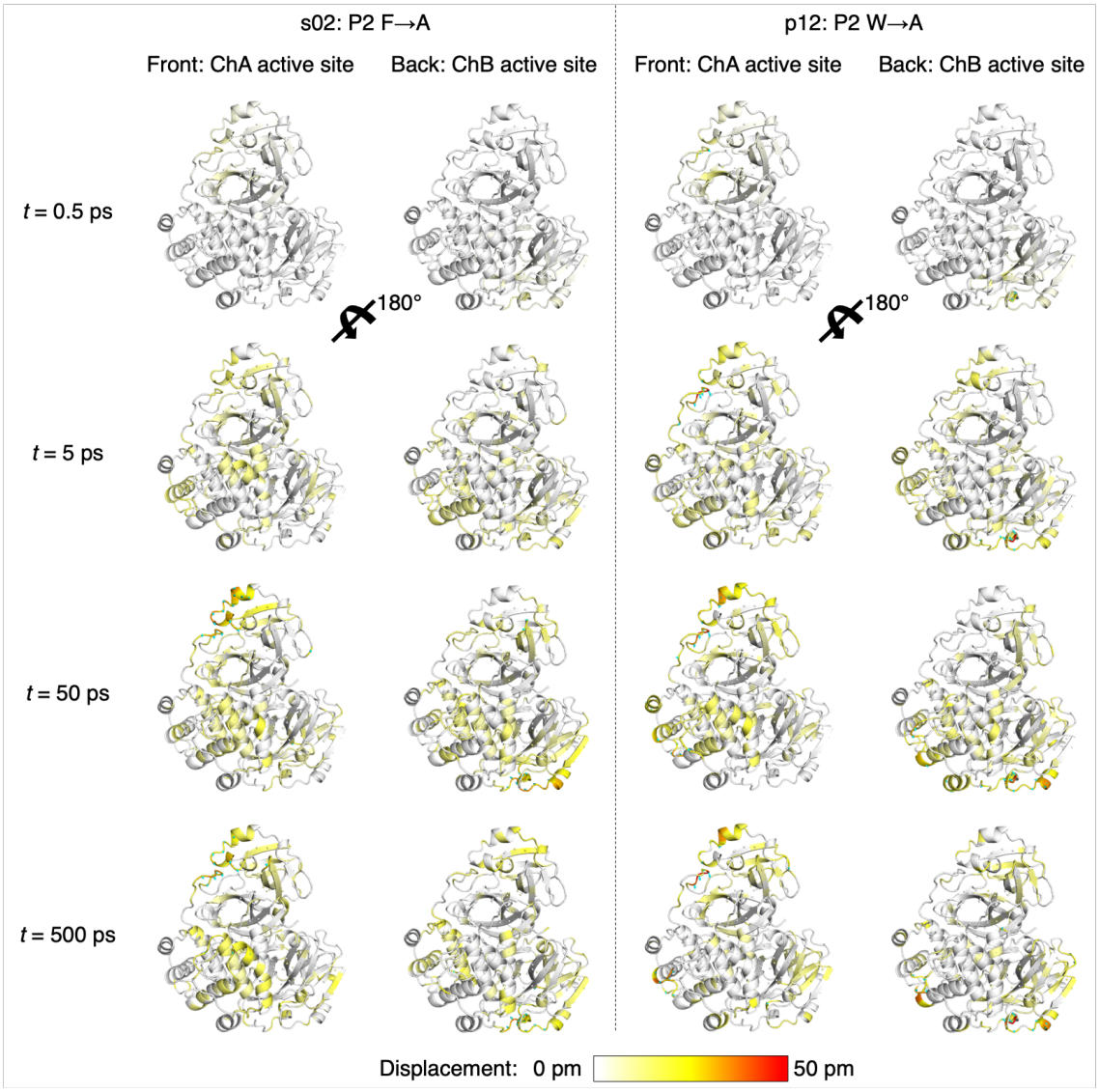


Figure S5.3.5: Views of the M^{pro} response to the P2 mutation in s01, s05, s02, and p12 from averaging Cα displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as cyan arrows with a scale-up factor of 5.[7]

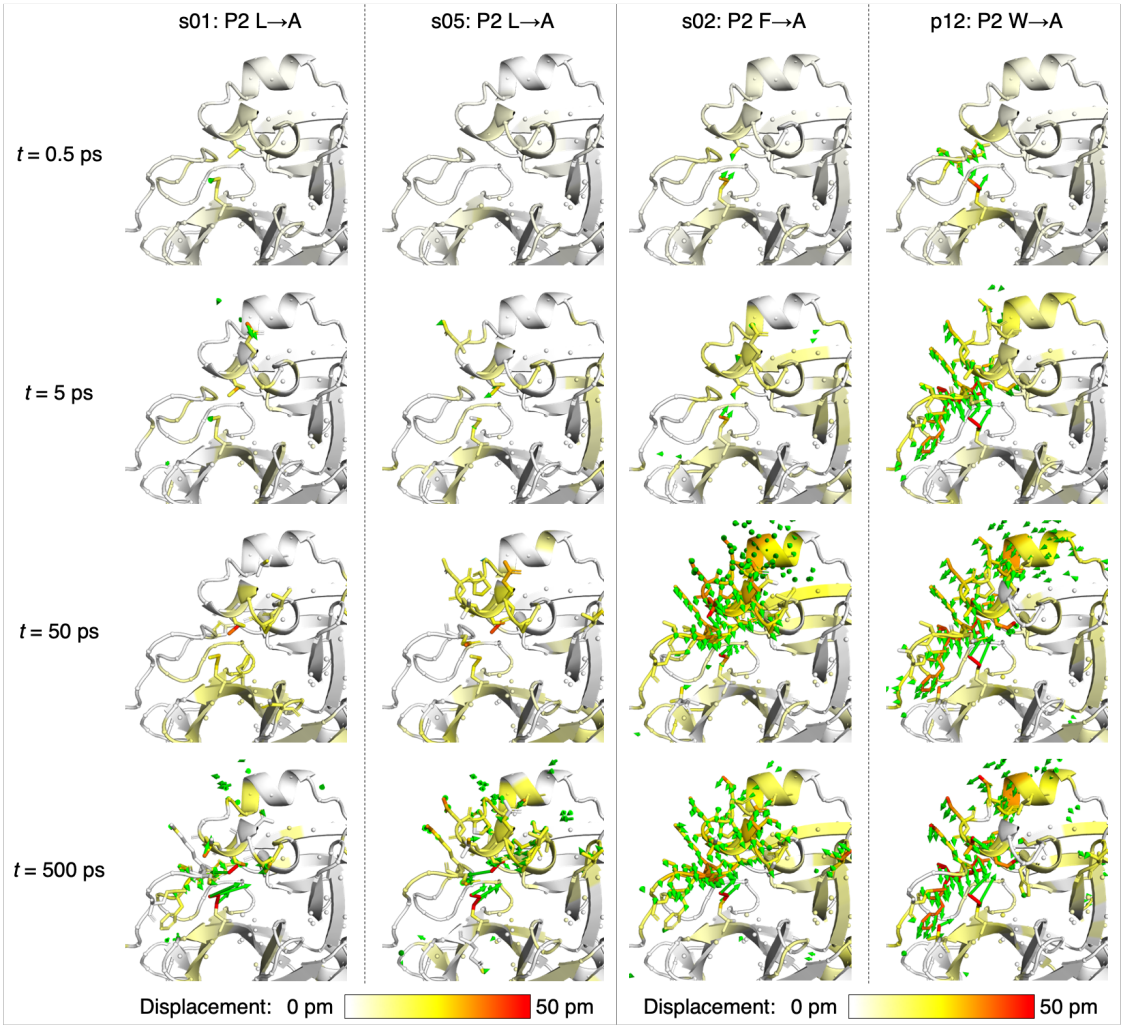
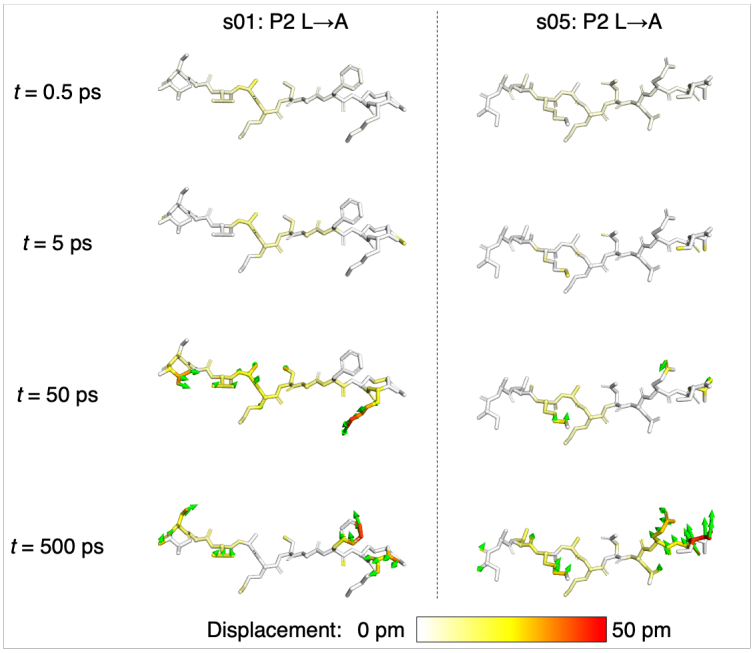


Figure S5.3.6: Views of the M^{pro} response to the P2 mutation in s01, s05, s02, and p12 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations, with a focus on the S2 subsite. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7] Residues that show such significant displacements and that are within 10 Å of the substrate P2 residue are shown as sticks.



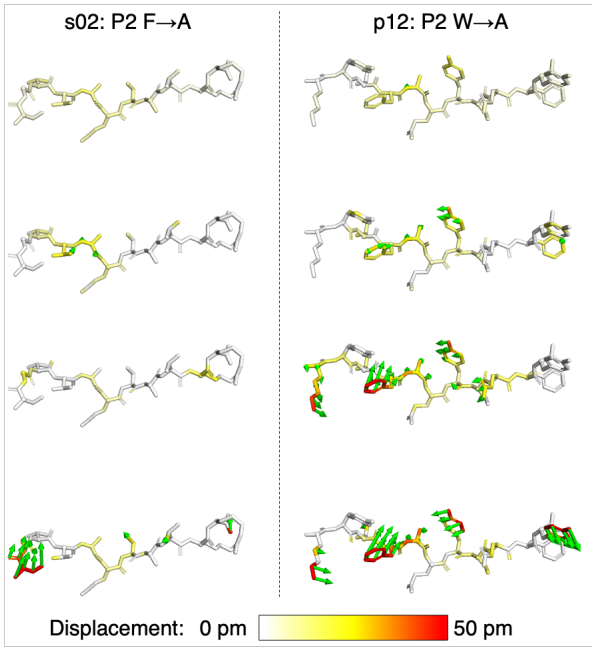


Figure S5.3.7: Views of the peptide response to the P2 mutation in s01, s05, s02, and p12 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7]

Section S5.4A: Effect of P3 mutation

Table S5.4A.1: The average distance differences in the five active-site distances ($\Delta d = d_{neq} - d_{eqm}$) following the P3 mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

Active Site	s01: P3 V->A				s05: P3 K->A		
Deprot	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.02	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.00	0.00		0.25	0.01	0.01
	0.5	0.00	0.00		0.5	0.02	0.02
	1	-0.03	0.03		1	-0.02	0.04
	2	-0.07	0.04		2	-0.01	0.06
	3	0.06	0.05		3	0.03	0.07
	5	-0.04	0.05		5	0.05	0.07
	10	0.07	0.06		10	0.07	0.07
	20	0.06	0.06		20	-0.04	0.07
	30	0.03	0.07		30	0.04	0.08
	50	0.01	0.07		50	-0.06	0.07
	100	0.01	0.06		100	0.08	0.08
	300	-0.09	0.07		300	0.12	0.08
	500	0.01	0.07		500	0.11	0.07
	1000	0.15	0.09		1000	0.04	0.07
Nuc	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.01	0.00
	0.1	0.00	0.00		0.1	0.01	0.00
	0.25	0.00	0.00		0.25	0.02	0.00
	0.5	0.00	0.00		0.5	0.02	0.01
	1	-0.01	0.01		1	0.01	0.02
	2	-0.01	0.02		2	-0.02	0.02
	3	-0.05	0.02		3	0.02	0.02
	5	-0.01	0.02		5	-0.02	0.02
	10	-0.02	0.02		10	0.01	0.02
	20	0.04	0.03		20	-0.03	0.03
	30	-0.04	0.03		30	0.03	0.03
	50	-0.01	0.04		50	-0.02	0.04
	100	-0.03	0.04		100	-0.03	0.04
	300	-0.09	0.05		300	-0.04	0.05
	500	0.10	0.04		500	0.03	0.05
	1000	0.00	0.05		1000	0.07	0.05
OH145	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.02	0.00
	0.1	0.00	0.00		0.1	-0.05	0.00
	0.25	-0.01	0.00		0.25	-0.04	0.01
	0.5	-0.01	0.01		0.5	-0.02	0.01
	1	-0.01	0.02		1	-0.01	0.02
	2	0.02	0.03		2	0.00	0.03
	3	0.01	0.03		3	-0.03	0.03
	5	0.04	0.03		5	-0.04	0.02
	10	0.00	0.03		10	-0.02	0.03
	20	-0.04	0.03		20	0.00	0.02
	30	0.01	0.03		30	0.00	0.03
	50	-0.05	0.03		50	-0.01	0.02
	100	-0.05	0.04		100	-0.04	0.03
	300	-0.04	0.03		300	0.03	0.03
	500	-0.07	0.03		500	-0.02	0.03
	1000	0.01	0.03		1000	0.01	0.03
OH144	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.01	0.00
	0.1	0.00	0.00		0.1	-0.02	0.00
	0.25	-0.01	0.00		0.25	-0.02	0.01
	0.5	-0.01	0.01		0.5	0.00	0.01
	1	0.01	0.02		1	-0.02	0.02
	2	0.02	0.03		2	-0.01	0.03
	3	0.02	0.04		3	0.00	0.03
	5	0.05	0.04		5	0.03	0.04
	10	0.08	0.03		10	0.04	0.03
	20	-0.07	0.04		20	0.03	0.03
	30	0.04	0.04		30	0.03	0.03
	50	-0.03	0.04		50	0.00	0.04
	100	-0.04	0.04		100	0.02	0.03
	300	-0.03	0.04		300	-0.03	0.04
	500	-0.04	0.04		500	0.03	0.04
	1000	-0.01	0.04		1000	0.02	0.04
OH143	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.04	0.00
	0.1	0.00	0.00		0.1	0.07	0.00
	0.25	0.00	0.00		0.25	0.02	0.01
	0.5	-0.01	0.01		0.5	0.03	0.02
	1	-0.03	0.02		1	0.00	0.02
	2	-0.01	0.03		2	0.02	0.03
	3	0.00	0.03		3	-0.01	0.03
	5	-0.04	0.03		5	-0.02	0.03
	10	0.00	0.03		10	0.01	0.03
	20	0.05	0.03		20	0.02	0.03
	30	0.01	0.03		30	0.05	0.03
	50	0.00	0.03		50	0.01	0.03
	100	0.08	0.03		100	0.06	0.03
	300	-0.03	0.03		300	0.00	0.03
	500	-0.02	0.03		500	0.01	0.03
	1000	0.03	0.03		1000	0.03	0.03

Table S5.4A.2: The average distance differences in the donor-acceptor distances of HBs 1-5 following the P3 mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB1-5	s01: P3 V->A				s05: P3 K->A		
HB1	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.01	0.00
	0.1	0.01	0.00		0.1	0.02	0.00
	0.25	0.00	0.00		0.25	0.02	0.01
	0.5	-0.01	0.01		0.5	0.01	0.01
	1	-0.02	0.01		1	0.01	0.02
	2	-0.01	0.02		2	-0.01	0.02
	3	0.00	0.02		3	0.02	0.03
	5	-0.01	0.02		5	0.02	0.03
	10	0.03	0.02		10	-0.03	0.03
	20	0.00	0.02		20	-0.03	0.03
	30	0.03	0.03		30	0.01	0.03
	50	0.04	0.03		50	0.00	0.03
	100	0.04	0.03		100	0.02	0.04
	300	0.08	0.04		300	0.00	0.04
	500	0.02	0.03		500	-0.02	0.04
	1000	0.01	0.04		1000	0.03	0.06
HB2	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.02	0.00
	0.1	0.02	0.00		0.1	0.01	0.00
	0.25	-0.01	0.01		0.25	0.05	0.01
	0.5	-0.03	0.01		0.5	0.02	0.01
	1	-0.02	0.01		1	0.03	0.02
	2	-0.06	0.02		2	0.01	0.02
	3	-0.03	0.02		3	-0.01	0.02
	5	-0.05	0.02		5	-0.02	0.02
	10	-0.01	0.02		10	0.01	0.02
	20	-0.02	0.02		20	-0.02	0.02
	30	-0.05	0.02		30	0.00	0.02
	50	-0.05	0.02		50	0.02	0.02
	100	0.00	0.02		100	0.00	0.02
	300	-0.03	0.02		300	0.00	0.02
	500	-0.02	0.02		500	-0.01	0.02
	1000	-0.06	0.02		1000	0.00	0.02
HB3	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	-0.01	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	-0.01	0.01		0.25	0.02	0.01
	0.5	-0.02	0.01		0.5	0.00	0.01
	1	-0.01	0.01		1	0.00	0.01
	2	-0.04	0.01		2	-0.02	0.02
	3	0.00	0.01		3	-0.01	0.01
	5	-0.05	0.02		5	-0.01	0.02
	10	0.01	0.01		10	0.00	0.01
	20	-0.03	0.01		20	-0.02	0.01
	30	0.00	0.01		30	0.03	0.01
	50	-0.02	0.01		50	0.00	0.02
	100	0.01	0.01		100	0.00	0.02
	300	-0.03	0.01		300	0.02	0.02
	500	-0.02	0.01		500	0.00	0.01
	1000	-0.02	0.01		1000	0.01	0.02
HB4	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.01	0.00
	0.1	-0.01	0.00		0.1	-0.01	0.01
	0.25	-0.01	0.01		0.25	0.02	0.02
	0.5	-0.02	0.01		0.5	0.00	0.02
	1	-0.02	0.02		1	0.02	0.02
	2	0.02	0.03		2	-0.02	0.03
	3	0.03	0.04		3	0.08	0.04
	5	-0.03	0.04		5	0.06	0.05
	10	0.01	0.04		10	-0.01	0.04
	20	0.08	0.05		20	0.01	0.06
	30	0.02	0.06		30	0.06	0.06
	50	-0.03	0.07		50	0.17	0.08
	100	0.03	0.09		100	0.20	0.11
	300	-0.05	0.09		300	0.51	0.14
	500	-0.01	0.13		500	0.74	0.13
	1000	-0.07	0.13		1000	0.63	0.15
HB5	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.01	0.00
	0.25	0.02	0.00		0.25	0.02	0.01
	0.5	0.02	0.01		0.5	0.03	0.01
	1	0.00	0.01		1	0.01	0.02
	2	-0.02	0.03		2	0.08	0.02
	3	0.02	0.03		3	0.03	0.03
	5	0.01	0.03		5	0.04	0.02
	10	0.02	0.03		10	0.03	0.03
	20	0.02	0.03		20	-0.03	0.03
	30	0.03	0.03		30	0.00	0.02
	50	-0.01	0.03		50	0.07	0.03
	100	0.08	0.03		100	0.00	0.03
	300	0.01	0.03		300	0.06	0.03
	500	-0.01	0.03		500	0.01	0.03
	1000	0.09	0.04		1000	0.02	0.03

Table S5.4A.3: The average distance differences in the donor-acceptor distances of HBs 8-12 following the P3 mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB8-12	s01: P3 V->A				s05: P3 K->A		
HB8	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.02	0.00
	0.1	0.00	0.00		0.1	-0.05	0.00
	0.25	-0.01	0.00		0.25	-0.03	0.01
	0.5	-0.01	0.01		0.5	-0.02	0.01
	1	-0.01	0.02		1	-0.01	0.02
	2	0.02	0.03		2	0.01	0.03
	3	0.01	0.03		3	-0.03	0.03
	5	0.04	0.03		5	-0.04	0.02
	10	0.00	0.03		10	-0.03	0.03
	20	-0.04	0.03		20	0.00	0.02
	30	0.00	0.03		30	0.00	0.03
	50	-0.05	0.03		50	-0.01	0.02
	100	-0.05	0.04		100	-0.04	0.03
	300	-0.05	0.03		300	0.04	0.03
	500	-0.08	0.03		500	-0.03	0.03
	1000	0.01	0.03		1000	0.01	0.03
HB9	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.02	0.00
	0.1	0.00	0.00		0.1	0.04	0.00
	0.25	0.00	0.00		0.25	0.01	0.01
	0.5	0.00	0.01		0.5	0.03	0.01
	1	-0.02	0.01		1	0.00	0.01
	2	0.01	0.02		2	0.02	0.02
	3	0.00	0.02		3	0.01	0.02
	5	-0.02	0.01		5	-0.01	0.02
	10	0.01	0.02		10	0.01	0.02
	20	0.01	0.02		20	0.01	0.02
	30	0.01	0.02		30	0.02	0.02
	50	0.00	0.02		50	0.01	0.02
	100	0.04	0.02		100	0.04	0.02
	300	-0.01	0.02		300	0.01	0.02
	500	-0.02	0.02		500	0.00	0.02
	1000	0.01	0.02		1000	0.01	0.02
HB10	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.00	0.00		0.25	0.00	0.00
	0.5	0.01	0.00		0.5	0.02	0.01
	1	0.00	0.01		1	0.02	0.01
	2	0.02	0.02		2	0.01	0.02
	3	0.01	0.02		3	0.03	0.02
	5	-0.01	0.02		5	0.01	0.02
	10	-0.01	0.02		10	0.01	0.02
	20	-0.03	0.02		20	-0.01	0.02
	30	0.00	0.02		30	0.02	0.02
	50	0.00	0.02		50	0.00	0.02
	100	-0.01	0.02		100	-0.01	0.02
	300	0.00	0.02		300	0.05	0.02
	500	-0.03	0.02		500	-0.02	0.02
	1000	-0.01	0.02		1000	0.02	0.02
HB11	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.02	0.00
	0.1	0.00	0.00		0.1	-0.03	0.00
	0.25	0.00	0.00		0.25	-0.01	0.00
	0.5	0.00	0.00		0.5	-0.01	0.01
	1	0.01	0.01		1	-0.01	0.01
	2	0.00	0.02		2	-0.03	0.01
	3	0.01	0.02		3	0.02	0.02
	5	-0.03	0.02		5	0.02	0.02
	10	-0.02	0.02		10	0.01	0.02
	20	0.03	0.03		20	0.00	0.02
	30	-0.02	0.02		30	-0.01	0.02
	50	-0.03	0.02		50	-0.01	0.01
	100	0.00	0.02		100	0.01	0.01
	300	-0.01	0.02		300	-0.01	0.02
	500	-0.01	0.02		500	-0.03	0.02
	1000	0.01	0.02		1000	0.01	0.01
HB12	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.00	0.00		0.25	0.01	0.00
	0.5	0.00	0.00		0.5	0.00	0.01
	1	0.00	0.02		1	-0.01	0.02
	2	0.01	0.03		2	-0.05	0.02
	3	0.03	0.03		3	0.00	0.02
	5	-0.04	0.03		5	0.01	0.02
	10	-0.04	0.03		10	-0.04	0.03
	20	-0.03	0.03		20	0.00	0.02
	30	-0.02	0.03		30	0.00	0.03
	50	-0.05	0.03		50	0.03	0.03
	100	0.02	0.04		100	-0.02	0.03
	300	-0.03	0.05		300	0.00	0.03
	500	0.00	0.04		500	-0.03	0.03
	1000	-0.05	0.05		1000	0.03	0.04

Table S5.4A.4: The ten most perturbed Cα atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P3 mutation in s01 and s05. Only displacements ≥0.01 Å are shown.

s01 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	4ALA_CA	C	0.03	0.00
	2	3ALA_CA	C	0.02	0.00
	3	2SER_CA	C	0.01	0.00
	4	5LEU_CA	C	0.01	0.00
	5				
	6				
	7				
	8				
	9				
	10				
0.5	1	4ALA_CA	C	0.08	0.01
	2	5LEU_CA	C	0.05	0.01
	3	3ALA_CA	C	0.03	0.01
	4	45THR_CA	A	0.02	0.00
	5	46SER_CA	A	0.02	0.00
	6	138GLY_CA	A	0.02	0.00
	7	167LEU_CA	A	0.02	0.01
	8	168PRO_CA	A	0.02	0.00
	9	169THR_CA	A	0.02	0.00
	10	170GLY_CA	A	0.02	0.00
5	1	94ALA_CA	A	0.15	0.04
	2	93THR_CA	A	0.14	0.04
	3	301SER_CA	B	0.12	0.03
	4	4ALA_CA	C	0.12	0.03
	5	24THR_CA	B	0.11	0.05
	6	283GLY_CA	B	0.11	0.04
	7	293PRO_CA	B	0.11	0.03
	8	300CYS_CA	B	0.11	0.03
	9	195GLY_CA	A	0.10	0.04
	10	283GLY_CA	A	0.10	0.04
50	1	72ASN_CA	B	0.26	0.09
	2	255ALA_CA	B	0.22	0.06
	3	1THR_CA	C	0.22	0.09
	4	191ALA_CA	B	0.21	0.08
	5	258GLY_CA	B	0.20	0.06
	6	194ALA_CA	B	0.19	0.07
	7	195GLY_CA	B	0.19	0.06
	8	196THR_CA	B	0.19	0.06
	9	246HIE_CA	B	0.19	0.05
	10	248ASP_CA	B	0.19	0.06
500	1	72ASN_CA	A	0.23	0.10
	2	47GLU_CA	A	0.20	0.08
	3	303VAL_CA	B	0.20	0.06
	4	72ASN_CA	B	0.19	0.10
	5	73VAL_CA	A	0.18	0.07
	6	92ASP_CA	A	0.17	0.08
	7	260ALA_CA	A	0.17	0.07
	8	251GLY_CA	B	0.17	0.06
	9	274ASN_CA	B	0.17	0.08
	10	275GLY_CA	B	0.17	0.07
1000	1	277ASN_CA	A	0.27	0.12
	2	277ASN_CA	B	0.25	0.09
	3	256GLN_CA	B	0.23	0.07
	4	255ALA_CA	B	0.22	0.07
	5	51ASN_CA	A	0.20	0.08
	6	52PRO_CA	A	0.20	0.08
	7	55GLU_CA	A	0.19	0.07
	8	283GLY_CA	A	0.19	0.06
	9	50LEU_CA	B	0.19	0.09
	10	224THR_CA	B	0.19	0.09

s05 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	4ALA_CA	C	0.03	0.00
	2	46SER_CA	A	0.01	0.00
	3	141LEU_CA	A	0.01	0.00
	4	143GLY_CA	A	0.01	0.00
	5	170GLY_CA	A	0.01	0.00
	6	188ARG_CA	A	0.01	0.00
	7	189GLN_CA	A	0.01	0.00
	8	190THR_CA	A	0.01	0.00
	9	1SER_CA	B	0.01	0.00
	10	2GLY_CA	B	0.01	0.00
0.5	1	7ASN_CA	C	0.06	0.01
	2	8ASN_CA	C	0.05	0.01
	3	9GLU_CA	C	0.05	0.01
	4	10LEU_CA	C	0.05	0.01
	5	45THR_CA	A	0.04	0.01
	6	170GLY_CA	A	0.04	0.01
	7	5LEU_CA	C	0.04	0.01
	8	23GLY_CA	A	0.03	0.01
	9	24THR_CA	A	0.03	0.01
	10	44CYS_CA	A	0.03	0.01
5	1	50LEU_CA	B	0.18	0.04
	2	46SER_CA	B	0.16	0.05
	3	51ASN_CA	B	0.16	0.04
	4	47GLU_CA	B	0.15	0.05
	5	24THR_CA	B	0.14	0.05
	6	189GLN_CA	B	0.14	0.04
	7	71GLY_CA	A	0.13	0.06
	8	45THR_CA	B	0.13	0.04
	9	49MET_CA	B	0.13	0.04
	10	64HIE_CA	B	0.13	0.05
50	1	11SER_CA	C	0.26	0.11
	2	251GLY_CA	A	0.25	0.06
	3	252PRO_CA	A	0.25	0.06
	4	253LEU_CA	A	0.22	0.05
	5	24THR_CA	B	0.22	0.06
	6	236LYS_CA	B	0.20	0.07
	7	232LEU_CA	A	0.19	0.06
	8	23GLY_CA	B	0.19	0.05
	9	275GLY_CA	B	0.19	0.07
	10	276MET_CA	B	0.19	0.07
500	1	72ASN_CA	A	0.36	0.11
	2	11SER_CA	C	0.30	0.16
	3	55GLU_CA	A	0.29	0.07
	4	93THR_CA	B	0.26	0.09
	5	277ASN_CA	B	0.25	0.11
	6	73VAL_CA	A	0.23	0.09
	7	278GLY_CA	A	0.23	0.09
	8	193ALA_CA	B	0.22	0.07
	9	194ALA_CA	B	0.22	0.07
	10	195GLY_CA	B	0.22	0.08
1000	1	11SER_CA	C	0.84	0.18
	2	10LEU_CA	C	0.48	0.11
	3	278GLY_CA	A	0.33	0.12
	4	9GLU_CA	C	0.31	0.07
	5	55GLU_CA	B	0.29	0.09
	6	56ASP_CA	B	0.29	0.11
	7	306GLN_CA	A	0.28	0.10
	8	46SER_CA	A	0.23	0.07
	9	59ILE_CA	B	0.23	0.10
	10	277ASN_CA	A	0.22	0.11

Table S5.4A.5: The ten most perturbed non-hydrogen atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P3 mutation in s01 and s05.

s01 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	4ALA_CB	C	0.09	0.01
	2	4ALA_N	C	0.04	0.00
	3	3ALA_CB	C	0.03	0.00
	4	4ALA_O	C	0.03	0.00
	5	3ALA_C	C	0.03	0.00
	6	4ALA_CA	C	0.03	0.00
	7	3ALA_CA	C	0.02	0.00
	8	166GLU_CB	A	0.02	0.00
	9	5LEU_CA	C	0.01	0.00
	10	3ALA_O	C	0.01	0.00
0.5	1	4ALA_CB	C	0.18	0.02
	2	189GLN_NE2	A	0.09	0.02
	3	4ALA_CA	C	0.08	0.01
	4	5LEU_O	C	0.07	0.02
	5	3ALA_O	C	0.07	0.02
	6	166GLU_OE2	A	0.07	0.01
	7	5LEU_CB	C	0.07	0.01
	8	5LEU_N	C	0.06	0.01
	9	189GLN_CD	A	0.06	0.01
	10	189GLN_OE1	A	0.06	0.02
5	1	232LEU_CD2	A	0.27	0.08
	2	232LEU_CD1	A	0.26	0.07
	3	10ARG_NH2	C	0.23	0.07
	4	76ARG_NH2	A	0.23	0.08
	5	222ARG_NH1	B	0.22	0.10
	6	53ASN_OD1	A	0.21	0.08
	7	236LYS_NZ	A	0.21	0.11
	8	92ASP_OD2	A	0.21	0.06
	9	4ALA_CB	C	0.21	0.04
	10	232LEU_CG	A	0.20	0.06
50	1	72ASN_ND2	B	0.47	0.14
	2	11LYS_NZ	C	0.40	0.18
	3	222ARG_NH2	A	0.39	0.19
	4	55GLU_OE1	B	0.38	0.14
	5	60ARG_NH1	A	0.37	0.12
	6	1THR_OG1	C	0.36	0.14
	7	248ASP_OD1	B	0.36	0.09
	8	72ASN_OD1	A	0.35	0.14
	9	222ARG_NH2	B	0.35	0.16
	10	72ASN_CG	B	0.35	0.12
500	1	244GLN_OE1	B	0.60	0.16
	2	64HIE_NE2	B	0.52	0.15
	3	222ARG_NH1	B	0.49	0.25
	4	97LYS_NZ	B	0.48	0.14
	5	72ASN_OD1	B	0.47	0.21
	6	244GLN_NE2	A	0.47	0.21
	7	9PHE_CZ	C	0.47	0.16
	8	72ASN_ND2	A	0.46	0.20
	9	47GLU_OE2	A	0.45	0.21
	10	64HIE_CE1	B	0.44	0.15
1000	1	277ASN_OD1	A	0.62	0.20
	2	222ARG_NH2	B	0.58	0.25
	3	244GLN_NE2	B	0.56	0.23
	4	142ASN_ND2	B	0.53	0.14
	5	279ARG_NH1	B	0.52	0.20
	6	55GLU_OE1	B	0.52	0.16
	7	222ARG_NH1	B	0.51	0.22
	8	222ARG_CZ	B	0.50	0.21
	9	306GLN_NE2	A	0.49	0.22
	10	55GLU_OE2	B	0.48	0.16

s05 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	7ASN_O	C	0.05	0.00
	2	142ASN_OD1	A	0.05	0.00
	3	5LEU_O	C	0.05	0.00
	4	7ASN_OD1	C	0.05	0.00
	5	4ALA_O	C	0.04	0.00
	6	189GLN_OE1	A	0.03	0.00
	7	4ALA_N	C	0.03	0.01
	8	4ALA_CB	C	0.03	0.01
	9	166GLU_OE1	A	0.03	0.00
	10	9GLU_OE1	C	0.03	0.00
0.5	1	142ASN_ND2	A	0.21	0.04
	2	9GLU_OE1	C	0.18	0.03
	3	9GLU_OE2	C	0.17	0.03
	4	142ASN_OD1	A	0.17	0.03
	5	9GLU_CD	C	0.15	0.02
	6	10LEU_O	C	0.10	0.02
	7	47GLU_OE2	A	0.10	0.02
	8	47GLU_OE1	A	0.09	0.02
	9	9GLU_CG	C	0.09	0.02
	10	142ASN_CG	A	0.09	0.02
5	1	142ASN_ND2	A	0.33	0.07
	2	279ARG_NH1	B	0.30	0.07
	3	107GLN_NE2	B	0.28	0.08
	4	189GLN_NE2	B	0.27	0.07
	5	279ARG_CZ	B	0.26	0.06
	6	107GLN_NE2	A	0.26	0.08
	7	279ARG_NH2	B	0.26	0.07
	8	24THR_CG2	B	0.25	0.09
	9	47GLU_OE1	B	0.25	0.10
	10	244GLN_NE2	B	0.24	0.09
50	1	142ASN_OD1	A	0.57	0.08
	2	236LYS_NZ	B	0.48	0.19
	3	142ASN_ND2	A	0.44	0.08
	4	236LYS_CE	B	0.44	0.16
	5	244GLN_NE2	B	0.43	0.14
	6	11SER_OC2	C	0.43	0.16
	7	279ARG_NH2	B	0.41	0.12
	8	279ARG_NH1	A	0.41	0.12
	9	306GLN_OE1	A	0.39	0.12
	10	236LYS_CD	B	0.35	0.13
500	1	189GLN_OE1	A	0.75	0.11
	2	256GLN_NE2	A	0.68	0.14
	3	72ASN_ND2	A	0.68	0.24
	4	55GLU_OE2	A	0.59	0.17
	5	76ARG_NH1	B	0.56	0.17
	6	11SER_OC1	C	0.55	0.24
	7	46SER_OG	A	0.55	0.12
	8	72ASN_CG	A	0.54	0.20
	9	9GLU_OE1	C	0.53	0.14
	10	47GLU_OE2	A	0.52	0.19
1000	1	11SER_OG	C	1.14	0.24
	2	11SER_CB	C	1.08	0.23
	3	11SER_OC2	C	0.95	0.25
	4	11SER_CA	C	0.84	0.19
	5	9GLU_OE2	C	0.81	0.14
	6	11SER_C	C	0.81	0.22
	7	11SER_N	C	0.71	0.15
	8	47GLU_OE1	A	0.68	0.19
	9	189GLN_OE1	A	0.68	0.12
	10	11SER_OC1	C	0.66	0.25

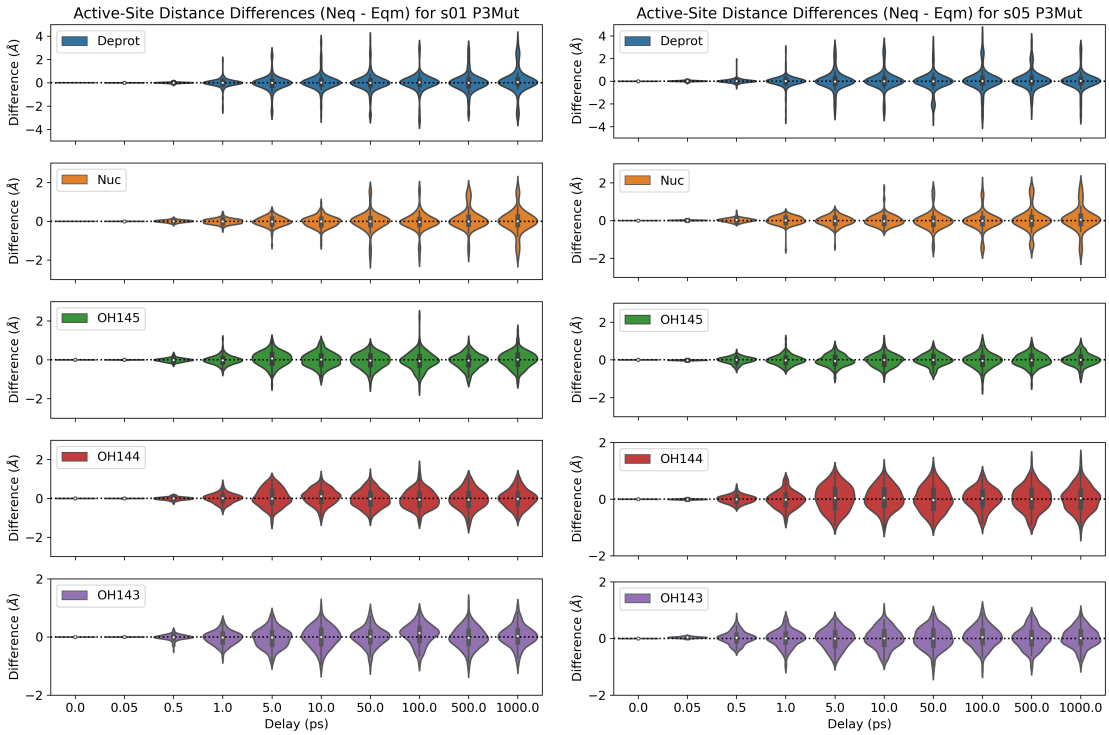


Figure S5.4A.1: Distributions of the distance differences in the five active-site distances ($\Delta d = d_{\text{neq}} - d_{\text{eqm}}$) following the P3 mutation in s01 and s05.

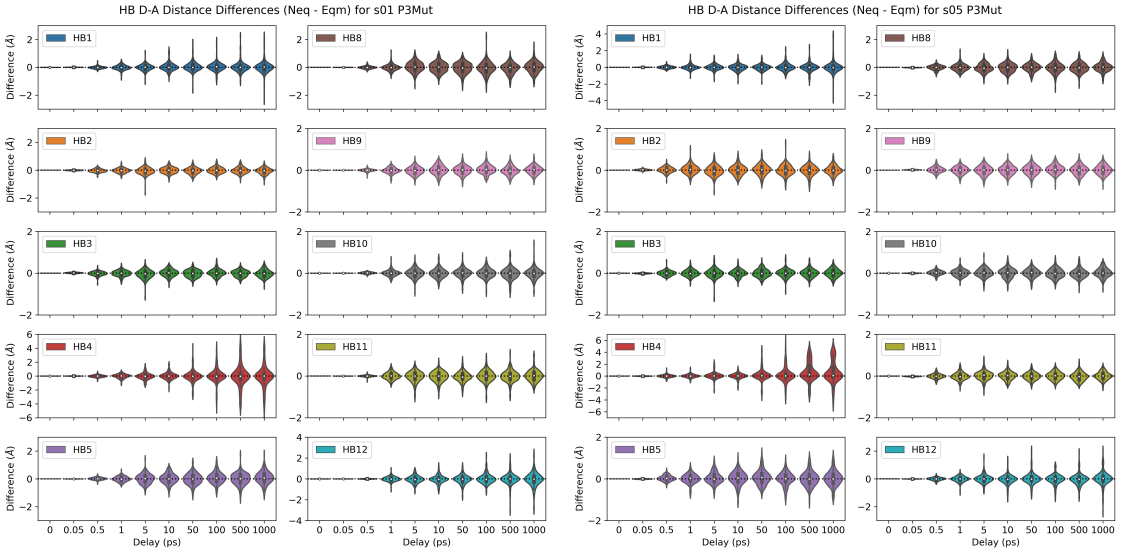


Figure S5.4A.2: Distributions of the distance differences in the ten donor-acceptor distances of HBs 1-5 and 8-12 following the P3 mutation in s01 and s05.

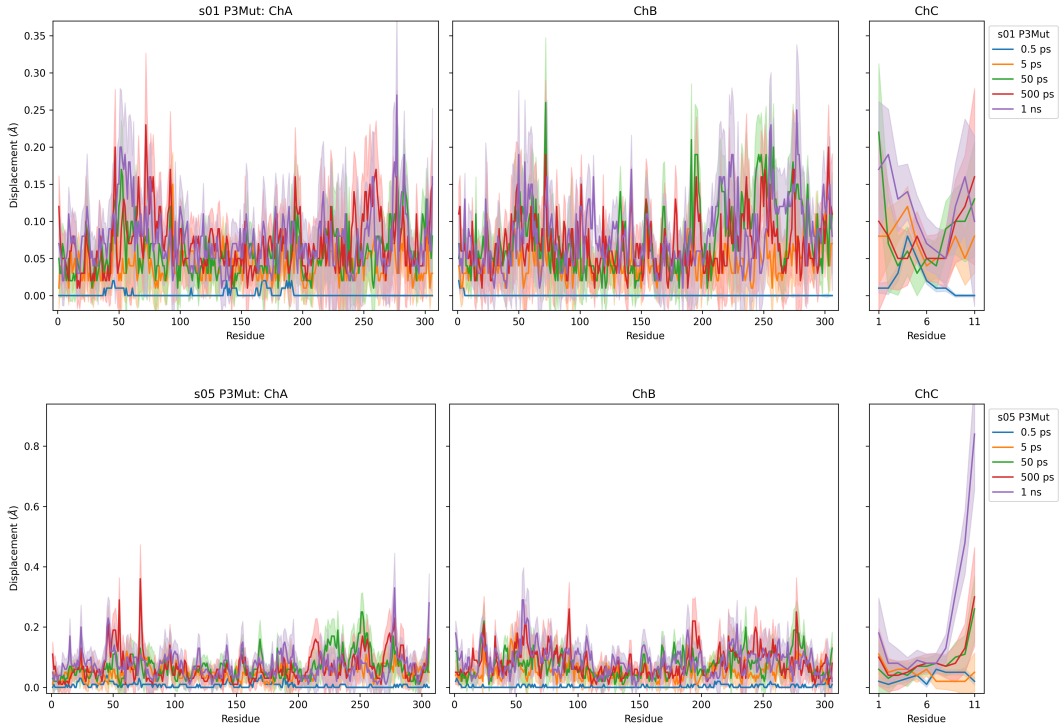


Figure S5.4A.3: Magnitudes of the average M^{pro} Cα displacement vectors, with ± SEM shown as shaded regions, following the P3 mutation in s01 and s05.

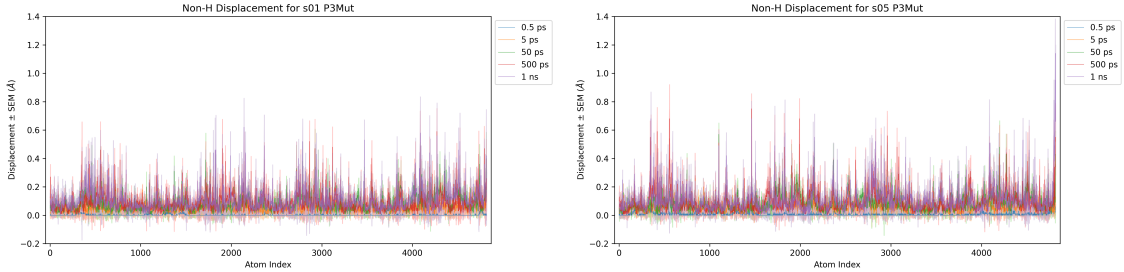


Figure S5.4A.4: Magnitudes of the average non-hydrogen atom displacement vectors in the M^{pro}-peptide complex, with ± SEM shown as shaded regions, following the P3 mutation in s01 and s05.

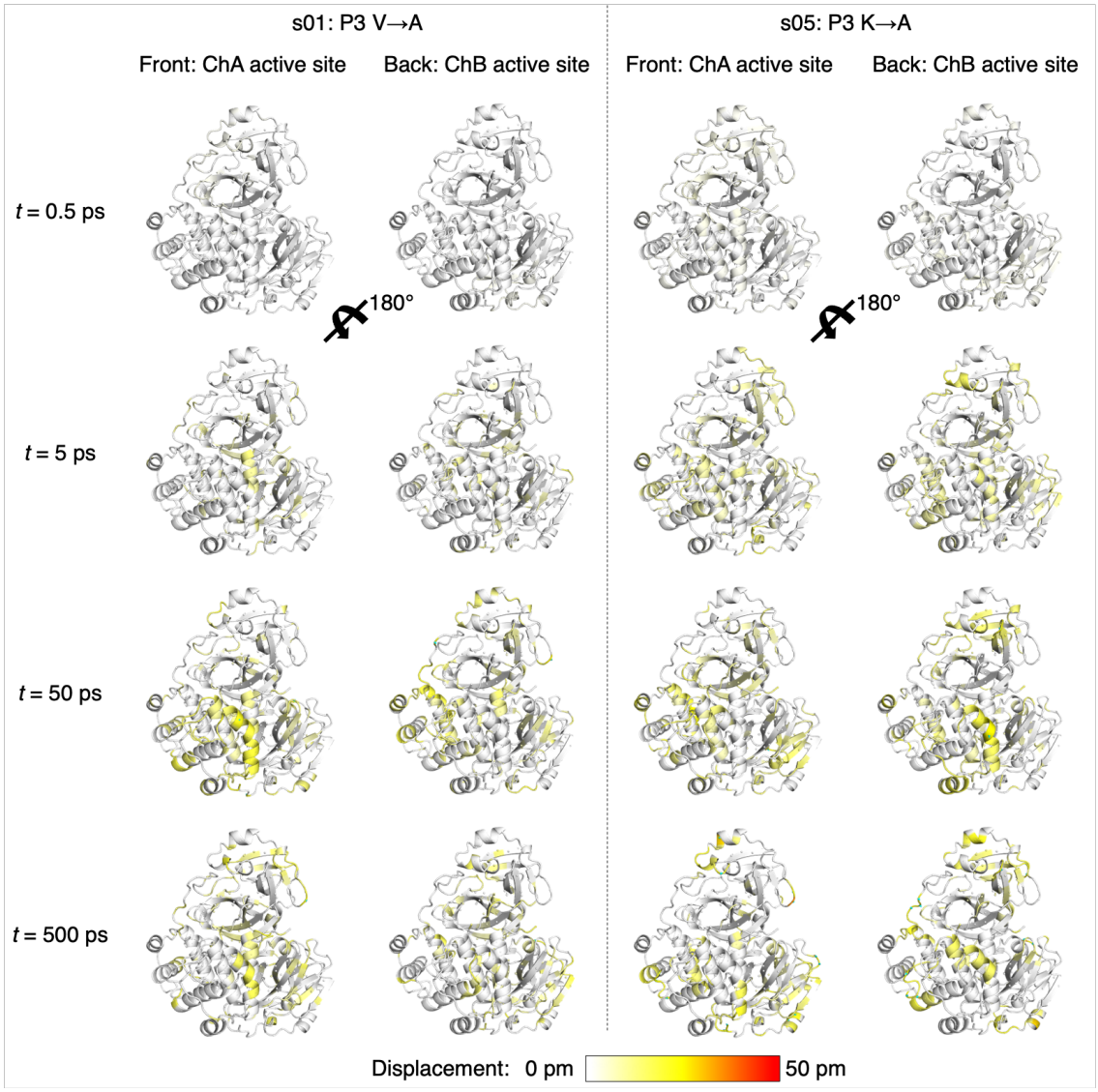


Figure S5.4A.5: Views of the M^{pro} response to the P3 mutation in s01 and s05 from averaging Cα displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as cyan arrows with a scale-up factor of 5.[7]

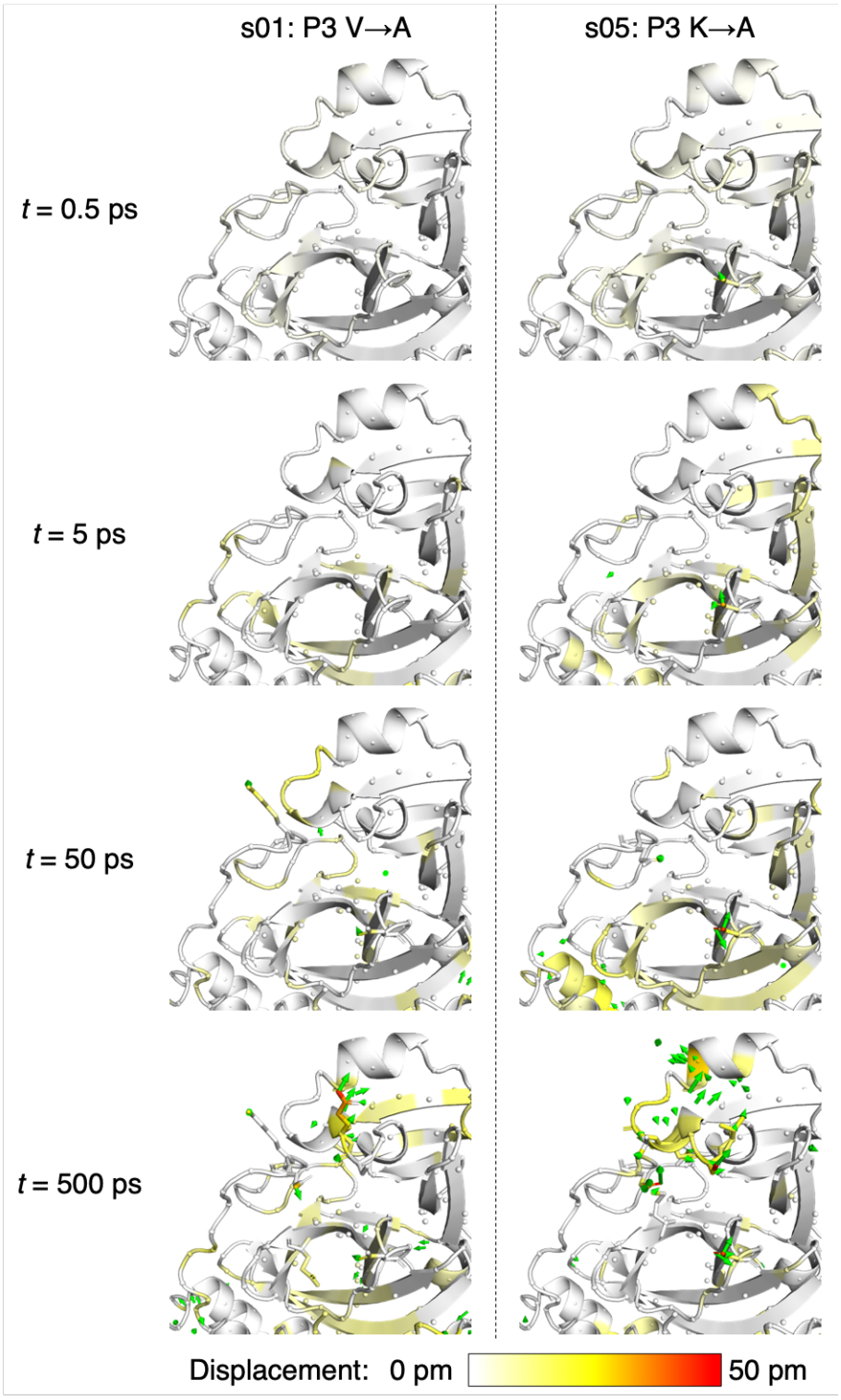


Figure S5.4A.6: Views of the M^{pro} response to the P3 mutation in s01 and s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations, with a focus on the S3 subsite. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7] Residues that show such significant displacements and that are within 10 Å of the substrate P3 residue are shown as sticks.

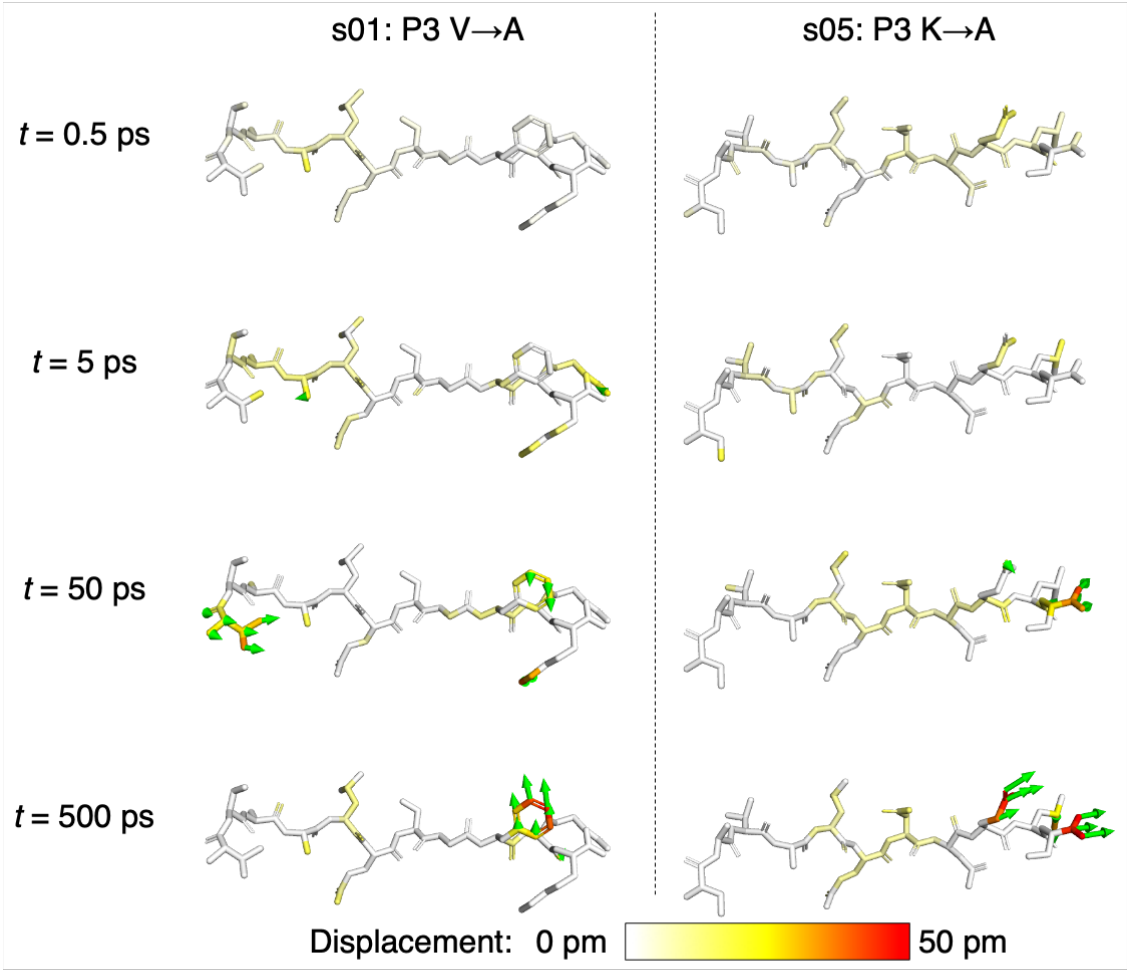


Figure S5.4A.7: Views of the peptide response to the P3 mutation in s01 and s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7]

Section S5.4B: Effect of P4 mutation

Table S5.4B.1: The average distance differences in the five active-site distances ($\Delta d = d_{neq} - d_{eqm}$) following the P4 mutation in s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

Active Site	s05: P4 V->A		
Deprot	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.01	0.01
	1	0.02	0.03
	2	0.00	0.05
	3	0.05	0.07
	5	0.01	0.07
	10	-0.01	0.06
	20	-0.05	0.07
	30	0.03	0.06
	50	0.00	0.07
	100	-0.03	0.07
	300	0.06	0.07
	500	0.02	0.07
	1000	0.12	0.08
Nuc	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.01	0.00
	0.5	0.00	0.00
	1	0.00	0.01
	2	0.01	0.02
	3	0.03	0.02
	5	-0.02	0.02
	10	-0.05	0.03
	20	-0.04	0.03
	30	0.01	0.03
	50	-0.02	0.04
	100	0.01	0.04
	300	-0.05	0.05
	500	-0.03	0.05
	1000	-0.01	0.05
OH145	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.01	0.00
	0.5	0.01	0.00
	1	0.00	0.01
	2	-0.01	0.02
	3	-0.01	0.03
	5	0.00	0.03
	10	-0.02	0.03
	20	0.01	0.02
	30	0.01	0.03
	50	-0.01	0.02
	100	-0.05	0.03
	300	0.01	0.03
	500	0.04	0.03
	1000	0.02	0.03
OH144	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	-0.01	0.02
	2	-0.05	0.03
	3	-0.01	0.03
	5	0.03	0.03
	10	0.02	0.03
	20	-0.02	0.03
	30	0.03	0.03
	50	-0.02	0.04
	100	-0.01	0.04
	300	-0.02	0.04
	500	-0.02	0.04
	1000	-0.05	0.04
OH143	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	-0.01	0.00
	1	0.01	0.02
	2	0.03	0.03
	3	-0.02	0.03
	5	0.00	0.03
	10	0.02	0.03
	20	-0.04	0.03
	30	0.00	0.03
	50	-0.01	0.03
	100	0.02	0.03
	300	0.00	0.03
	500	-0.02	0.03
	1000	-0.03	0.03

Table S5.4B.2: The average distance differences in the donor-acceptor distances of HBs 1-5 following the P4 mutation in s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB1-5	s05: P4 V->A		
HB1	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	-0.01	0.00
	0.1	-0.01	0.01
	0.25	-0.06	0.01
	0.5	-0.11	0.02
	1	-0.13	0.03
	2	-0.10	0.04
	3	-0.09	0.03
	5	-0.11	0.04
	10	-0.12	0.03
	20	-0.15	0.04
	30	-0.14	0.04
	50	-0.14	0.04
	100	-0.12	0.04
	300	-0.14	0.06
	500	-0.19	0.06
	1000	-0.10	0.07
HB2	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	-0.02	0.00
	0.25	-0.05	0.01
	0.5	-0.06	0.01
	1	-0.03	0.01
	2	-0.06	0.02
	3	-0.04	0.02
	5	-0.08	0.02
	10	-0.06	0.02
	20	-0.05	0.02
	30	-0.06	0.02
	50	-0.02	0.02
	100	-0.04	0.02
	300	-0.05	0.02
	500	-0.04	0.02
	1000	-0.06	0.02
HB3	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	-0.02	0.01
	1	-0.01	0.01
	2	-0.01	0.01
	3	0.00	0.01
	5	-0.04	0.02
	10	-0.02	0.01
	20	-0.02	0.01
	30	0.01	0.01
	50	0.00	0.01
	100	-0.02	0.01
	300	0.00	0.01
	500	-0.01	0.01
	1000	-0.01	0.02
HB4	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.01	0.00
	0.5	0.01	0.01
	1	-0.01	0.02
	2	-0.06	0.03
	3	0.00	0.03
	5	0.01	0.04
	10	-0.07	0.04
	20	-0.08	0.05
	30	-0.08	0.06
	50	-0.15	0.06
	100	-0.09	0.08
	300	-0.13	0.09
	500	0.00	0.10
	1000	-0.09	0.14
HB5	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.01	0.00
	0.5	0.02	0.01
	1	-0.02	0.01
	2	0.04	0.02
	3	0.03	0.02
	5	0.07	0.02
	10	0.00	0.02
	20	0.02	0.03
	30	0.01	0.03
	50	0.02	0.03
	100	-0.01	0.03
	300	0.08	0.03
	500	0.03	0.03
	1000	0.03	0.03

Table S5.4B.3: The average distance differences in the donor-acceptor distances of HBs 8-12 following the P4 mutation in s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB8-12	s05: P4 V->A		
HB8	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.01	0.00
	0.5	0.01	0.00
	1	0.00	0.01
	2	-0.01	0.02
	3	-0.01	0.02
	5	0.00	0.03
	10	-0.02	0.03
	20	0.02	0.02
	30	0.00	0.03
	50	-0.01	0.02
	100	-0.06	0.03
	300	0.02	0.03
	500	0.03	0.03
	1000	0.02	0.03
HB9	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	0.01	0.01
	2	0.00	0.02
	3	-0.02	0.02
	5	0.00	0.02
	10	0.01	0.02
	20	-0.01	0.01
	30	0.00	0.02
	50	0.00	0.02
	100	0.01	0.02
	300	-0.01	0.02
	500	-0.02	0.02
	1000	-0.02	0.02
HB10	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	0.02	0.01
	2	0.00	0.02
	3	0.03	0.02
	5	0.02	0.02
	10	-0.01	0.02
	20	0.00	0.02
	30	0.00	0.02
	50	0.01	0.02
	100	-0.01	0.02
	300	0.01	0.02
	500	-0.03	0.02
	1000	0.05	0.02
HB11	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	0.01	0.01
	2	-0.02	0.02
	3	0.01	0.02
	5	-0.01	0.02
	10	0.01	0.01
	20	0.03	0.02
	30	0.00	0.02
	50	0.01	0.02
	100	-0.02	0.02
	300	0.01	0.02
	500	0.01	0.02
	1000	0.03	0.02
HB12	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	0.00	0.01
	2	-0.04	0.02
	3	-0.01	0.02
	5	-0.05	0.02
	10	0.00	0.02
	20	0.02	0.03
	30	0.01	0.03
	50	0.01	0.03
	100	-0.01	0.03
	300	0.01	0.03
	500	-0.06	0.04
	1000	0.02	0.04

Table S5.4B.4: The ten most perturbed Cα atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P4 mutation in s05. Only displacements ≥0.01 Å are shown.

s05 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	3ALA_CA	C	0.03	0.00
	2	2ALA_CA	C	0.02	0.00
	3	167LEU_CA	A	0.01	0.00
	4	188ARG_CA	A	0.01	0.00
	5	189GLN_CA	A	0.01	0.00
	6	190THR_CA	A	0.01	0.00
	7	192GLN_CA	A	0.01	0.00
	8	1SER_CA	C	0.01	0.00
	9	4LYS_CA	C	0.01	0.00
	10				
0.5	1	4LYS_CA	C	0.16	0.01
	2	192GLN_CA	A	0.15	0.01
	3	3ALA_CA	C	0.15	0.01
	4	188ARG_CA	A	0.10	0.01
	5	167LEU_CA	A	0.08	0.01
	6	171VAL_CA	A	0.08	0.01
	7	189GLN_CA	A	0.08	0.01
	8	190THR_CA	A	0.08	0.01
	9	193ALA_CA	A	0.07	0.01
	10	186VAL_CA	A	0.06	0.01
5	1	192GLN_CA	A	0.28	0.04
	2	193ALA_CA	A	0.26	0.06
	3	188ARG_CA	A	0.18	0.03
	4	190THR_CA	A	0.18	0.04
	5	191ALA_CA	A	0.18	0.04
	6	189GLN_CA	A	0.17	0.03
	7	1SER_CA	C	0.17	0.07
	8	186VAL_CA	A	0.16	0.03
	9	2ALA_CA	C	0.16	0.04
	10	3ALA_CA	C	0.16	0.03
50	1	193ALA_CA	A	0.69	0.07
	2	194ALA_CA	A	0.58	0.07
	3	192GLN_CA	A	0.50	0.05
	4	195GLY_CA	A	0.48	0.07
	5	1SER_CA	C	0.42	0.11
	6	191ALA_CA	A	0.30	0.05
	7	196THR_CA	A	0.28	0.07
	8	2ALA_CA	C	0.28	0.06
	9	11SER_CA	C	0.27	0.13
	10	72ASN_CA	A	0.26	0.09
500	1	194ALA_CA	A	0.88	0.09
	2	193ALA_CA	A	0.87	0.09
	3	195GLY_CA	A	0.63	0.09
	4	192GLN_CA	A	0.42	0.05
	5	168PRO_CA	A	0.30	0.06
	6	1SER_CA	C	0.29	0.10
	7	11SER_CA	C	0.28	0.19
	8	189GLN_CA	A	0.27	0.04
	9	2ALA_CA	C	0.27	0.08
	10	191ALA_CA	B	0.26	0.11
1000	1	193ALA_CA	A	0.91	0.08
	2	194ALA_CA	A	0.79	0.08
	3	195GLY_CA	A	0.62	0.09
	4	11SER_CA	C	0.53	0.20
	5	192GLN_CA	A	0.51	0.06
	6	1SER_CA	C	0.44	0.15
	7	196THR_CA	A	0.39	0.08
	8	2ALA_CA	C	0.37	0.09
	9	168PRO_CA	A	0.32	0.06
	10	191ALA_CA	A	0.31	0.06

Table S5.4B.5: The ten most perturbed non-hydrogen atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P4 mutation in s05.

s05 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	3ALA_CB	C	0.11	0.01
	2	192GLN_CB	A	0.05	0.00
	3	3ALA_N	C	0.04	0.00
	4	3ALA_CA	C	0.03	0.00
	5	4LYS_N	C	0.02	0.00
	6	2ALA_O	C	0.02	0.00
	7	167LEU_CD2	A	0.02	0.00
	8	190THR_O	A	0.02	0.00
	9	188ARG_O	A	0.02	0.00
	10	165MET_CE	A	0.02	0.00
0.5	1	167LEU_CD1	A	0.33	0.04
	2	167LEU_CD2	A	0.32	0.03
	3	165MET_CE	A	0.29	0.04
	4	167LEU_CG	A	0.28	0.02
	5	165MET_SD	A	0.27	0.03
	6	3ALA_CB	C	0.23	0.02
	7	3ALA_O	C	0.22	0.02
	8	192GLN_CB	A	0.21	0.01
	9	4LYS_N	C	0.20	0.01
	10	192GLN_NE2	A	0.20	0.02
5	1	165MET_CE	A	0.54	0.08
	2	193ALA_CB	A	0.33	0.07
	3	167LEU_CD2	A	0.32	0.05
	4	192GLN_CB	A	0.31	0.04
	5	3ALA_CB	C	0.29	0.04
	6	192GLN_CA	A	0.28	0.04
	7	192GLN_C	A	0.28	0.05
	8	193ALA_N	A	0.28	0.05
	9	192GLN_O	A	0.27	0.05
	10	193ALA_CA	A	0.26	0.06
50	1	167LEU_CD2	A	0.87	0.08
	2	193ALA_CB	A	0.83	0.10
	3	193ALA_CA	A	0.69	0.07
	4	194ALA_N	A	0.64	0.07
	5	193ALA_N	A	0.63	0.06
	6	193ALA_C	A	0.62	0.07
	7	192GLN_O	A	0.60	0.07
	8	165MET_CE	A	0.59	0.09
	9	191ALA_O	A	0.59	0.08
	10	194ALA_CA	A	0.58	0.07
500	1	193ALA_CB	A	1.13	0.11
	2	167LEU_CD2	A	1.11	0.08
	3	194ALA_N	A	0.90	0.09
	4	194ALA_CB	A	0.89	0.10
	5	194ALA_CA	A	0.88	0.09
	6	193ALA_CA	A	0.87	0.09
	7	193ALA_C	A	0.82	0.08
	8	194ALA_O	A	0.80	0.09
	9	194ALA_C	A	0.79	0.08
	10	193ALA_O	A	0.76	0.08
1000	1	193ALA_CB	A	1.13	0.11
	2	167LEU_CD2	A	1.11	0.09
	3	11SER_OG	C	0.92	0.26
	4	193ALA_CA	A	0.91	0.08
	5	194ALA_N	A	0.86	0.08
	6	222ARG_NH2	B	0.85	0.33
	7	196THR_OG1	A	0.81	0.14
	8	194ALA_O	A	0.81	0.09
	9	193ALA_C	A	0.80	0.08
	10	194ALA_CA	A	0.79	0.08

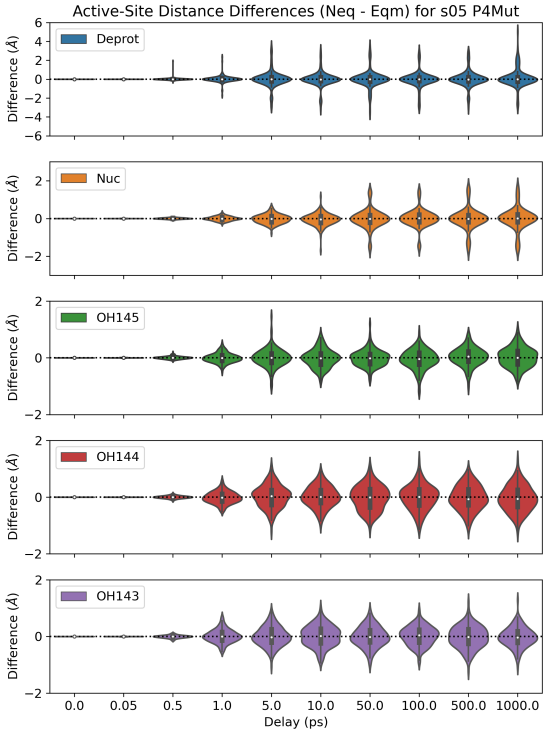


Figure S5.4B.1: Distributions of the distance differences in the five active-site distances ($\Delta d = d_{\mathrm{neq}} - d_{\mathrm{eqm}}$) following the P4 mutation in s05.

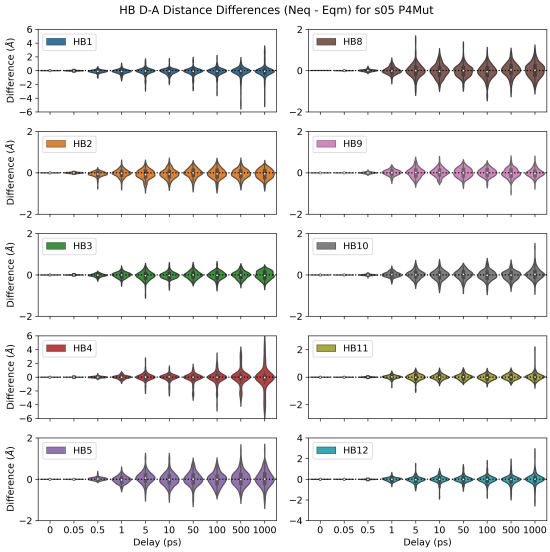


Figure S5.4B.2: Distributions of the distance differences in the ten donor-acceptor distances of HBs 1-5 and 8-12 following the P4 mutation in s05.

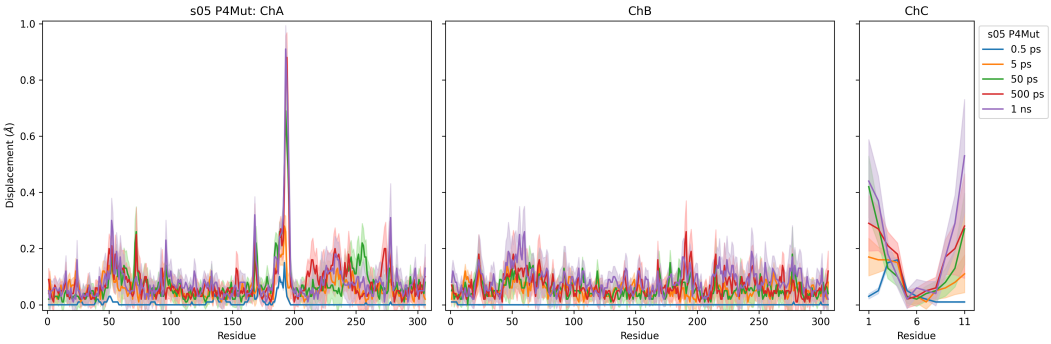


Figure S5.4B.3: Magnitudes of the average M^{pro} Cα displacement vectors, with ± SEM shown as shaded regions, following the P4 mutation in s05.

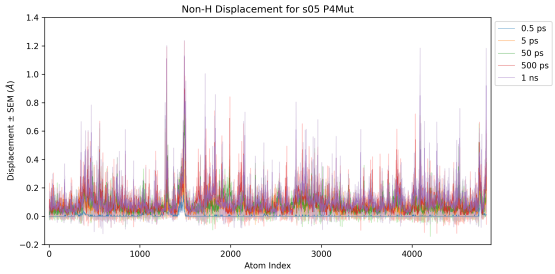


Figure S5.4B.4: Magnitudes of the average non-hydrogen atom displacement vectors in the M^{pro}-peptide complex, with ± SEM shown as shaded regions, following the P4 mutation in s05.

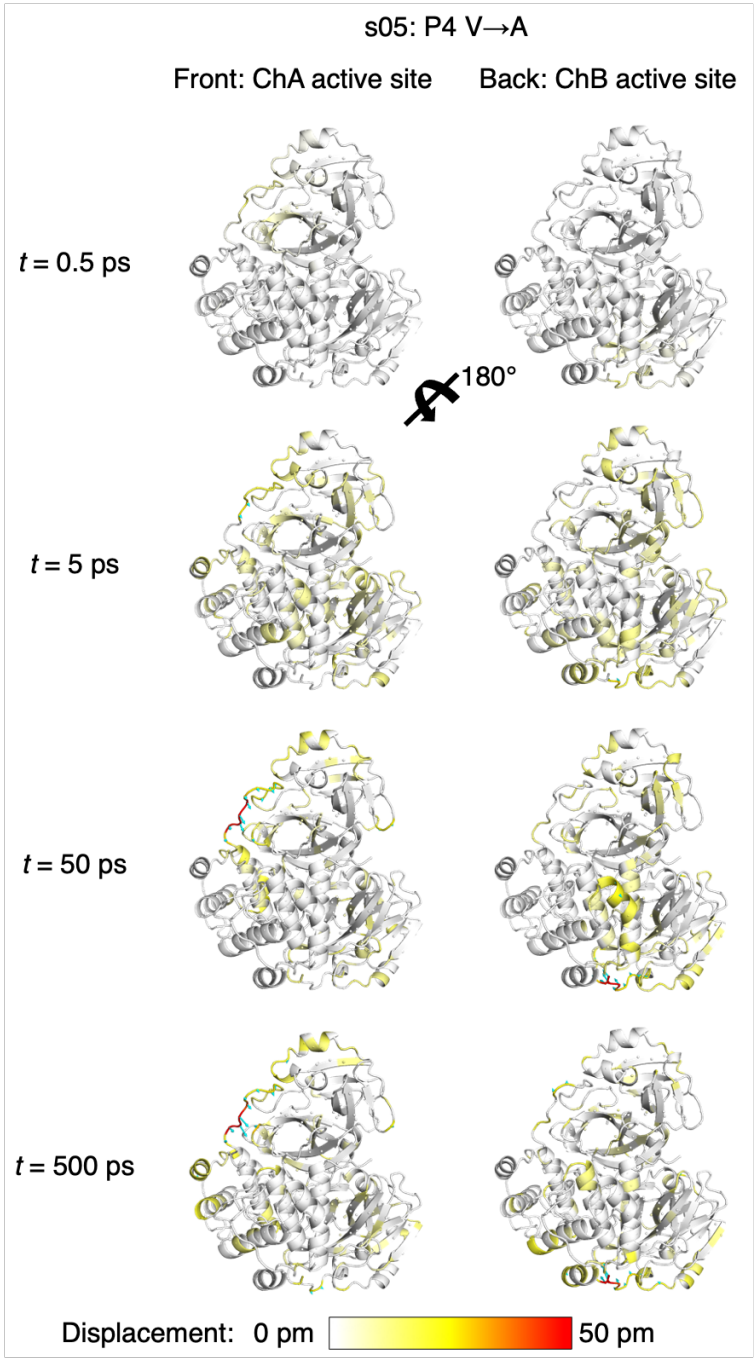


Figure S5.4B.5: Views of the M^{pro} response to the P4 mutation in s05 from averaging Cα displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as cyan arrows with a scale-up factor of 5.[7]

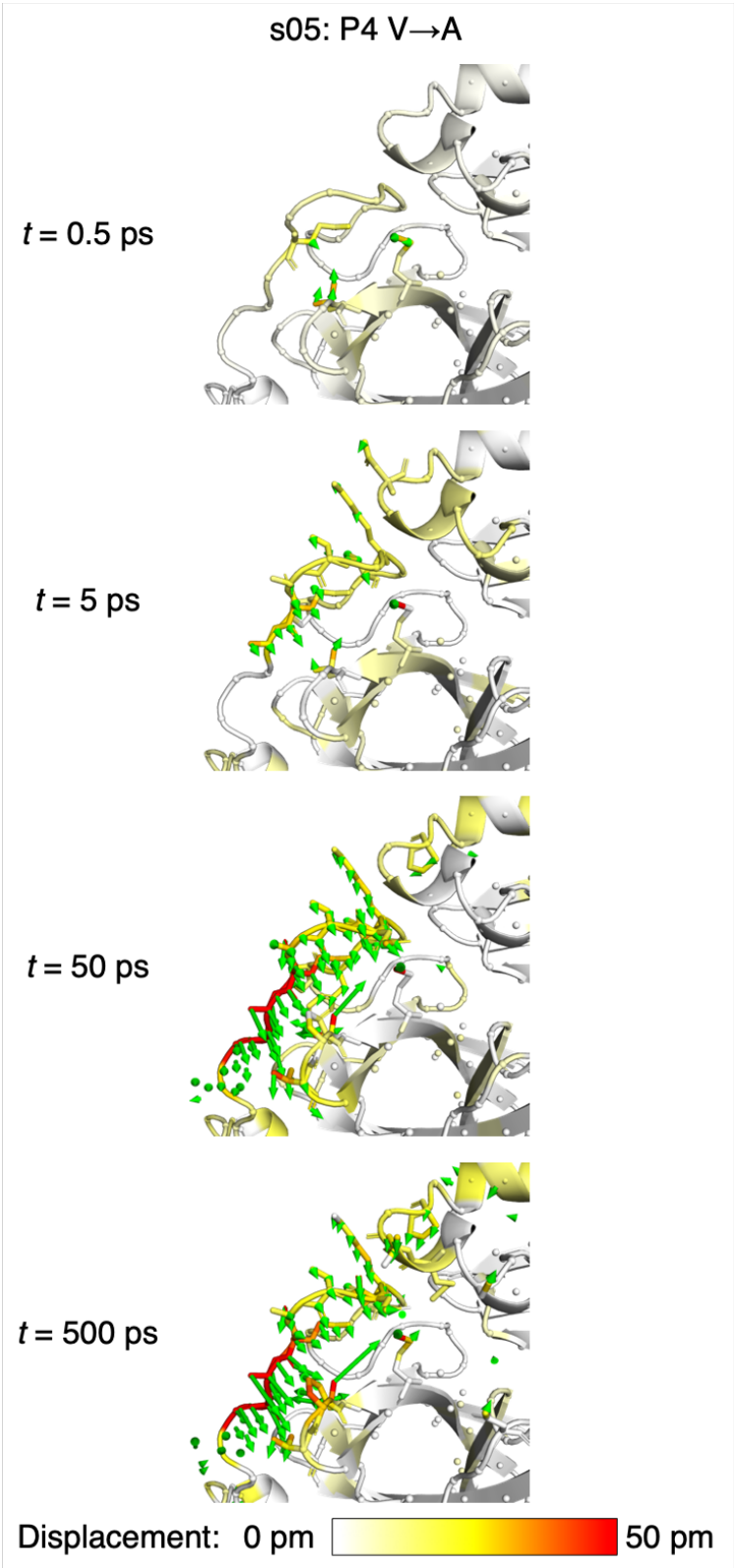


Figure S5.4B.6: Views of the M^{pro} response to the P4 mutation in s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations, with a focus on the S4 subsite. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7] Residues that show such significant displacements and that are within 10 Å of the substrate P4 residue are shown as sticks.

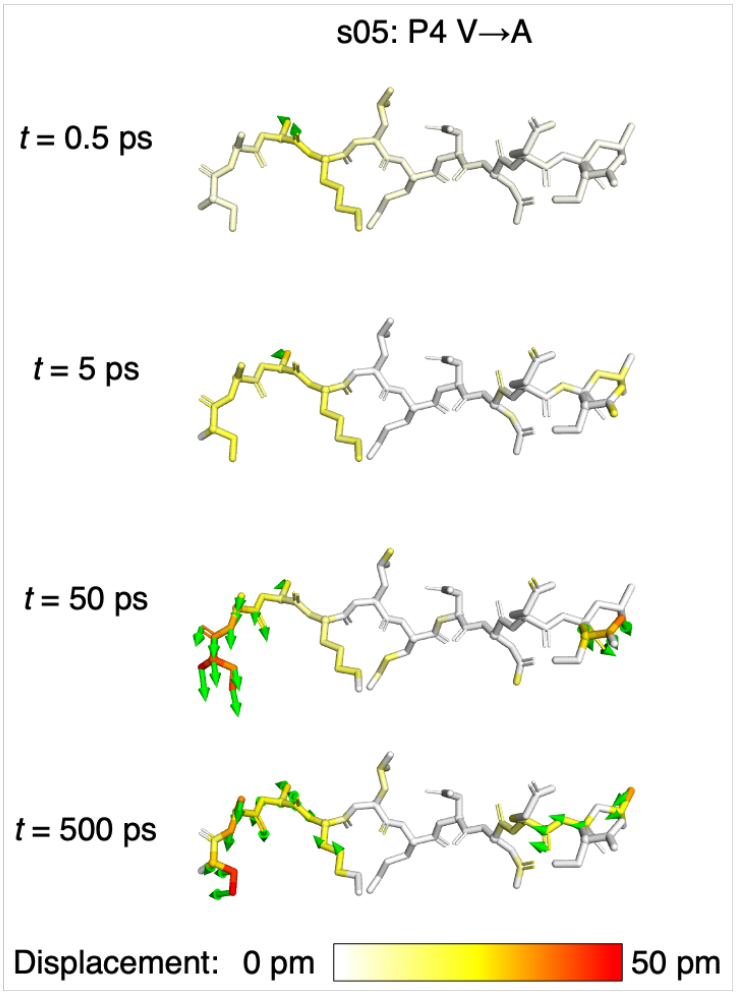


Figure S5.4B.7: Views of the peptide response to the P4 mutation in s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7]

Section S5.5A: Effect of P1′ mutation

Table S5.5A.1: The average distance differences in the five active-site distances ($\Delta d = d_{neq} - d_{eqm}$) following the P1′ mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

Active Site	s01: P1' S->A				s05: P1' N->A		
Deprot	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.01	0.00
	0.1	-0.01	0.00		0.1	0.00	0.01
	0.25	-0.02	0.01		0.25	-0.06	0.02
	0.5	0.00	0.02		0.5	-0.04	0.02
	1	-0.06	0.04		1	-0.07	0.04
	2	-0.12	0.04		2	-0.05	0.06
	3	-0.02	0.04		3	-0.07	0.07
	5	-0.10	0.06		5	-0.22	0.06
	10	0.02	0.06		10	-0.10	0.06
	20	0.00	0.04		20	-0.17	0.07
	30	-0.10	0.06		30	-0.11	0.07
	50	-0.04	0.06		50	-0.17	0.07
	100	-0.05	0.07		100	-0.25	0.06
	300	-0.23	0.06		300	-0.12	0.06
	500	-0.14	0.05		500	-0.10	0.05
	1000	-0.12	0.06		1000	-0.06	0.06
Nuc	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.01	0.00
	0.25	-0.01	0.01		0.25	0.01	0.01
	0.5	0.01	0.01		0.5	-0.01	0.01
	1	-0.01	0.01		1	0.00	0.01
	2	0.00	0.02		2	0.02	0.02
	3	-0.04	0.02		3	0.02	0.02
	5	-0.03	0.02		5	-0.04	0.02
	10	0.00	0.02		10	0.01	0.02
	20	0.06	0.03		20	-0.02	0.03
	30	-0.02	0.03		30	0.00	0.03
	50	-0.04	0.04		50	0.03	0.04
	100	-0.01	0.04		100	0.01	0.04
	300	-0.04	0.05		300	0.01	0.05
	500	0.07	0.05		500	-0.08	0.05
	1000	0.04	0.05		1000	-0.07	0.06
OH145	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.01	0.00
	0.1	0.00	0.00		0.1	0.02	0.00
	0.25	0.00	0.01		0.25	0.05	0.01
	0.5	0.03	0.01		0.5	0.09	0.02
	1	0.01	0.02		1	0.06	0.02
	2	0.00	0.03		2	0.09	0.03
	3	0.04	0.03		3	0.08	0.03
	5	0.04	0.03		5	0.10	0.03
	10	0.06	0.03		10	0.06	0.03
	20	-0.03	0.03		20	0.07	0.03
	30	0.04	0.03		30	0.08	0.03
	50	-0.02	0.03		50	0.09	0.03
	100	-0.04	0.03		100	0.07	0.03
	300	-0.02	0.03		300	0.09	0.03
	500	-0.05	0.03		500	0.09	0.03
	1000	0.02	0.03		1000	0.04	0.03
OH144	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	-0.02	0.01		0.1	-0.02	0.01
	0.25	-0.04	0.01		0.25	0.01	0.01
	0.5	0.00	0.01		0.5	0.04	0.01
	1	0.00	0.02		1	0.00	0.02
	2	0.01	0.03		2	0.00	0.04
	3	0.00	0.04		3	0.04	0.03
	5	0.07	0.04		5	0.05	0.04
	10	0.06	0.03		10	0.07	0.03
	20	-0.09	0.04		20	0.02	0.03
	30	0.00	0.04		30	0.07	0.03
	50	0.01	0.03		50	0.04	0.04
	100	0.00	0.04		100	0.01	0.03
	300	0.04	0.04		300	0.07	0.04
	500	-0.03	0.04		500	0.00	0.04
	1000	0.00	0.04		1000	0.01	0.04
OH143	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	-0.01	0.00		0.05	-0.02	0.00
	0.1	-0.01	0.00		0.1	-0.03	0.01
	0.25	-0.01	0.01		0.25	-0.05	0.01
	0.5	-0.03	0.01		0.5	-0.05	0.02
	1	-0.04	0.02		1	-0.06	0.02
	2	-0.01	0.03		2	-0.06	0.03
	3	-0.01	0.03		3	-0.09	0.03
	5	-0.03	0.03		5	-0.09	0.03
	10	-0.04	0.03		10	-0.03	0.03
	20	0.00	0.03		20	-0.05	0.03
	30	-0.04	0.03		30	-0.07	0.03
	50	0.04	0.03		50	-0.09	0.03
	100	0.02	0.03		100	-0.04	0.03
	300	-0.04	0.03		300	-0.06	0.03
	500	0.01	0.03		500	-0.04	0.03
	1000	-0.02	0.03		1000	0.01	0.03

Table S5.5A.2: The average distance differences in the donor-acceptor distances of HBs 1-5 following the P1′ mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB1-5	s01: P1' S->A				s05: P1' N->A		
HB1	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.00	0.00		0.25	0.01	0.00
	0.5	0.00	0.00		0.5	0.00	0.01
	1	0.00	0.01		1	-0.02	0.02
	2	0.00	0.02		2	-0.01	0.03
	3	0.02	0.02		3	-0.01	0.03
	5	-0.02	0.02		5	0.02	0.03
	10	0.06	0.03		10	0.02	0.03
	20	0.07	0.03		20	-0.01	0.03
	30	0.04	0.04		30	0.02	0.03
	50	0.01	0.02		50	0.04	0.03
	100	0.01	0.03		100	0.01	0.04
	300	0.05	0.03		300	0.00	0.04
	500	0.00	0.04		500	-0.01	0.04
	1000	-0.09	0.04		1000	0.08	0.06
HB2	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.01	0.00		0.25	0.01	0.00
	0.5	0.00	0.00		0.5	0.01	0.00
	1	0.00	0.01		1	0.01	0.01
	2	-0.01	0.02		2	0.00	0.02
	3	0.01	0.02		3	-0.02	0.02
	5	-0.01	0.02		5	-0.05	0.02
	10	0.00	0.02		10	0.03	0.02
	20	-0.01	0.02		20	0.00	0.02
	30	-0.03	0.02		30	-0.01	0.02
	50	0.00	0.02		50	0.01	0.02
	100	0.01	0.02		100	0.01	0.02
	300	0.00	0.02		300	-0.01	0.02
	500	0.02	0.02		500	0.00	0.02
	1000	-0.02	0.02		1000	-0.02	0.02
HB3	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.01	0.00
	0.25	0.00	0.00		0.25	0.02	0.00
	0.5	0.00	0.00		0.5	0.02	0.01
	1	0.01	0.01		1	0.01	0.01
	2	-0.02	0.01		2	0.00	0.02
	3	0.00	0.01		3	0.01	0.01
	5	-0.02	0.01		5	-0.02	0.02
	10	0.01	0.01		10	0.03	0.01
	20	-0.04	0.01		20	-0.02	0.01
	30	0.02	0.01		30	0.02	0.01
	50	-0.01	0.01		50	0.02	0.01
	100	0.02	0.01		100	0.02	0.01
	300	0.01	0.01		300	0.03	0.01
	500	0.02	0.01		500	0.03	0.01
	1000	0.00	0.01		1000	0.01	0.02
HB4	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.01	0.00
	0.25	0.01	0.00		0.25	0.01	0.01
	0.5	-0.01	0.01		0.5	0.00	0.01
	1	0.01	0.03		1	0.01	0.02
	2	0.06	0.03		2	-0.05	0.03
	3	0.05	0.03		3	0.01	0.03
	5	-0.01	0.03		5	0.01	0.03
	10	-0.04	0.04		10	-0.07	0.04
	20	-0.03	0.04		20	-0.10	0.06
	30	-0.05	0.06		30	-0.09	0.06
	50	-0.09	0.07		50	-0.11	0.06
	100	-0.06	0.09		100	-0.14	0.08
	300	-0.16	0.11		300	0.19	0.12
	500	-0.33	0.13		500	0.33	0.12
	1000	-0.35	0.15		1000	-0.05	0.14
HB5	t (ps)	Av. Diff (Å)	SEM (Å)		t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.01	0.00
	0.25	0.00	0.00		0.25	0.01	0.00
	0.5	0.00	0.01		0.5	0.02	0.01
	1	-0.03	0.02		1	-0.02	0.02
	2	-0.03	0.03		2	0.01	0.02
	3	-0.03	0.03		3	0.03	0.03
	5	-0.01	0.03		5	0.05	0.03
	10	0.03	0.03		10	0.00	0.02
	20	0.00	0.03		20	0.00	0.03
	30	0.03	0.03		30	0.02	0.02
	50	0.03	0.03		50	0.02	0.03
	100	0.02	0.03		100	-0.01	0.03
	300	0.00	0.03		300	0.11	0.03
	500	-0.01	0.04		500	0.05	0.03
	1000	0.04	0.04		1000	-0.03	0.03

Table S5.5A.3: The average distance differences in the donor-acceptor distances of HBs 8-12 following the P1′ mutation in s01 and s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB8-12	s01: P1' S->A				s05: P1' N->A		
HB8	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.01	0.00
	0.1	0.00	0.00		0.1	0.02	0.00
	0.25	0.00	0.01		0.25	0.05	0.01
	0.5	0.03	0.01		0.5	0.09	0.02
	1	0.01	0.02		1	0.06	0.02
	2	-0.01	0.03		2	0.09	0.03
	3	0.04	0.03		3	0.08	0.03
	5	0.04	0.03		5	0.09	0.03
	10	0.06	0.03		10	0.06	0.03
	20	-0.03	0.03		20	0.07	0.03
	30	0.03	0.03		30	0.08	0.03
	50	-0.02	0.03		50	0.09	0.03
	100	-0.04	0.03		100	0.06	0.03
	300	-0.02	0.03		300	0.09	0.03
	500	-0.06	0.03		500	0.09	0.03
	1000	0.02	0.03		1000	0.04	0.03
HB9	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	-0.01	0.00		0.05	-0.01	0.00
	0.1	-0.01	0.00		0.1	-0.02	0.00
	0.25	-0.01	0.00		0.25	-0.03	0.01
	0.5	-0.01	0.01		0.5	-0.02	0.01
	1	-0.01	0.01		1	-0.04	0.01
	2	0.00	0.02		2	-0.03	0.02
	3	-0.01	0.02		3	-0.05	0.02
	5	-0.01	0.02		5	-0.05	0.02
	10	-0.02	0.02		10	-0.01	0.02
	20	0.00	0.02		20	-0.02	0.02
	30	-0.02	0.02		30	-0.03	0.02
	50	0.02	0.02		50	-0.04	0.02
	100	0.01	0.02		100	-0.02	0.02
	300	-0.02	0.02		300	-0.03	0.02
	500	0.00	0.02		500	-0.02	0.02
	1000	-0.01	0.02		1000	0.01	0.02
HB10	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.01	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	0.01	0.00		0.25	0.02	0.01
	0.5	0.01	0.01		0.5	0.02	0.01
	1	0.00	0.01		1	0.00	0.01
	2	0.03	0.02		2	0.01	0.02
	3	0.02	0.02		3	0.03	0.02
	5	0.00	0.02		5	0.03	0.02
	10	0.00	0.02		10	0.02	0.02
	20	-0.02	0.02		20	0.01	0.02
	30	-0.01	0.02		30	0.04	0.02
	50	0.02	0.02		50	0.04	0.02
	100	0.00	0.02		100	0.02	0.02
	300	-0.01	0.02		300	0.02	0.02
	500	-0.03	0.02		500	-0.02	0.02
	1000	-0.01	0.02		1000	0.05	0.02
HB11	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	-0.02	0.00
	0.1	0.01	0.00		0.1	-0.06	0.01
	0.25	0.00	0.01		0.25	-0.05	0.01
	0.5	0.01	0.01		0.5	-0.06	0.01
	1	-0.02	0.02		1	-0.07	0.02
	2	-0.02	0.02		2	-0.07	0.01
	3	0.00	0.02		3	-0.04	0.02
	5	-0.01	0.02		5	-0.04	0.02
	10	0.01	0.02		10	-0.03	0.02
	20	0.01	0.03		20	-0.06	0.02
	30	-0.02	0.02		30	-0.05	0.02
	50	0.02	0.03		50	-0.01	0.02
	100	0.00	0.02		100	-0.04	0.02
	300	0.01	0.02		300	-0.03	0.02
	500	-0.03	0.02		500	-0.01	0.02
	1000	0.02	0.02		1000	0.00	0.02
HB12	t (ps)	Av Diff (Å)	SEM (Å)		t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00		0.05	0.00	0.00
	0.1	0.00	0.00		0.1	0.00	0.00
	0.25	-0.01	0.00		0.25	0.01	0.00
	0.5	-0.02	0.01		0.5	0.01	0.01
	1	0.00	0.02		1	0.01	0.02
	2	-0.01	0.02		2	-0.02	0.02
	3	0.06	0.03		3	0.03	0.02
	5	0.00	0.03		5	-0.04	0.03
	10	-0.05	0.03		10	0.05	0.03
	20	0.03	0.03		20	0.05	0.03
	30	-0.02	0.03		30	0.04	0.04
	50	-0.08	0.03		50	0.03	0.03
	100	0.02	0.04		100	0.00	0.03
	300	-0.05	0.05		300	0.03	0.03
	500	0.02	0.04		500	0.01	0.03
	1000	-0.06	0.05		1000	0.04	0.04

Table S5.5A.4: The ten most perturbed Cα atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P1′ mutation in s01 and s05.

s01 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	7ALA_CA	C	0.02	0.00
	2	8GLY_CA	C	0.01	0.00
	3				
	4				
	5				
	6				
	7				
	8				
	9				
	10				
0.5	1	27LEU_CA	A	0.02	0.00
	2	42VAL_CA	A	0.02	0.00
	3	46SER_CA	A	0.02	0.01
	4	187ASP_CA	A	0.02	0.00
	5	5LEU_CA	C	0.02	0.01
	6	7ALA_CA	C	0.02	0.01
	7	8GLY_CA	C	0.02	0.01
	8	19GLN_CA	A	0.01	0.00
	9	20VAL_CA	A	0.01	0.00
	10	22CYS_CA	A	0.01	0.00
5	1	72ASN_CA	B	0.17	0.05
	2	277ASN_CA	B	0.15	0.05
	3	71GLY_CA	B	0.12	0.06
	4	196THR_CA	B	0.12	0.05
	5	218TRP_CA	B	0.12	0.04
	6	278GLY_CA	B	0.12	0.05
	7	219PHE_CA	B	0.11	0.03
	8	275GLY_CA	B	0.11	0.05
	9	279ARG_CA	B	0.11	0.04
	10	283GLY_CA	B	0.11	0.04
50	1	72ASN_CA	B	0.25	0.08
	2	191ALA_CA	B	0.23	0.08
	3	232LEU_CA	A	0.19	0.06
	4	1THR_CA	C	0.19	0.09
	5	73VAL_CA	B	0.18	0.07
	6	142ASN_CA	B	0.14	0.04
	7	221ASN_CA	B	0.14	0.05
	8	225THR_CA	B	0.14	0.06
	9	226THR_CA	B	0.14	0.06
	10	83GLN_CA	A	0.13	0.04
500	1	274ASN_CA	A	0.24	0.09
	2	72ASN_CA	B	0.24	0.10
	3	73VAL_CA	B	0.24	0.08
	4	277ASN_CA	B	0.24	0.10
	5	278GLY_CA	B	0.22	0.11
	6	191ALA_CA	B	0.21	0.09
	7	59ILE_CA	A	0.20	0.06
	8	275GLY_CA	A	0.20	0.08
	9	74GLN_CA	B	0.20	0.06
	10	169THR_CA	B	0.20	0.07
1000	1	244GLN_CA	A	0.24	0.08
	2	236LYS_CA	A	0.23	0.08
	3	11LYS_CA	C	0.23	0.14
	4	235MET_CA	A	0.22	0.08
	5	245ASP_CA	A	0.22	0.07
	6	51ASN_CA	A	0.21	0.07
	7	232LEU_CA	A	0.21	0.07
	8	283GLY_CA	A	0.21	0.07
	9	243THR_CA	A	0.20	0.07
	10	92ASP_CA	A	0.19	0.07

s05 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	25THR_CA	A	0.01	0.00
	2	6GLN_CA	C	0.01	0.00
	3	7ALA_CA	C	0.01	0.00
	4	8ASN_CA	C	0.01	0.00
	5				
	6				
	7				
	8				
	9				
	10				
0.5	1	7ALA_CA	C	0.11	0.02
	2	44CYS_CA	A	0.07	0.01
	3	6GLN_CA	C	0.07	0.01
	4	41HID_CA	A	0.06	0.01
	5	43ILE_CA	A	0.06	0.01
	6	5LEU_CA	C	0.06	0.01
	7	8ASN_CA	C	0.06	0.01
	8	9GLU_CA	C	0.06	0.01
	9	42VAL_CA	A	0.05	0.01
	10	49MET_CA	A	0.05	0.01
5	1	285ALA_CA	B	0.16	0.03
	2	51ASN_CA	B	0.15	0.04
	3	286LEU_CA	B	0.15	0.03
	4	49MET_CA	A	0.13	0.04
	5	72ASN_CA	A	0.13	0.05
	6	284SER_CA	B	0.13	0.03
	7	44CYS_CA	A	0.12	0.03
	8	50LEU_CA	A	0.12	0.04
	9	64HIE_CA	A	0.12	0.04
	10	235MET_CA	A	0.12	0.04
50	1	72ASN_CA	A	0.29	0.08
	2	275GLY_CA	B	0.22	0.06
	3	274ASN_CA	B	0.21	0.07
	4	277ASN_CA	B	0.21	0.09
	5	71GLY_CA	A	0.20	0.08
	6	236LYS_CA	A	0.20	0.06
	7	24THR_CA	B	0.20	0.07
	8	72ASN_CA	B	0.18	0.08
	9	64HIE_CA	A	0.17	0.06
	10	227LEU_CA	A	0.17	0.05
500	1	278GLY_CA	A	0.33	0.10
	2	10LEU_CA	C	0.32	0.09
	3	277ASN_CA	A	0.30	0.11
	4	72ASN_CA	A	0.28	0.10
	5	244GLN_CA	B	0.28	0.07
	6	9GLU_CA	C	0.27	0.07
	7	245ASP_CA	B	0.25	0.07
	8	274ASN_CA	A	0.24	0.09
	9	243THR_CA	B	0.24	0.07
	10	11SER_CA	C	0.24	0.16
1000	1	11SER_CA	C	0.39	0.18
	2	59ILE_CA	B	0.21	0.09
	3	60ARG_CA	B	0.21	0.09
	4	10LEU_CA	C	0.21	0.09
	5	1SER_CA	C	0.20	0.12
	6	15GLY_CA	A	0.19	0.05
	7	222ARG_CA	B	0.19	0.10
	8	234ALA_CA	A	0.18	0.06
	9	215GLY_CA	B	0.18	0.10
	10	9GLU_CA	C	0.17	0.06

Table S5.5A.5: The ten most perturbed non-hydrogen atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P1′ mutation in s01 and s05. Only displacements ≥0.01 Å are shown.

s01 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	7ALA_CB	C	0.06	0.01
	2	5LEU_O	C	0.02	0.00
	3	26THR_O	A	0.02	0.00
	4	7ALA_CA	C	0.02	0.00
	5	6GLN_O	C	0.01	0.00
	6	8GLY_N	C	0.01	0.00
	7	7ALA_C	C	0.01	0.00
	8	7ALA_N	C	0.01	0.00
	9	41HID_NE2	A	0.01	0.00
	10	7ALA_O	C	0.01	0.00
0.5	1	7ALA_CB	C	0.08	0.02
	2	41HID_NE2	A	0.08	0.02
	3	41HID_CE1	A	0.08	0.02
	4	49MET_CE	A	0.07	0.02
	5	5LEU_O	C	0.07	0.02
	6	41HID_ND1	A	0.06	0.01
	7	41HID_CD2	A	0.06	0.01
	8	7ALA_O	C	0.06	0.02
	9	26THR_O	A	0.06	0.01
	10	25THR_CG2	A	0.05	0.02
5	1	72ASN_ND2	B	0.30	0.09
	2	72ASN_CG	B	0.26	0.08
	3	72ASN_CB	B	0.24	0.07
	4	10ARG_NH2	C	0.23	0.07
	5	306GLN_OE1	B	0.23	0.07
	6	72ASN_OD1	B	0.22	0.09
	7	294PHE_CZ	A	0.21	0.07
	8	277ASN_O	B	0.21	0.06
	9	53ASN_OD1	B	0.20	0.07
	10	47GLU_OE2	B	0.20	0.09
50	1	222ARG_NH2	B	0.42	0.19
	2	235MET_CE	A	0.39	0.15
	3	55GLU_OE1	B	0.39	0.13
	4	222ARG_NH1	A	0.38	0.19
	5	72ASN_ND2	B	0.37	0.15
	6	222ARG_NH1	B	0.36	0.19
	7	11LYS_NZ	C	0.36	0.19
	8	142ASN_OD1	B	0.33	0.10
	9	217ARG_NH1	B	0.32	0.12
	10	191ALA_CB	B	0.32	0.10
500	1	222ARG_NH2	B	0.77	0.22
	2	64HIE_NE2	B	0.69	0.15
	3	222ARG_NH1	B	0.69	0.22
	4	189GLN_NE2	B	0.61	0.14
	5	222ARG_CZ	B	0.59	0.18
	6	222ARG_NH2	A	0.57	0.23
	7	64HIE_CD2	B	0.56	0.12
	8	64HIE_CE1	B	0.55	0.15
	9	107GLN_OE1	B	0.49	0.17
	10	189GLN_OE1	B	0.49	0.15
1000	1	236LYS_NZ	A	0.84	0.24
	2	236LYS_CE	A	0.77	0.21
	3	279ARG_NH1	B	0.70	0.22
	4	60ARG_NH2	A	0.63	0.19
	5	47GLU_OE2	B	0.60	0.22
	6	222ARG_NH1	B	0.60	0.27
	7	236LYS_CD	A	0.57	0.16
	8	244GLN_NE2	A	0.56	0.24
	9	235MET_CE	B	0.54	0.27
	10	235MET_SD	B	0.53	0.19

s05 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	5LEU_O	C	0.06	0.00
	2	8ASN_O	C	0.04	0.00
	3	7ALA_CB	C	0.04	0.01
	4	7ALA_O	C	0.04	0.00
	5	4LYS_NZ	C	0.02	0.00
	6	8ASN_N	C	0.02	0.00
	7	25THR_CG2	A	0.02	0.00
	8	26THR_O	A	0.02	0.00
	9	5LEU_C	C	0.02	0.00
	10	49MET_CE	A	0.02	0.00
0.5	1	7ALA_O	C	0.28	0.03
	2	7ALA_C	C	0.18	0.02
	3	49MET_CE	A	0.17	0.02
	4	5LEU_O	C	0.16	0.01
	5	49MET_SD	A	0.13	0.02
	6	8ASN_O	C	0.13	0.02
	7	41HID_NE2	A	0.12	0.01
	8	7ALA_CB	C	0.12	0.02
	9	25THR_CG2	A	0.12	0.02
	10	7ALA_CA	C	0.12	0.02
5	1	7ALA_O	C	0.34	0.05
	2	49MET_CE	A	0.26	0.04
	3	228ASN_OD1	A	0.24	0.08
	4	217ARG_NH1	A	0.23	0.08
	5	107GLN_NE2	A	0.23	0.08
	6	107GLN_NE2	B	0.23	0.09
	7	189GLN_NE2	B	0.23	0.07
	8	235MET_SD	A	0.23	0.08
	9	41HID_NE2	A	0.22	0.03
	10	49MET_SD	A	0.22	0.04
50	1	72ASN_OD1	A	0.51	0.14
	2	279ARG_NH2	B	0.50	0.11
	3	277ASN_ND2	B	0.41	0.13
	4	72ASN_CG	A	0.40	0.13
	5	223PHE_CZ	A	0.39	0.12
	6	222ARG_NH2	A	0.39	0.16
	7	7ALA_O	C	0.38	0.05
	8	279ARG_CZ	B	0.37	0.10
	9	72ASN_ND2	A	0.36	0.16
	10	72ASN_CB	A	0.35	0.11
500	1	256GLN_NE2	A	0.69	0.15
	2	222ARG_NH2	B	0.62	0.28
	3	235MET_CE	B	0.56	0.20
	4	10LEU_CD1	C	0.55	0.14
	5	189GLN_NE2	B	0.51	0.13
	6	107GLN_NE2	B	0.49	0.14
	7	60ARG_NH2	A	0.49	0.17
	8	245ASP_OD1	B	0.49	0.15
	9	60ARG_NH1	A	0.48	0.18
	10	189GLN_OE1	B	0.47	0.14
1000	1	222ARG_NH2	B	0.89	0.30
	2	235MET_CE	A	0.80	0.24
	3	222ARG_CZ	B	0.72	0.27
	4	235MET_SD	A	0.70	0.19
	5	222ARG_NH1	B	0.68	0.31
	6	76ARG_NH1	B	0.65	0.22
	7	76ARG_NH2	B	0.63	0.23
	8	222ARG_NE	B	0.62	0.23
	9	76ARG_CZ	B	0.52	0.19
	10	11SER_OG	C	0.51	0.27

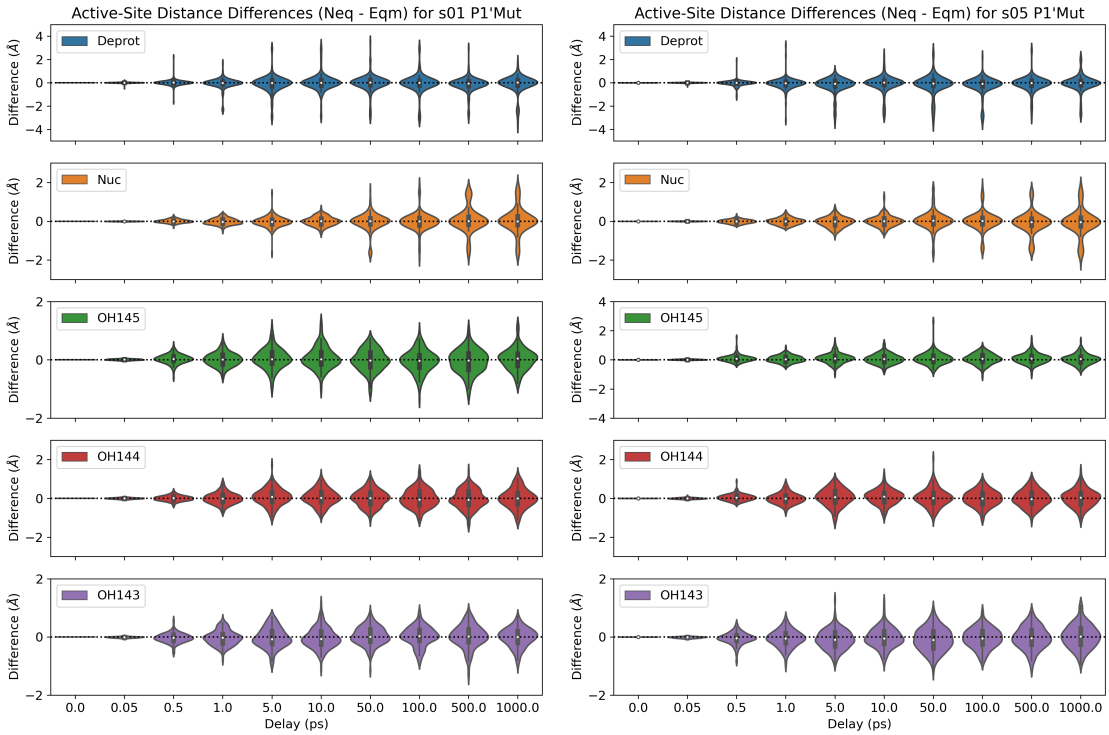


Figure S5.5A.1: Distributions of the distance differences in the five active-site distances ($\Delta d = d_{\mathrm{neq}} - d_{\mathrm{eqm}}$) following the P1′ mutation in s01 and s05.

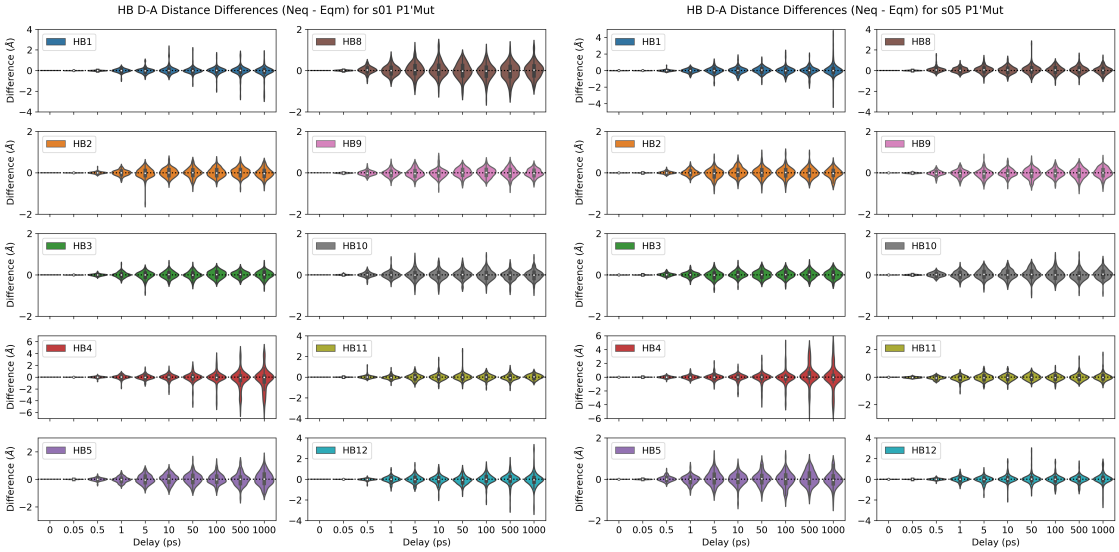


Figure S5.5A.2: Distributions of the distance differences in the ten donor-acceptor distances of HBs 1-5 and 8-12 following the P1′ mutation in s01 and s05.

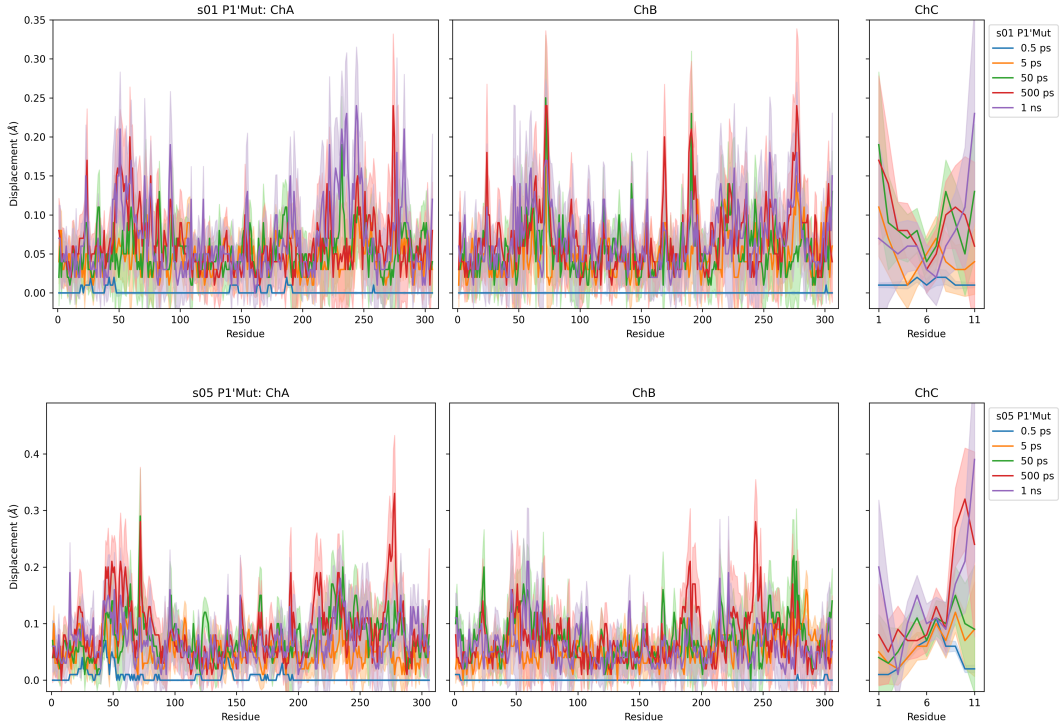


Figure S5.5A.3: Magnitudes of the average M^{pro} Cα displacement vectors, with ± SEM shown as shaded regions, following the P1′ mutation in s01 and s05.

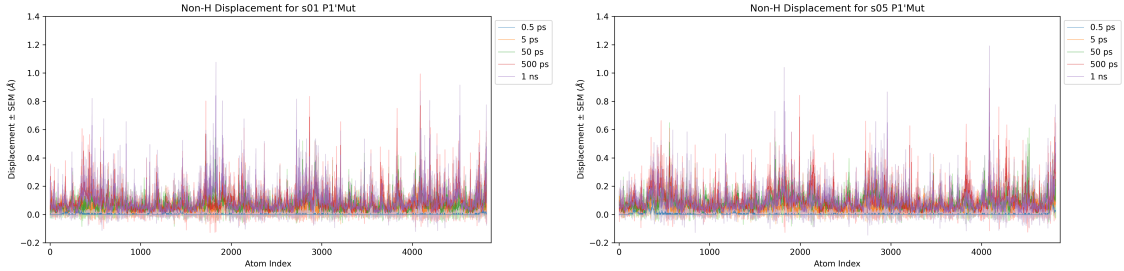


Figure S5.5A.4: Magnitudes of the average non-hydrogen atom displacement vectors in the M^{pro}-peptide complex, with ± SEM shown as shaded regions, following the P1′ mutation in s01 and s05.

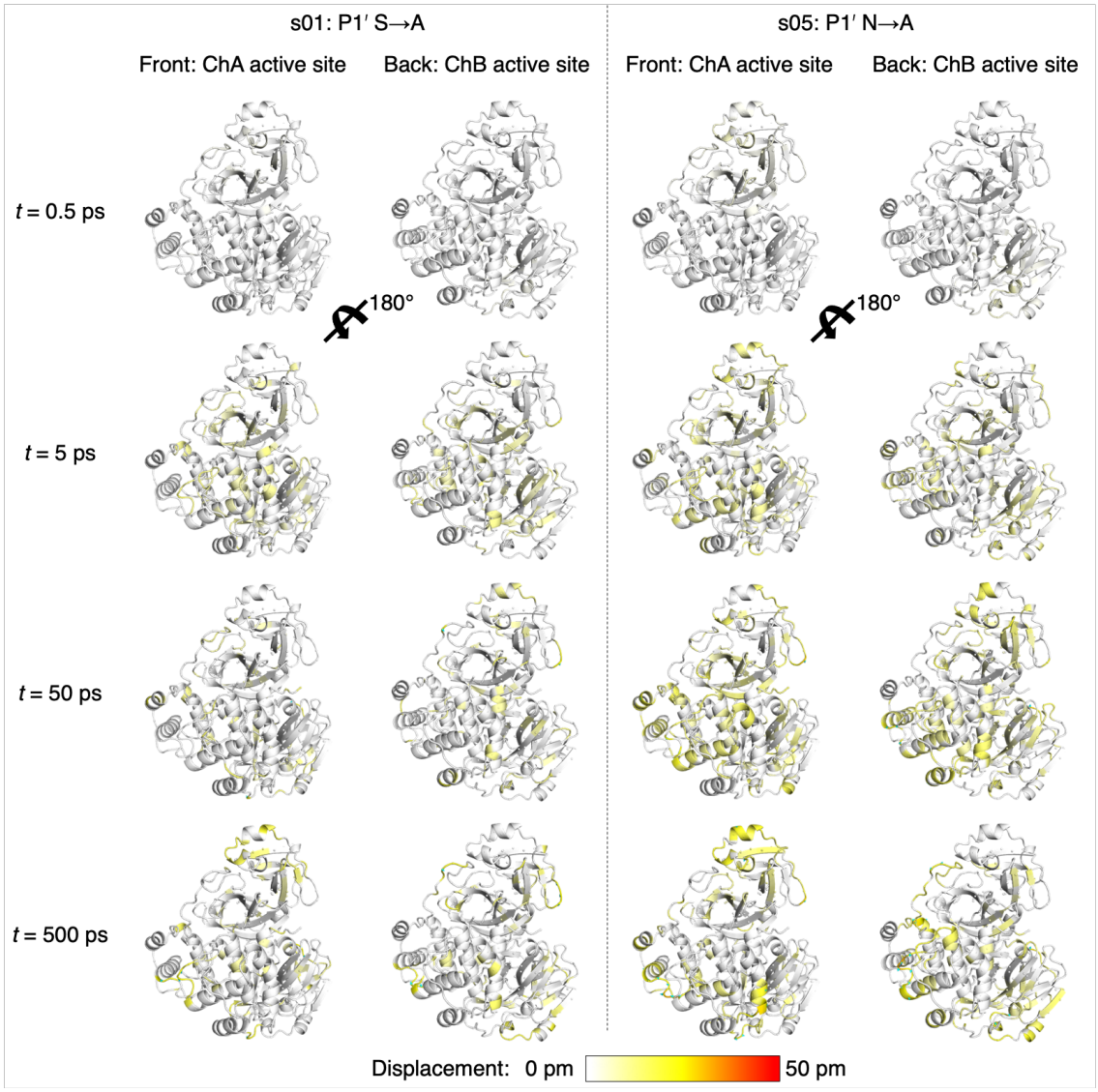


Figure S5.5A.5: Views of the M^{pro} response to the P1′ mutation in s01 and s05 from averaging Cα displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as cyan arrows with a scale-up factor of 5.[7]

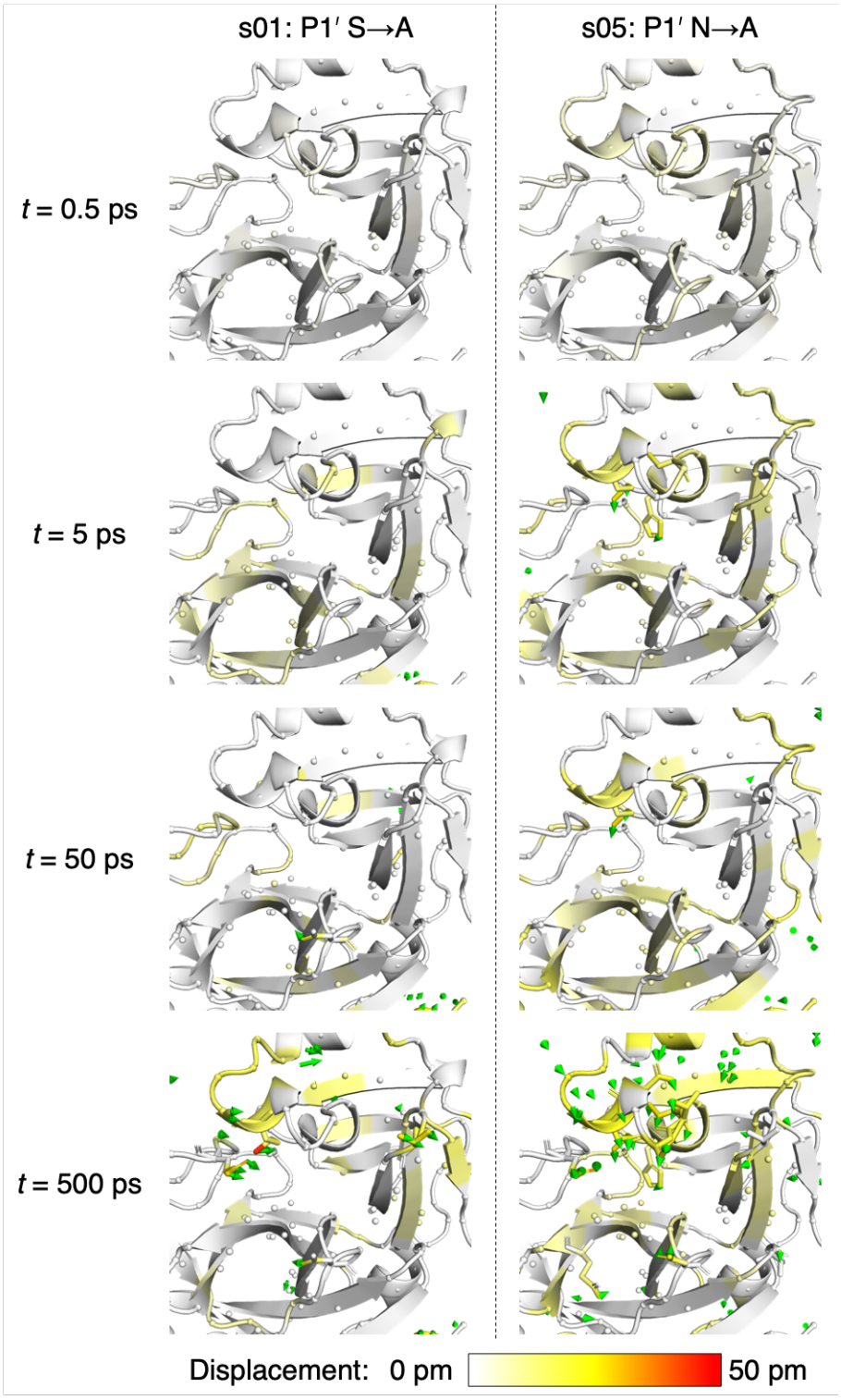


Figure S5.5A.6: Views of the M^{pro} response to the P1′ mutation in s01 and s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations, with a focus on the S1′ subsite. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7] Residues that show such significant displacements and that are within 10 Å of the substrate P1′ residue are shown as sticks.

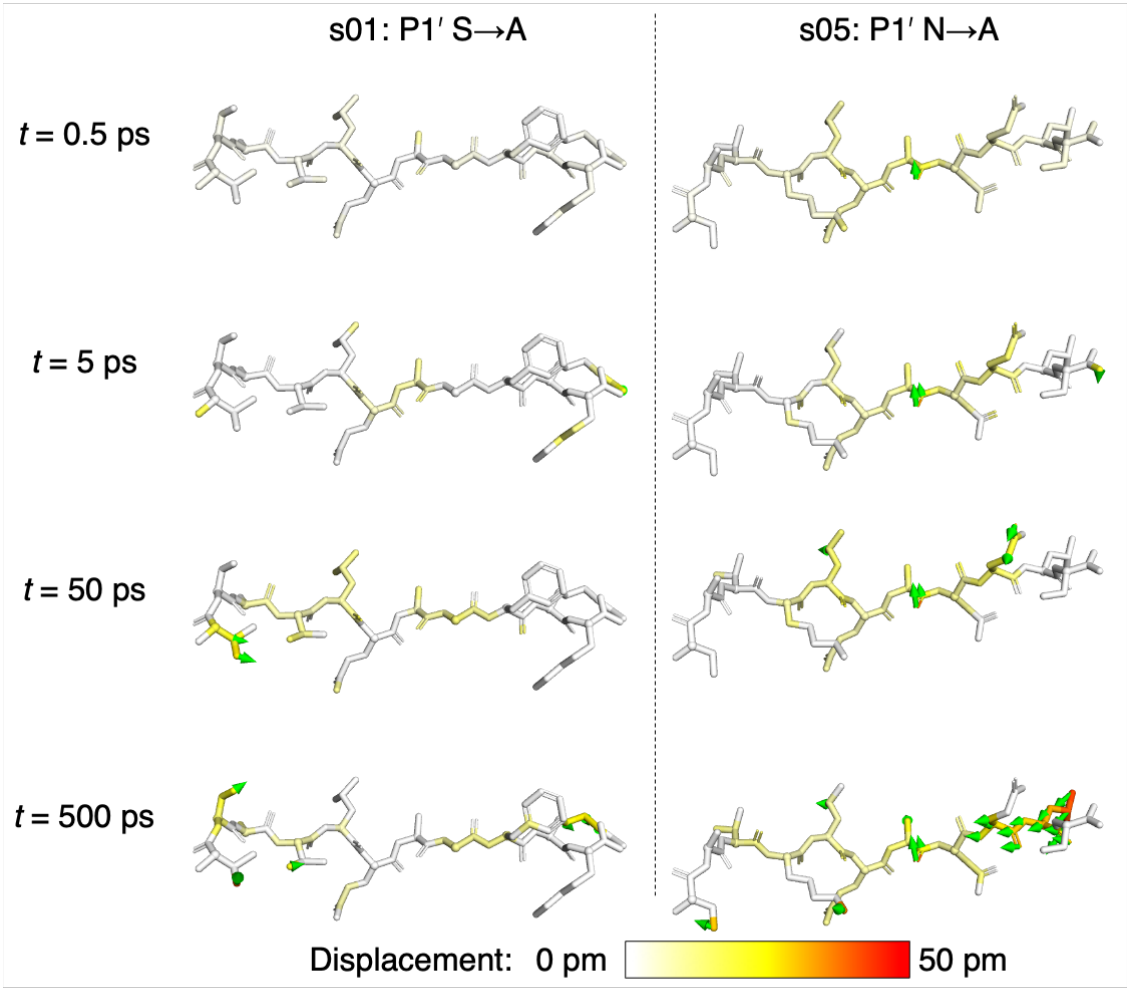


Figure S5.5A.7: Views of the peptide response to the P1′ mutation in s01 and s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7]

Section S5.5B: Effect of P2′ mutation

Table S5.5B.1: The average distance differences in the five active-site distances ($\Delta d = d_{neq} - d_{eqm}$) following the P2′ mutation in s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

Active Site	s05: P2' N->A		
Deprot	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.01
	1	-0.02	0.03
	2	0.00	0.05
	3	-0.07	0.06
	5	-0.05	0.07
	10	-0.04	0.05
	20	0.03	0.06
	30	0.01	0.08
	50	0.01	0.07
	100	-0.01	0.08
	300	-0.01	0.07
	500	-0.02	0.07
	1000	0.12	0.07
Nuc	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.01
	1	0.01	0.01
	2	-0.02	0.02
	3	0.00	0.02
	5	-0.01	0.02
	10	-0.02	0.03
	20	0.00	0.04
	30	0.00	0.04
	50	-0.04	0.04
	100	0.01	0.05
	300	0.04	0.05
	500	0.03	0.04
	1000	-0.04	0.05
OH145	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	-0.01	0.00
	0.25	-0.01	0.00
	0.5	0.00	0.01
	1	-0.03	0.02
	2	-0.02	0.02
	3	-0.01	0.02
	5	0.00	0.03
	10	-0.01	0.02
	20	-0.04	0.02
	30	0.00	0.03
	50	-0.01	0.02
	100	-0.02	0.03
	300	-0.02	0.03
	500	0.00	0.02
	1000	-0.01	0.03
OH144	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	-0.02	0.00
	0.25	-0.02	0.01
	0.5	-0.02	0.01
	1	-0.01	0.02
	2	-0.04	0.03
	3	-0.05	0.03
	5	0.02	0.04
	10	0.06	0.03
	20	0.00	0.03
	30	0.00	0.04
	50	-0.01	0.03
	100	-0.06	0.03
	300	0.00	0.04
	500	0.03	0.03
	1000	-0.06	0.04
OH143	t (ps)	Av. Diff (Å)	SEM (Å)
	0.05	0.01	0.00
	0.1	0.02	0.00
	0.25	0.03	0.01
	0.5	-0.01	0.01
	1	0.01	0.02
	2	0.04	0.03
	3	-0.01	0.03
	5	-0.03	0.03
	10	-0.01	0.03
	20	0.00	0.03
	30	0.03	0.03
	50	-0.03	0.03
	100	0.02	0.03
	300	0.00	0.03
	500	-0.01	0.03
	1000	-0.02	0.03

Table S5.5B.2: The average distance differences in the donor-acceptor distances of HBs 1-5 following the P2′ mutation in s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB1-5	s05: P2' N->A		
HB1	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	-0.01	0.00
	1	-0.01	0.02
	2	0.00	0.02
	3	-0.01	0.03
	5	0.02	0.03
	10	-0.03	0.03
	20	-0.04	0.03
	30	0.01	0.03
	50	0.03	0.03
	100	0.01	0.03
	300	-0.01	0.04
	500	-0.01	0.05
	1000	-0.03	0.05
HB2	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	0.01	0.01
	2	-0.03	0.02
	3	0.03	0.02
	5	-0.05	0.02
	10	0.02	0.02
	20	-0.01	0.02
	30	0.00	0.02
	50	0.02	0.02
	100	-0.01	0.02
	300	0.01	0.02
	500	0.03	0.02
	1000	-0.02	0.02
HB3	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.00
	1	-0.01	0.01
	2	-0.02	0.02
	3	0.02	0.01
	5	-0.04	0.02
	10	0.01	0.01
	20	-0.02	0.02
	30	-0.01	0.01
	50	0.01	0.01
	100	-0.01	0.01
	300	0.00	0.01
	500	0.02	0.02
	1000	-0.01	0.02
HB4	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.00
	0.5	0.00	0.01
	1	0.01	0.02
	2	-0.02	0.03
	3	0.01	0.03
	5	0.05	0.04
	10	-0.07	0.04
	20	0.00	0.05
	30	0.01	0.05
	50	0.02	0.07
	100	0.05	0.08
	300	0.08	0.11
	500	0.11	0.11
	1000	-0.24	0.13
HB5	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	-0.01	0.00
	0.5	0.00	0.01
	1	-0.06	0.02
	2	0.03	0.02
	3	-0.04	0.02
	5	-0.01	0.02
	10	-0.03	0.03
	20	-0.05	0.03
	30	-0.01	0.02
	50	0.02	0.03
	100	-0.04	0.03
	300	-0.03	0.02
	500	-0.04	0.03
	1000	-0.01	0.03

Table S5.5B.3: The average distance differences in the donor-acceptor distances of HBs 8-12 following the P2′ mutation in s05. Positive and negative differences with magnitudes >0.1 Å are in green and red, respectively.

HB8-12	s05: P2' N->A		
HB8	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	-0.01	0.00
	0.25	-0.01	0.00
	0.5	0.00	0.01
	1	-0.02	0.02
	2	-0.02	0.02
	3	-0.01	0.02
	5	0.00	0.03
	10	-0.02	0.02
	20	-0.03	0.02
	30	0.00	0.03
	50	-0.01	0.02
	100	-0.03	0.03
	300	-0.02	0.03
	500	0.00	0.02
	1000	0.00	0.03
HB9	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.01	0.00
	0.1	0.01	0.00
	0.25	0.01	0.00
	0.5	0.00	0.01
	1	0.01	0.01
	2	0.02	0.02
	3	0.00	0.02
	5	-0.01	0.02
	10	0.01	0.02
	20	0.01	0.02
	30	0.00	0.02
	50	-0.01	0.02
	100	0.02	0.02
	300	0.02	0.02
	500	0.00	0.02
	1000	-0.01	0.02
HB10	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	-0.01	0.00
	0.25	0.00	0.01
	0.5	0.00	0.01
	1	0.01	0.01
	2	-0.04	0.02
	3	0.00	0.02
	5	0.02	0.02
	10	-0.01	0.02
	20	-0.03	0.02
	30	0.00	0.02
	50	-0.01	0.02
	100	-0.02	0.02
	300	0.00	0.02
	500	-0.06	0.02
	1000	0.02	0.02
HB11	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	-0.01	0.00
	0.1	0.01	0.00
	0.25	-0.01	0.01
	0.5	0.00	0.01
	1	0.03	0.02
	2	-0.01	0.02
	3	0.00	0.01
	5	0.01	0.02
	10	0.00	0.02
	20	-0.01	0.02
	30	0.00	0.02
	50	0.02	0.02
	100	0.02	0.02
	300	0.00	0.02
	500	0.01	0.02
	1000	0.06	0.02
HB12	t (ps)	Av Diff (Å)	SEM (Å)
	0.05	0.00	0.00
	0.1	0.00	0.00
	0.25	0.00	0.01
	0.5	0.01	0.01
	1	0.01	0.02
	2	-0.05	0.02
	3	0.02	0.02
	5	-0.02	0.03
	10	0.01	0.03
	20	0.03	0.03
	30	0.00	0.03
	50	0.02	0.03
	100	0.03	0.04
	300	0.05	0.03
	500	0.01	0.04
	1000	0.07	0.04

Table S5.5B.4: The ten most perturbed Cα atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P2′ mutation in s05. Only displacements ≥0.01 Å are shown.

s05 t(ps)	CA Rank	Atom	Chain	v (Å)	SEM (Å)
0.05	1	8ALA_CA	C	0.02	0.00
	2	143GLY_CA	A	0.01	0.00
	3	7ASN_CA	C	0.01	0.00
	4	9GLU_CA	C	0.01	0.00
	5				
	6				
	7				
	8				
	9				
	10				
0.5	1	8ALA_CA	C	0.04	0.01
	2	9GLU_CA	C	0.04	0.01
	3	26THR_CA	A	0.03	0.01
	4	142ASN_CA	A	0.03	0.01
	5	143GLY_CA	A	0.03	0.01
	6	144SER_CA	A	0.03	0.01
	7	27LEU_CA	A	0.02	0.00
	8	46SER_CA	A	0.02	0.00
	9	47GLU_CA	A	0.02	0.00
	10	119ASN_CA	A	0.02	0.01
5	1	251GLY_CA	B	0.17	0.04
	2	252PRO_CA	B	0.17	0.04
	3	183GLY_CA	B	0.14	0.04
	4	50LEU_CA	B	0.13	0.04
	5	51ASN_CA	B	0.13	0.04
	6	235MET_CA	B	0.13	0.04
	7	11SER_CA	C	0.13	0.07
	8	218TRP_CA	A	0.12	0.04
	9	221ASN_CA	A	0.12	0.04
	10	222ARG_CA	A	0.12	0.05
50	1	11SER_CA	C	0.28	0.13
	2	277ASN_CA	B	0.24	0.10
	3	215GLY_CA	B	0.23	0.07
	4	278GLY_CA	B	0.23	0.09
	5	279ARG_CA	B	0.21	0.07
	6	50LEU_CA	A	0.20	0.06
	7	23GLY_CA	B	0.20	0.06
	8	24THR_CA	B	0.20	0.07
	9	275GLY_CA	B	0.20	0.08
	10	276MET_CA	B	0.19	0.07
500	1	278GLY_CA	A	0.26	0.11
	2	191ALA_CA	B	0.26	0.10
	3	23GLY_CA	B	0.25	0.07
	4	215GLY_CA	B	0.24	0.09
	5	155ASP_CA	A	0.23	0.06
	6	215GLY_CA	A	0.23	0.08
	7	277ASN_CA	A	0.23	0.11
	8	276MET_CA	A	0.22	0.07
	9	190THR_CA	B	0.22	0.07
	10	252PRO_CA	B	0.22	0.08
1000	1	11SER_CA	C	0.81	0.22
	2	10LEU_CA	C	0.47	0.11
	3	9GLU_CA	C	0.39	0.07
	4	276MET_CA	A	0.24	0.10
	5	215GLY_CA	A	0.23	0.09
	6	194ALA_CA	B	0.21	0.07
	7	215GLY_CA	B	0.21	0.11
	8	55GLU_CA	B	0.20	0.09
	9	273GLN_CA	B	0.20	0.10
	10	274ASN_CA	B	0.20	0.11

Table S5.5B.5: The ten most perturbed non-hydrogen atoms in the M^{pro}-peptide complex in terms of magnitude of the average displacement vector *v*, at selected time points following the P2′ mutation in s05.

s05 t(ps)	*Non-H Rank*	*Atom*	*Chain*	*v (Å)*	*SEM (Å)*
0.05	1	8ALA_CB	C	0.05	0.01
	2	8ALA_N	C	0.02	0.01
	3	8ALA_CA	C	0.02	0.00
	4	119ASN_OD1	A	0.02	0.00
	5	8ALA_C	C	0.02	0.00
	6	9GLU_CB	C	0.01	0.00
	7	8ALA_O	C	0.01	0.00
	8	7ASN_O	C	0.01	0.00
	9	7ASN_C	C	0.01	0.00
	10	9GLU_N	C	0.01	0.00
0.5	1	119ASN_OD1	A	0.11	0.03
	2	9GLU_O	C	0.10	0.02
	3	8ALA_CB	C	0.08	0.03
	4	7ASN_O	C	0.06	0.02
	5	119ASN_ND2	A	0.06	0.03
	6	9GLU_C	C	0.05	0.01
	7	9GLU_CB	C	0.05	0.01
	8	46SER_OG	A	0.05	0.01
	9	9GLU_CA	C	0.04	0.01
	10	19GLN_NE2	A	0.04	0.02
5	1	11SER_OC1	C	0.31	0.10
	2	47GLU_OE2	A	0.31	0.09
	3	252PRO_CG	B	0.27	0.06
	4	47GLU_OE1	B	0.26	0.11
	5	107GLN_NE2	B	0.25	0.09
	6	256GLN_NE2	A	0.25	0.07
	7	142ASN_ND2	B	0.24	0.08
	8	223PHE_CZ	B	0.23	0.08
	9	223PHE_CE2	B	0.22	0.08
	10	11SER_C	C	0.22	0.09
50	1	11SER_OC2	C	0.56	0.17
	2	11SER_C	C	0.42	0.16
	3	279ARG_NH2	B	0.41	0.12
	4	11SER_OC1	C	0.38	0.19
	5	256GLN_NE2	A	0.36	0.11
	6	279ARG_CZ	B	0.35	0.11
	7	217ARG_NH2	B	0.34	0.14
	8	279ARG_NH1	B	0.34	0.12
	9	82MET_CE	A	0.33	0.08
	10	217ARG_NH1	A	0.32	0.11
500	1	235MET_CE	A	0.54	0.23
	2	72ASN_OD1	B	0.50	0.19
	3	256GLN_NE2	A	0.49	0.16
	4	306GLN_OE1	B	0.48	0.17
	5	90LYS_NZ	B	0.46	0.15
	6	277ASN_OD1	A	0.44	0.18
	7	277ASN_ND2	A	0.44	0.18
	8	248ASP_OD2	A	0.43	0.13
	9	279ARG_NH1	A	0.43	0.19
	10	306GLN_OE1	A	0.43	0.17
1000	1	11SER_OG	C	1.06	0.29
	2	11SER_CB	C	0.99	0.28
	3	11SER_CA	C	0.81	0.22
	4	11SER_C	C	0.70	0.24
	5	9GLU_O	C	0.69	0.10
	6	11SER_OC1	C	0.68	0.24
	7	11SER_OC2	C	0.68	0.28
	8	11SER_N	C	0.66	0.18
	9	235MET_CE	A	0.66	0.25
	10	9GLU_OE2	C	0.66	0.14

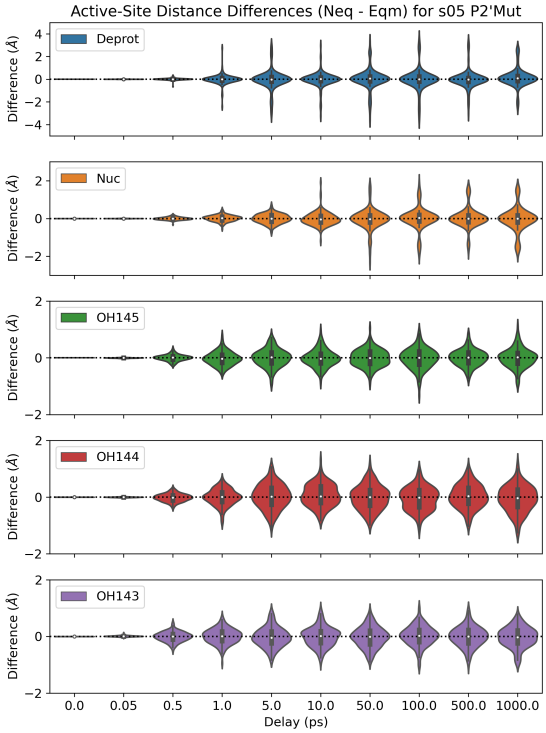


Figure S5.5B.1: Distributions of the distance differences in the five active-site distances ($\Delta d = d_{\mathrm{neq}} - d_{\mathrm{eqm}}$) following the P2′ mutation in s05.

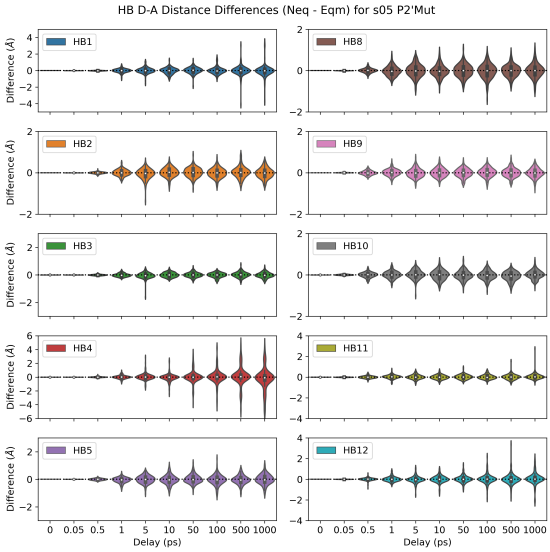


Figure S5.5B.2: Distributions of the distance differences in the ten donor-acceptor distances of HBs 1-5 and 8-12 following the P2′ mutation in s05.

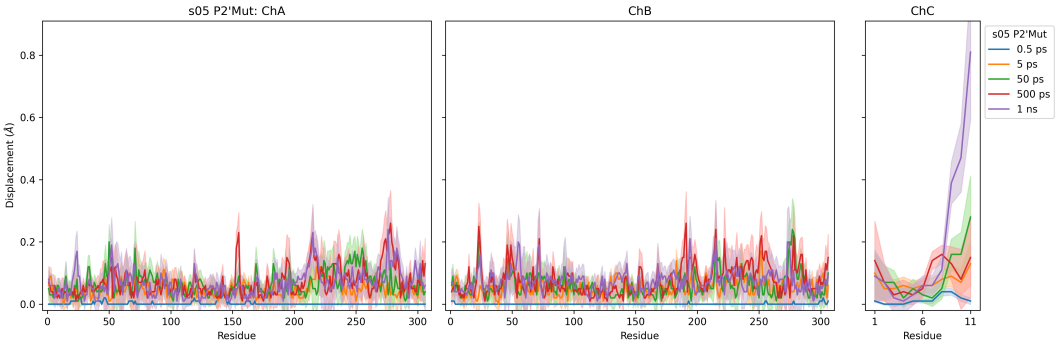


Figure S5.5B.3: Magnitudes of the average M^{pro} Cα displacement vectors, with ± SEM shown as shaded regions, following the P2′ mutation in s05.

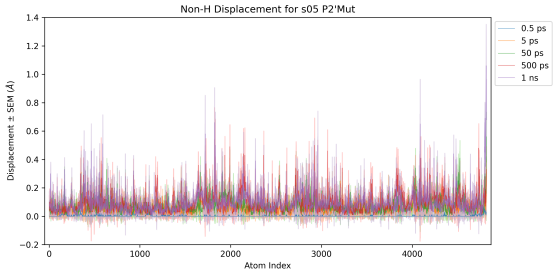


Figure S5.5B.4: Magnitudes of the average non-hydrogen atom displacement vectors in the M^{pro}-peptide complex, with ± SEM shown as shaded regions, following the P2′ mutation in s05.

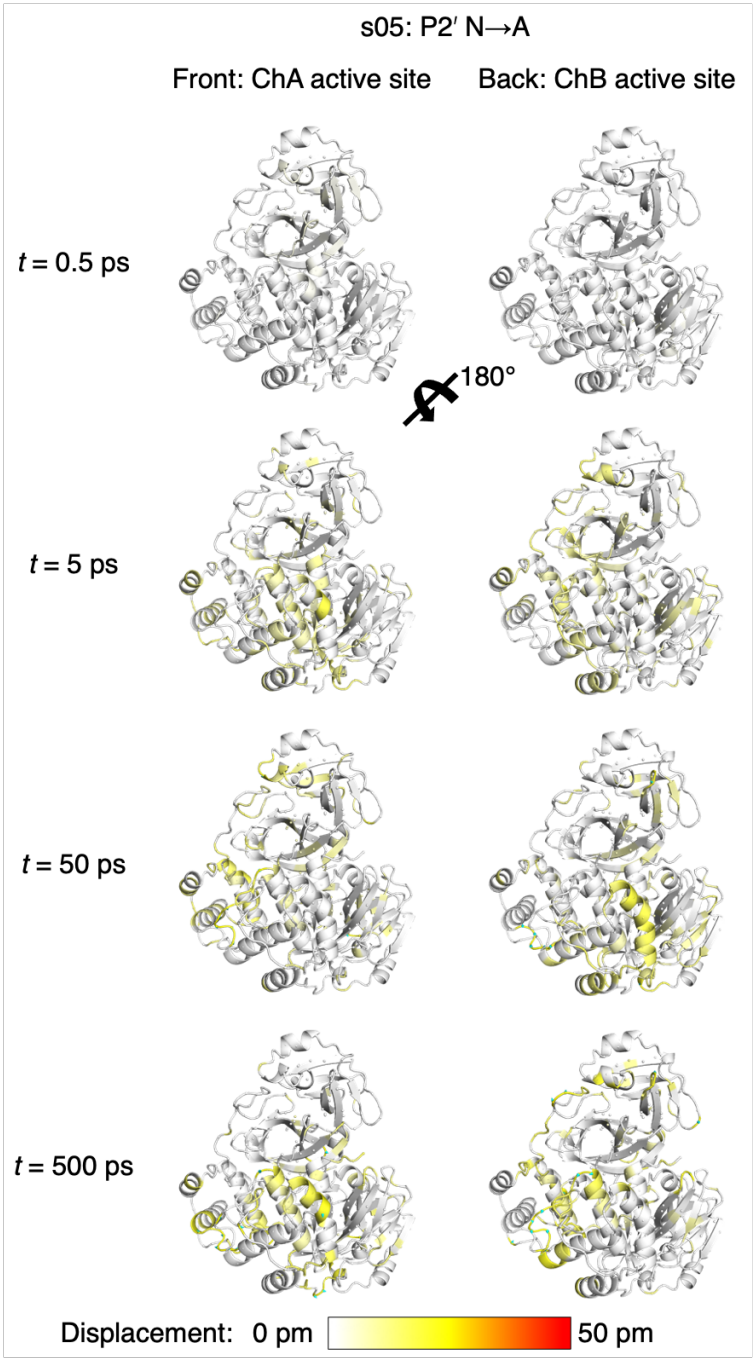


Figure S5.5B.5: Views of the M^{pro} response to the P2′ mutation in s05 from averaging Cα displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as cyan arrows with a scale-up factor of 5.[7]

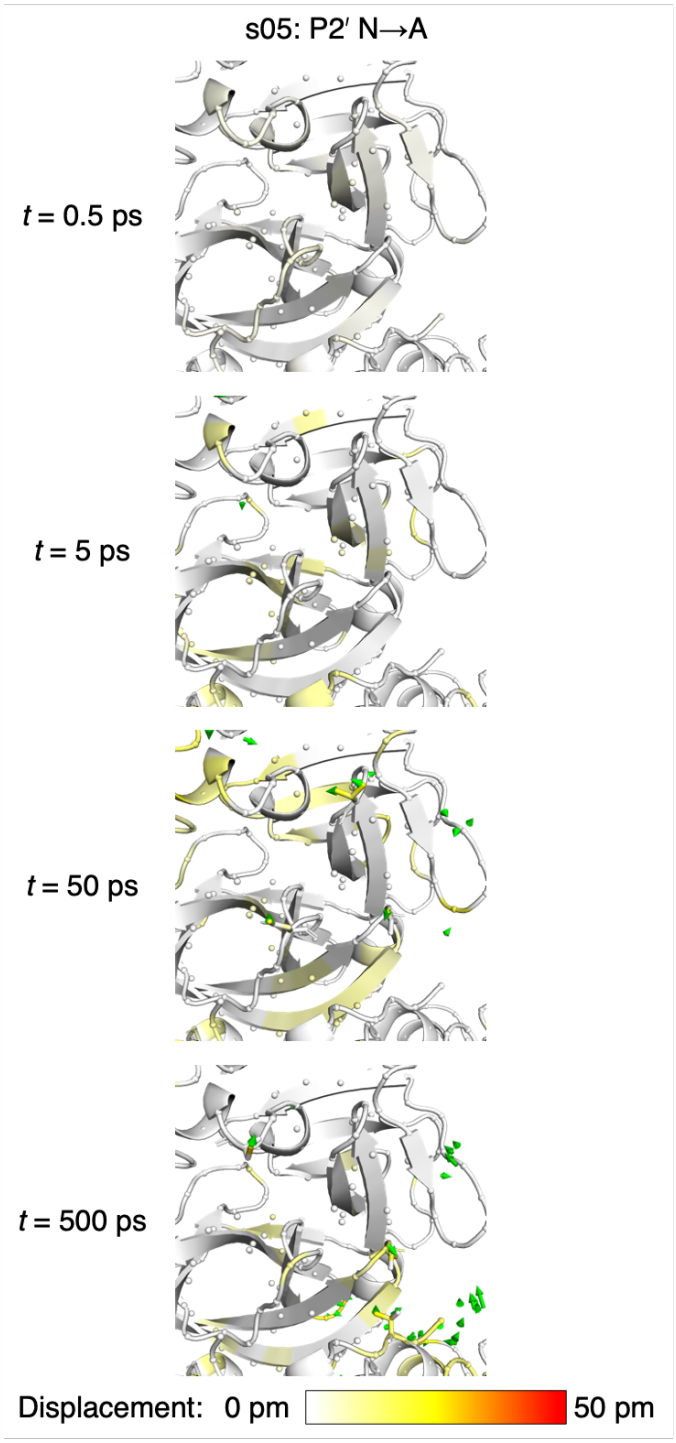


Figure S5.5B.6: Views of the M^{pro} response to the P2′ mutation in s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations, with a focus on the S2′ subsite. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7] Residues that show such significant displacements and that are within 10 Å of the substrate P2′ residue are shown as sticks.

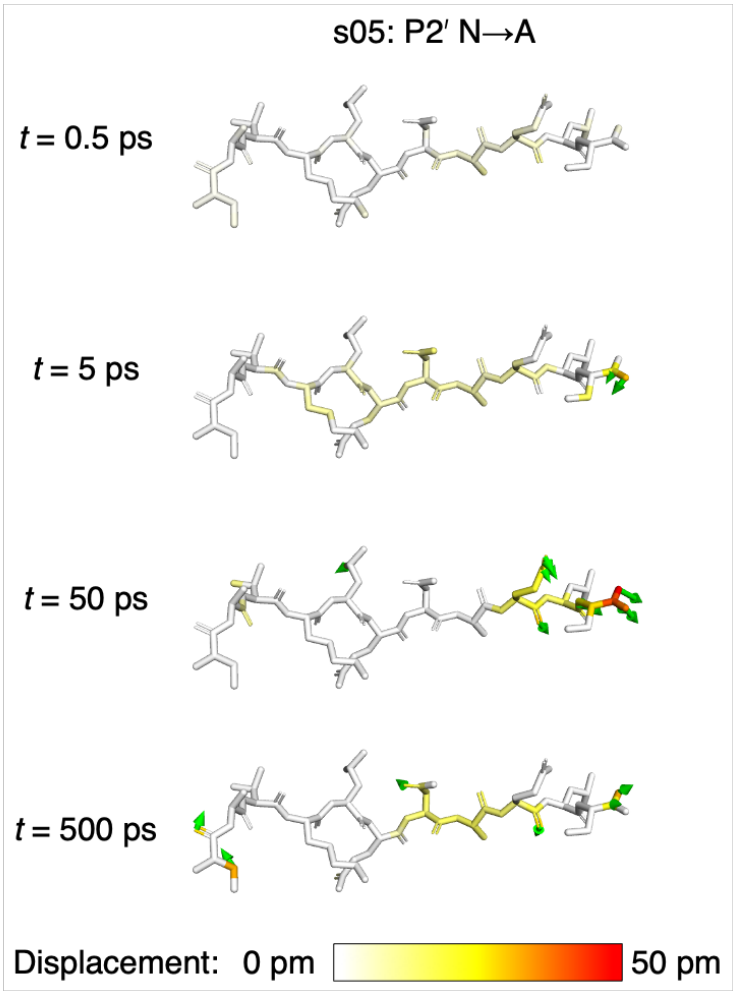


Figure S5.5B.7: Views of the peptide response to the P2′ mutation in s05 from averaging non-hydrogen atom displacement vectors, shown using the structure prior to MD simulations. Displacement magnitudes are represented on a white-yellow-red scale.[6] Significant vectors with length ≥20 pm are displayed as green arrows with a scale-up factor of 5.[7]

Section S5.6: Supplementary Methods

Table S5.6.1: Assignment of ionic and protonation states of titratable residues of SARS-CoV-2 M^{pro} (PDB 6YB7)[3] employed in this study. The setup is identical to that previously reported.[1, 2] Neutral histidine residues that are Nδ-protonated and Nε-protonated are referred to as "HID" and "HIE" respectively, according to AMBER force field nomenclature.[8] None of the N-terminal (NT) and C-terminal (CT) in M^{pro} and peptides are capped.

Residue	LYS (+1)	ARG (+1)	ASP (-1)	GLU (-1)	HID (0)	HIE (0)	NT (+1)	CT (-1)	**TOTAL**
ChA/ChB	5	4	33	14	41	64	1	306	
	12	40	34	47	80	163			
	61	60	48	55		164			
	88	76	56	166		172			
	90	105	92	178		246			
	97	131	153	240					
	100	188	155	270					
	102	217	176	288					
	137	222	187	290					
	236	279	197						
	269	298	216						
			229						
			245						
			248						
			263						
			289						
			295						
Charge	**+11**	**+11**	**-17**	**-9**	**0**	**0**	**+1**	**-1**	**-4**
Peptides									
s01	11	10					1	11	
Charge	**+1**	**+1**					**+1**	**-1**	**+2**
s02	10	11					1	11	
Charge	**+1**	**+1**					**+1**	**-1**	**+2**
p12	1						1	11	
Charge	**+1**						**+1**	**-1**	**+1**
s01-QP1A	11	10					1	11	
Charge	**+1**	**+1**					**+1**	**-1**	**+2**
s05-QP1A	4			9			1	11	
Charge	**+1**			**-1**			**+1**	**-1**	**0**

References

1. Chan, H. T. H.; Moesser, M. A.; Walters, R. K.; Malla, T. R.; Twidale, R. M.*, et al.* Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem. Sci.* **2021,** *12*, 13686.
2. Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **2023,** *3*, 1767.
3. Owen, C. D.; Lukacik, P.; Strain-Damerell, C. M.; Douangamath, A.; Powell, A. J.*, et al.* COVID-19 main protease with unliganded active site. *To be published* **2020,** *PDB 6YB7*, doi: 10.2210/pdb6yb7/pdb.
4. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.*, et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015,** *43*, D364.
5. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577.
6. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 2.3.0.*
7. Law, S. M. modevectors.py. https://pymolwiki.org/index.php/Modevectors (accessed 2022-06-30).
8. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.

Appendix for Chapter 6

(Supplementary Information to Manuscript IV)

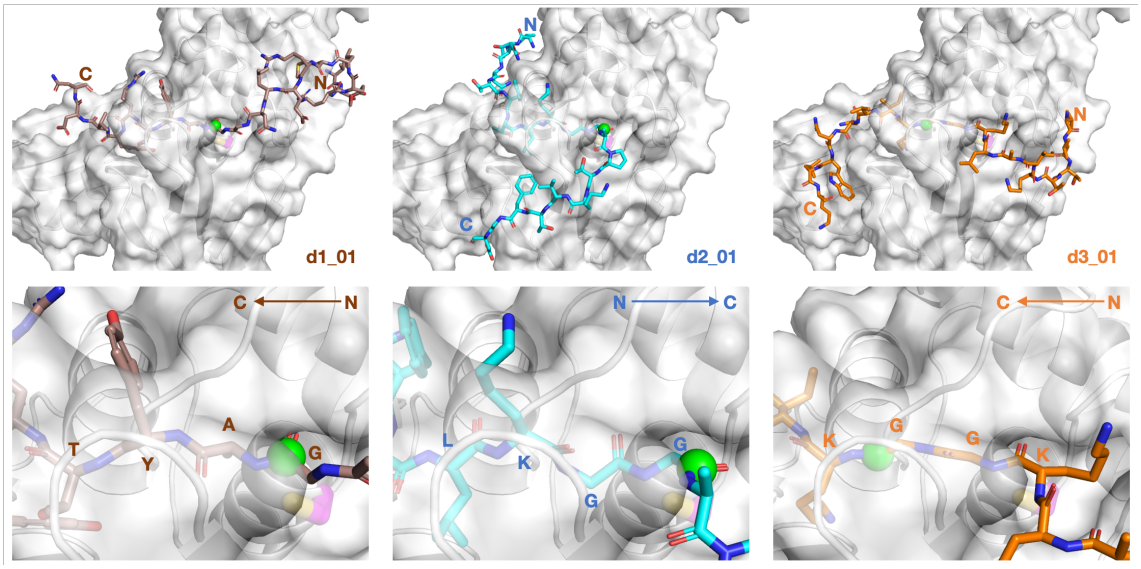


Figure S6.1: Top-ranked binding poses from AutoDock CrankPep (ADCP)[1] for the non-structural protein (nsp) peptides **1**, **2**, and **3**,[2] with the N- and C-terminals labelled. The P1 scissile amide carbonyl carbon is shown as a green sphere. The bottom row shows magnified views around the S4-S1 subsites, with each residue labelled and the N-/C-terminal peptide direction indicated.

Table S6.1: Deviations of P4-P1 Cα atom positions (compared to those in VIR251 in PDB 6WX4)[3] and root mean square deviation (RMSD) values for the 100 top-ranked solutions from ADCP[1] docking of peptides **1**, **2**, and **3**. Solutions that pass the filter (all four deviations within 2 Å) are in green, while reverse-bound solutions for peptide **3** that place Gly-Gly in S2-S1 are in orange.

Peptide 1							*Peptide 2*							*Peptide 3*					
Ranking	Deviation (Å)				RMSD (Å)		Ranking	Deviation (Å)				RMSD (Å)		Ranking	Deviation (Å)				RMSD (Å)
	P4	*P3*	*P2*	*P1*				*P4*	*P3*	*P2*	*P1*				*P4*	*P3*	*P2*	*P1*	
1	17.719	13.064	7.576	1.417	11.662		1	0.555	0.638	0.616	0.445	0.568		1	11.898	7.809	1.303	6.506	7.851
2	13.666	8.103	7.947	9.937	10.177		2	0.546	0.646	0.614	0.383	0.557		2	23.181	16.380	10.563	4.065	15.279
3	18.191	12.249	5.673	0.914	11.335		3	13.315	8.287	3.650	4.068	8.304		3	21.375	17.295	14.571	11.297	16.552
4	17.483	13.424	7.936	1.724	11.745		4	18.696	18.812	14.792	13.679	16.653		4	20.953	15.725	19.831	18.326	18.811
5	6.922	0.564	0.830	0.933	3.528		5	23.861	20.525	16.843	12.695	18.944		5	29.626	27.569	22.286	20.224	25.216
6	5.533	0.709	0.507	0.679	2.821		6	0.625	0.544	0.682	0.756	0.656		6	16.900	17.037	16.321	17.637	16.980
7	31.863	29.764	30.688	30.875	30.807		7	12.635	8.459	8.616	12.230	10.666		7	15.259	8.595	3.408	4.180	9.162
8	32.268	25.838	25.227	19.290	26.064		8	37.973	33.416	31.510	27.160	32.746		8	34.033	32.487	30.714	26.474	31.056
9	17.040	10.391	4.744	2.624	10.340		9	1.239	1.065	0.387	0.476	0.872		9	27.685	26.231	23.336	22.612	25.052
10	25.450	22.260	19.441	14.266	20.765		10	29.630	26.500	23.873	20.509	25.351		10	40.175	35.169	31.132	25.347	33.402
11	17.323	11.897	5.589	0.513	10.876		11	13.272	11.405	10.346	9.681	11.258		11	27.886	26.856	27.162	26.899	27.204
12	17.102	17.723	14.568	10.983	15.325		12	21.599	18.559	18.474	15.576	18.674		12	15.469	9.031	3.766	3.816	9.349
13	16.175	9.664	4.350	3.021	9.786		13	28.438	23.705	20.353	17.767	22.916		13	17.990	16.932	13.953	12.753	15.554
14	28.948	26.041	23.593	21.042	25.078		14	4.552	2.358	2.622	2.189	3.080		14	20.601	20.131	19.251	20.858	20.219
15	27.717	25.926	25.155	22.993	25.504		15	17.537	11.482	9.744	8.085	12.244		15	23.907	17.900	18.867	20.651	20.459
16	1.399	1.056	0.932	0.820	1.074		16	24.056	20.174	16.827	17.105	19.757		16	15.972	10.648	4.278	2.913	9.941
17	18.706	20.327	15.567	15.243	17.591		17	30.482	30.402	30.337	31.269	30.625		17	18.113	18.376	15.694	13.355	16.510
18	21.105	19.892	19.762	18.792	19.905		18	17.712	17.224	11.983	9.072	14.459		18	28.422	23.714	23.635	21.246	24.393
19	29.355	30.862	30.857	31.482	30.649		19	27.797	21.113	16.537	12.024	20.227		19	24.405	21.608	18.356	15.256	20.200
20	6.181	0.631	0.771	0.980	3.169		20	30.334	29.271	25.621	20.723	26.753		20	15.195	8.625	3.605	3.970	9.138
21	31.736	27.538	22.098	20.509	25.857		21	38.124	33.812	32.604	27.433	33.212		21	29.275	27.440	22.585	20.852	25.273
22	31.833	27.764	22.741	21.052	26.194		22	11.333	10.431	9.675	9.663	10.298		22	27.867	27.591	23.014	22.175	25.294

23	34.470	32.255	32.393	31.838	32.755		23	20.064	14.913	14.285	11.854	15.569		23	23.167	19.542	17.246	15.462	19.073
24	27.746	25.320	21.586	20.650	23.996		24	12.906	11.650	5.525	1.711	9.162		24	22.629	18.852	17.926	13.329	18.483
25	30.216	30.142	26.237	24.589	27.904		25	25.053	22.424	20.733	15.158	21.155		25	44.617	41.349	39.960	36.192	40.642
26	20.123	18.924	13.763	15.361	17.237		26	23.032	19.377	15.844	11.336	17.927		26	44.700	41.208	39.550	35.504	40.376
27	31.402	26.795	24.208	21.814	26.296		27	41.061	38.721	38.606	34.456	38.285		27	22.662	16.161	10.385	4.213	15.003
28	29.973	27.894	24.003	25.343	26.902		28	29.117	28.875	25.548	19.656	26.080		28	23.907	18.682	15.821	13.244	18.346
29	29.966	27.525	24.618	25.643	27.015		29	34.970	27.981	25.388	18.620	27.373		29	22.391	15.797	13.637	11.300	16.314
30	29.860	32.634	30.199	29.803	30.646		30	36.073	29.573	28.459	24.032	29.847		30	32.843	30.974	32.785	32.223	32.215
31	15.447	9.754	13.325	11.967	12.792		31	21.790	17.449	15.413	15.124	17.646		31	22.878	21.267	20.096	21.467	21.450
32	32.234	25.686	24.249	18.890	25.708		32	17.676	17.419	15.769	11.733	15.829		32	24.833	22.145	19.383	14.554	20.583
33	19.313	18.175	16.090	11.954	16.622		33	29.421	25.649	22.888	18.467	24.435		33	23.664	23.184	27.114	28.236	25.641
34	31.060	26.630	23.550	21.756	25.990		34	18.540	18.569	15.021	13.975	16.654		34	22.422	22.828	23.759	26.057	23.808
35	30.071	27.136	24.316	23.089	26.292		35	37.315	31.198	27.664	21.703	30.008		35	24.498	25.178	27.726	29.359	26.762
36	29.439	30.119	26.938	23.384	27.597		36	19.309	19.874	14.822	13.717	17.144		36	22.776	16.403	10.496	3.984	15.115
37	28.470	23.456	20.084	14.413	22.203		37	33.199	27.787	25.667	20.789	27.227		37	13.372	10.599	8.571	8.791	10.511
38	10.893	4.367	3.239	8.895	7.539		38	0.745	0.526	0.701	1.004	0.763		38	15.813	14.702	13.782	14.830	14.799
39	34.381	31.994	32.021	31.557	32.507		39	13.785	8.636	4.172	3.511	8.578		39	35.034	35.680	34.739	32.458	34.499
40	22.689	23.255	25.439	27.057	24.672		40	28.610	25.173	22.788	19.112	24.170		40	35.552	33.192	32.097	29.156	32.580
41	21.999	20.491	16.194	14.350	18.520		41	21.853	16.738	17.953	14.406	17.941		41	17.951	18.080	16.320	13.311	16.527
42	29.654	29.093	24.277	23.912	26.865		42	24.896	20.004	16.804	11.509	18.939		42	29.979	25.653	23.150	20.742	25.115
43	29.046	25.650	22.535	22.643	25.110		43	35.004	31.941	30.647	27.939	31.486		43	15.998	9.970	4.367	2.993	9.790
44	17.333	12.783	7.424	1.329	11.410		44	23.816	18.964	13.084	7.623	17.001		44	19.676	17.881	15.871	14.505	17.096
45	29.782	30.492	27.242	24.646	28.135		45	11.978	8.283	8.695	12.517	10.540		45	33.535	34.767	33.957	30.967	33.337
46	2.977	2.419	3.897	4.738	3.618		46	34.266	27.318	24.928	18.255	26.810		46	18.557	19.095	17.362	16.749	17.965
47	23.776	17.510	14.650	7.907	16.949		47	28.640	26.898	28.080	27.713	27.840		47	20.068	18.848	15.834	14.661	17.490
48	18.357	12.796	6.210	1.131	11.625		48	9.581	9.353	8.949	8.396	9.081		48	9.365	3.661	9.214	15.616	10.367
49	28.459	25.941	22.612	22.700	25.047		49	19.449	18.647	13.456	9.572	15.801		49	18.545	17.061	14.079	12.946	15.818
50	28.902	27.319	27.084	26.634	27.498		50	22.239	18.108	16.285	15.849	18.295		50	28.853	30.215	29.093	26.412	28.677
51	16.950	12.619	7.377	1.610	11.220		51	9.066	9.094	10.232	15.039	11.133		51	16.569	9.669	10.577	9.529	11.945
52	18.827	18.476	15.425	15.600	17.155		52	20.329	17.853	15.926	13.272	17.042		52	15.162	8.471	3.408	4.337	9.111
53	22.520	20.156	14.798	11.219	17.736		53	24.660	19.423	20.589	17.433	20.695		53	14.336	10.410	5.476	2.397	9.349
54	17.635	12.795	7.484	2.020	11.563		54	26.931	23.806	22.163	19.554	23.268		54	22.998	19.905	15.490	10.235	17.817
55	6.025	1.049	1.360	0.606	3.147		55	28.746	28.796	25.555	20.933	26.205		55	19.235	19.814	16.979	12.397	17.353
56	34.977	34.766	32.347	32.463	33.661		56	18.878	17.061	16.788	12.578	16.489		56	11.885	7.879	1.406	6.353	7.837
57	23.732	23.631	23.334	23.154	23.464		57	21.119	18.738	18.268	18.582	19.210		57	20.482	15.989	11.219	7.434	14.631
58	19.883	18.503	15.127	15.560	17.383		58	27.911	32.009	27.883	25.605	28.446		58	10.578	4.250	9.663	13.476	10.062
59	20.828	19.422	15.715	11.400	17.234		59	29.455	26.473	23.649	20.074	25.153		59	24.914	22.227	19.383	15.904	20.877
60	24.589	23.006	24.288	24.204	24.029		60	22.431	21.058	22.253	21.483	21.813		60	22.795	15.683	9.992	3.958	14.841
61	21.450	18.219	16.360	13.440	17.609		61	17.976	15.879	12.612	9.718	14.394		61	22.172	22.355	24.973	26.153	23.974
62	19.482	21.221	15.798	14.975	18.053		62	0.762	0.717	0.910	0.753	0.789		62	24.748	21.858	22.862	20.350	22.511
63	28.894	28.875	27.866	26.015	27.937		63	36.155	30.113	28.964	25.422	30.410		63	12.526	7.426	8.893	10.875	10.117
64	22.148	16.510	14.849	15.612	17.517		64	26.073	22.619	23.296	23.210	23.837		64	14.472	11.588	9.870	8.401	11.310
65	30.362	33.225	30.485	27.293	30.414		65	29.145	25.869	23.702	19.528	24.808		65	18.018	17.833	12.184	10.912	15.085
66	30.776	30.913	26.655	24.429	28.329		66	0.673	0.699	0.662	0.965	0.760		66	18.800	18.891	15.754	12.679	16.728
67	10.358	10.700	16.707	21.025	15.354		67	24.837	18.770	18.748	16.523	19.960		67	25.175	20.676	17.202	13.105	19.551
68	22.208	21.067	18.147	15.014	19.312		68	28.037	22.982	24.016	19.861	23.903		68	20.348	15.872	16.241	13.253	16.624
69	29.541	25.938	22.445	23.849	25.583		69	0.954	0.818	0.796	0.585	0.800		69	16.008	11.465	11.262	13.937	13.311
70	28.591	24.011	21.080	15.032	22.717		70	29.292	27.457	25.633	24.716	26.832		70	23.274	19.565	17.454	15.895	19.247
71	22.100	20.028	19.769	19.011	20.260		71	16.651	13.998	13.291	12.763	14.255		71	21.379	18.088	16.459	13.842	17.655
72	18.897	14.300	9.538	10.896	13.886		72	23.389	17.825	15.693	12.069	17.725		72	13.018	7.663	6.783	9.846	9.633
73	30.024	27.437	23.294	22.186	25.928		73	23.414	20.060	16.998	15.048	19.144		73	22.374	17.668	12.173	13.009	16.809
74	16.973	13.134	7.715	1.772	11.437		74	35.819	30.127	29.447	25.548	30.457		74	22.923	19.600	15.574	10.362	17.745
75	21.811	17.673	14.451	12.799	17.035		75	0.616	0.929	1.324	1.358	1.100		75	15.969	10.744	4.330	2.964	9.974
76	22.697	16.972	16.652	17.556	18.633		76	36.136	37.749	34.738	32.044	35.229		76	22.043	20.931	17.130	13.422	18.692
77	14.722	11.995	11.241	13.165	12.848		77	35.228	31.906	31.795	28.447	31.934		77	29.216	25.742	23.515	18.497	24.552
78	20.211	19.284	16.563	15.319	17.954		78	33.054	34.975	33.758	31.149	33.263		78	11.896	8.710	13.649	15.934	12.821
79	7.662	8.211	13.077	18.129	12.508		79	19.717	16.627	16.026	16.071	17.178		79	21.040	17.156	14.913	15.772	17.379
80	23.075	23.584	26.542	27.221	25.170		80	28.899	29.363	26.879	25.853	27.786		80	12.987	9.846	7.451	8.250	9.864
81	2.517	2.559	3.597	4.657	3.447		81	30.093	29.367	24.445	21.578	26.604		81	24.373	21.063	19.039	19.258	21.042
82	23.282	23.631	22.898	25.139	23.753		82	29.200	28.013	23.478	21.799	25.806		82	21.434	14.656	16.151	15.620	17.169
83	22.426	23.086	25.967	26.888	24.664		83	13.186	11.383	10.268	9.852	11.247		83	14.957	8.277	3.525	4.132	8.969
84	0.869	1.315	1.760	2.971	1.898		84	38.157	35.915	31.511	25.938	33.210		84	40.267	35.113	31.168	25.455	33.443
85	33.064	32.368	27.822	25.192	29.789		85	20.008	20.531	15.623	14.318	17.825		85	21.326	16.886	15.093	15.389	17.353
86	15.788	10.397	13.368	12.356	13.122		86	22.414	21.413	20.852	20.397	21.282		86	31.943	30.820	28.342	24.297	28.999
87	18.188	12.302	5.782	0.741	11.359		87	22.991	16.185	13.104	6.818	15.881		87	12.227	8.545	10.873	10.704	10.669
88	26.414	25.839	21.916	20.254	23.748		88	23.757	19.784	15.347	10.038	17.973		88	12.386	7.134	8.384	13.237	10.605
89	22.957	18.111	14.596	12.203	17.443		89	35.981	29.233	25.258	18.578	27.984		89	34.346	35.043	33.712	32.345	33.876
90	29.125	26.231	23.740	21.325	25.272		90	26.829	23.477	24.243	23.460	24.541		90	12.972	10.037	10.174	9.871	10.839
91	18.260	12.371	5.664	0.985	11.397		91	14.824	13.455	17.307	18.009	16.005		91	24.566	21.802	18.858	15.277	20.419
92	18.579	17.262	14.294	13.134	15.969		92	29.372	26.222	23.826	20.238	25.137		92	21.103	17.107	15.082	13.656	16.970
93	10.793	4.126	3.746	9.602	7.742		93	22.013	22.121	22.860	25.346	23.124		93	27.402	24.602	21.427	16.428	22.831
94	19.196	17.789	15.619	11.840	16.348		94	34.660	28.950	28.664	23.895	29.292		94	13.994	14.006	13.201	14.261	13.871
95	29.126	30.097	26.909	23.291	27.481		95	13.801	8.487	11.597	11.875	11.597		95	23.152	17.972	16.728	12.874	18.059
96	22.880	17.966	12.117	12.871	17.020		96	22.596	19.734	16.859	13.379	18.461		96	13.646	10.899	10.554	8.872	11.125
97	39.615	36.685	32.698	28.006	34.528		97	0.880	1.168	0.999	0.996	1.016		97	16.995	17.135	16.648	17.737	17.133
98	35.822	36.519	34.517	32.731	34.927		98	25.711	22.574	23.292	22.684	23.600		98	21.506	16.417	18.341	14.722	17.924
99	27.975	26.127	24.801	22.370	25.400		99	26.936	24.052	25.067	23.187	24.850		99	23.718	21.634	22.663	21.199	22.325
100	18.301	12.586	6.158	1.055	11.537		100	17.729	13.154	8.905	4.514	12.114		100	18.771	17.789	15.972	14.502	16.839

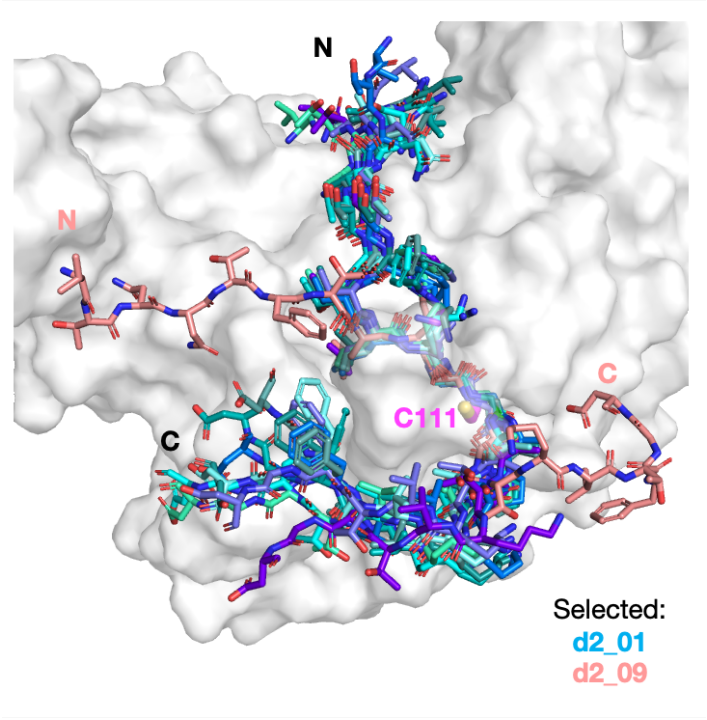


Figure S6.2: Docked conformations of peptide **2** that passed the applied filter, *i.e.*, the four Cα atoms of P4-P1 have to be within 2 Å of the corresponding Cα atoms of VIR251.[3] The poses include ranked solutions 01 (cyan), 02 (green-cyan), 06 (aquamarine), 09 (salmon), 38 (teal), 62 (deep teal), 66 (light teal), 69 (marine), 75 (slate), and 97 (purple blue). Except for d2_09, which adopts a unique extended conformation, d2_01 is representative of all the remaining poses.

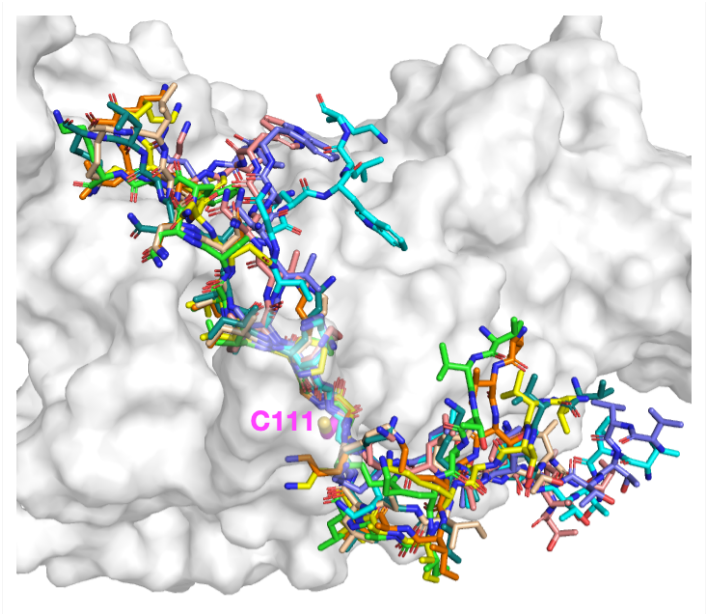


Figure S6.3: Docked reverse-bound poses of peptide **3**, which position the original P2-Gly residue in the S1 subsite, P1-Gly in the S2 subsite, P1′-Lys in the S3 subsite, and P2′-Ile in the S4 subsite. The poses include ranked solutions 07 (orange), 12 (green), 16 (cyan), 20 (yellow), 43 (salmon), 52 (teal), 75 (slate), and 83 (wheat). The highest ranked solution among them, d3_07, was selected as being representative and was used for subsequent modelling.

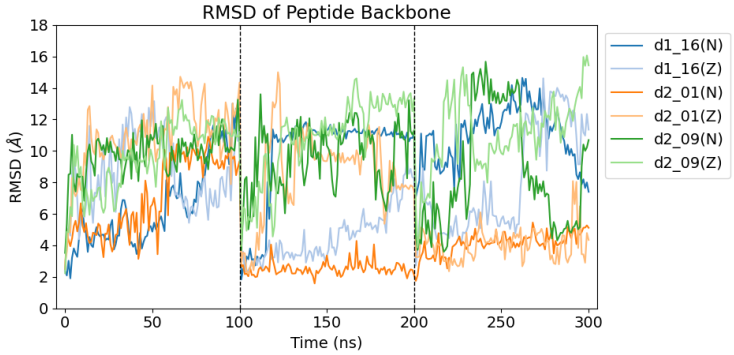


Figure S6.4: RMSD of peptide backbone atoms (N, Cα, C) in the combined 3 × 100 ns molecular dynamics (MD) trajectories (fitted based on PL^{pro} backbone) initiated from the d1_16, d2_01, and d2_09 poses in either the neutral (N) or zwitterionic (Z) state of Cys111-His272, relative to their corresponding starting peptide poses.

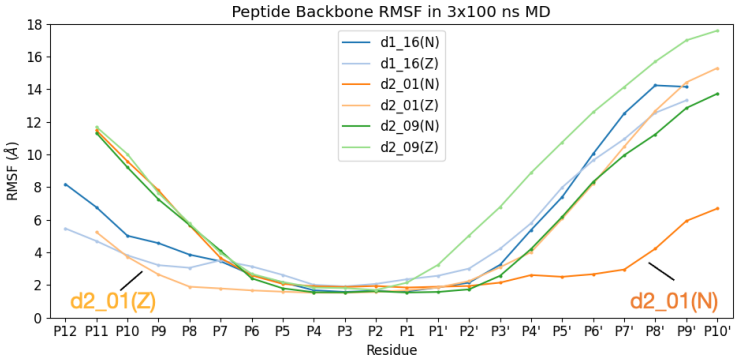


Figure S6.5: Per-residue root mean square fluctuation (RMSF) of peptide backbone atoms in the combined 3 × 100 ns MD (fitted based on PL^{pro} backbone) initiated from the d1_16, d2_01, and d2_09 poses in either the N or Z state of Cys111-His272. Note that the C-terminal NH_2 group is treated as a separate residue.

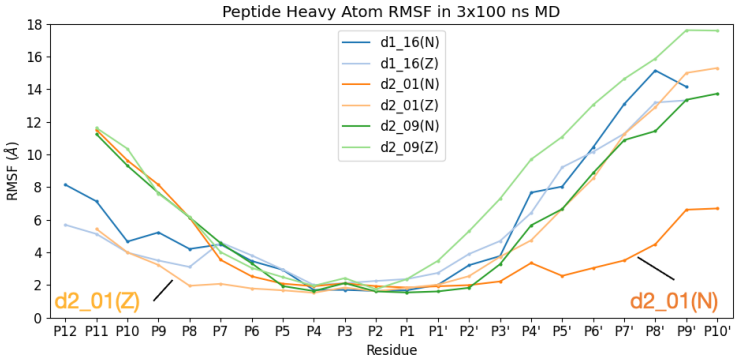


Figure S6.6: Per-residue RMSF of peptide non-hydrogen atoms in the combined 3 × 100 ns MD (fitted based on PL^{pro} backbone) initiated from the d1_16, d2_01, and d2_09 poses in either the N or Z state of Cys111-His272.

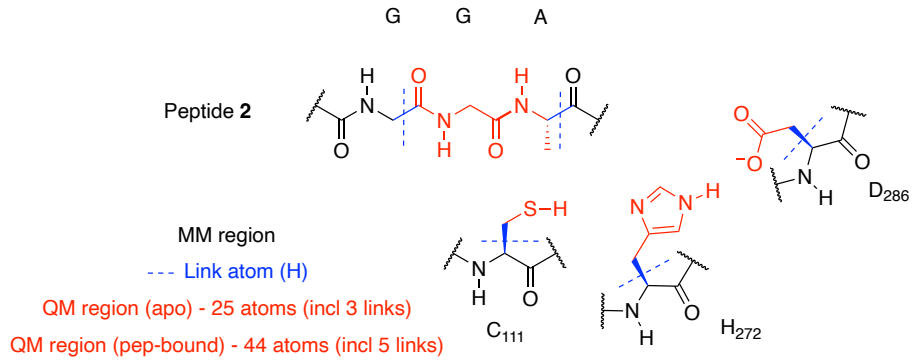


Figure S6.7: Definition of the quantum mechanical (QM) region in the quantum mechanics/molecular mechanics-umbrella sampling (QM/MM-US) calculations for proton transfer processes in the Cys111-His272-Asp286 catalytic triad.

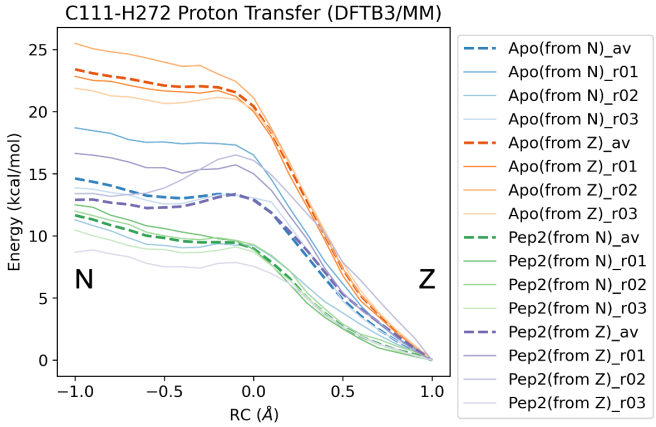


Figure S6.8: DFTB3/MM-US free energy profiles from the weighted histogram analysis method (WHAM)[4] for the proton transfer between Cys111 and His272, with reaction coordinate (RC) = -1 being the N state and +1 being the Z state. The energy profiles averaged from three replicas are shown as dotted lines.

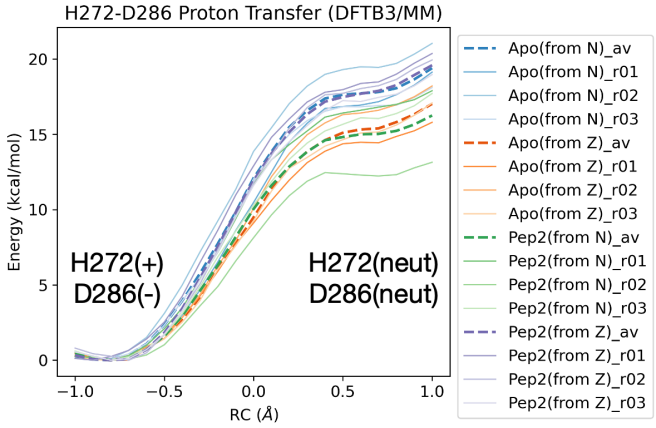


Figure S6.9: WHAM[4]-derived DFTB3/MM-US free energy profiles for the proton transfer between His272 and Asp286, between RC = -1 (doubly protonated His272 and deprotonated Asp286) and +1 (neutral His272 and neutral Asp286). The energy profiles averaged from three replicas are shown as dotted lines.

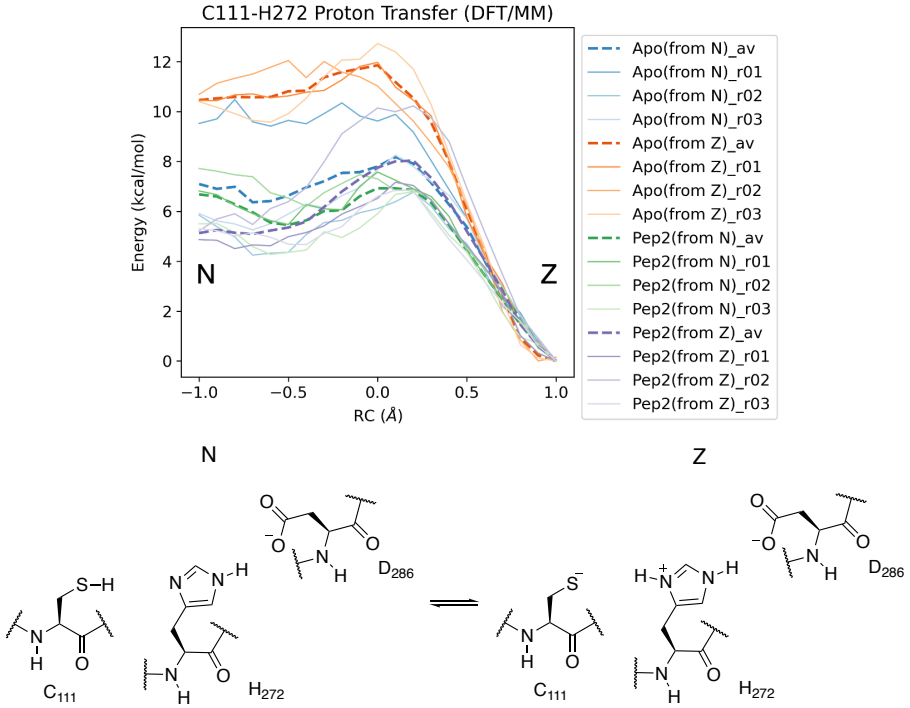


Figure S6.10: WHAM[4]-derived PBE0-D3BJ/6-31G(d)/MM-US free energy profiles for the proton transfer between Cys111 and His272, with RC = -1 being the N state and +1 being the Z state. The energy profiles averaged from three replicas are shown as dotted lines.

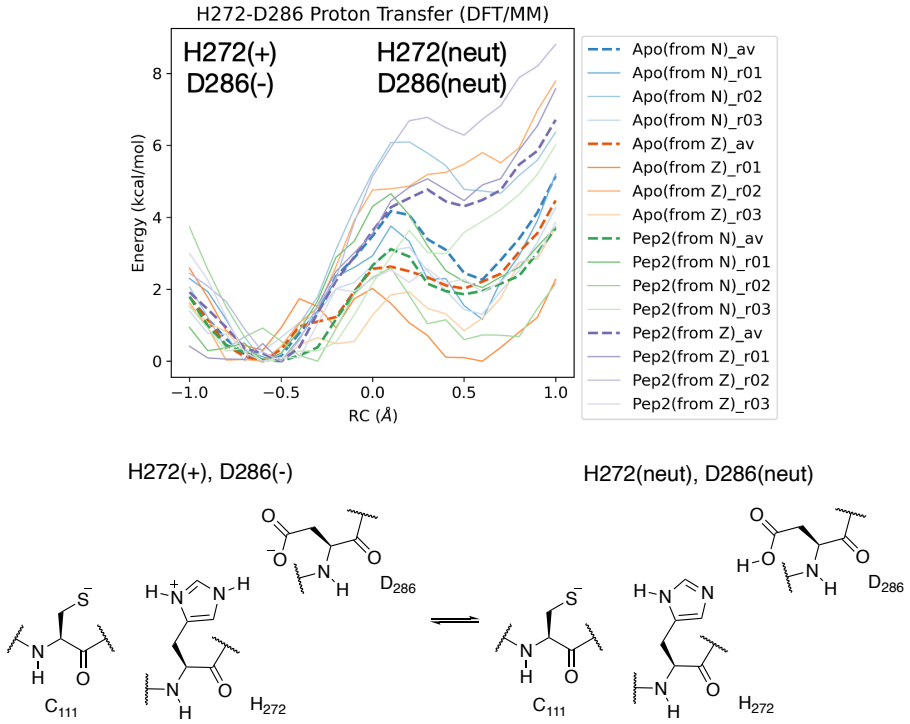


Figure S6.11: WHAM[4]-derived PBE0-D3BJ/6-31G(d)/MM-US free energy profiles for the proton transfer between His272 and Asp286, between RC = -1 (doubly protonated His272 and deprotonated Asp286) and +1 (neutral His272 and neutral Asp286). The energy profiles averaged from three replicas are shown as dotted lines.

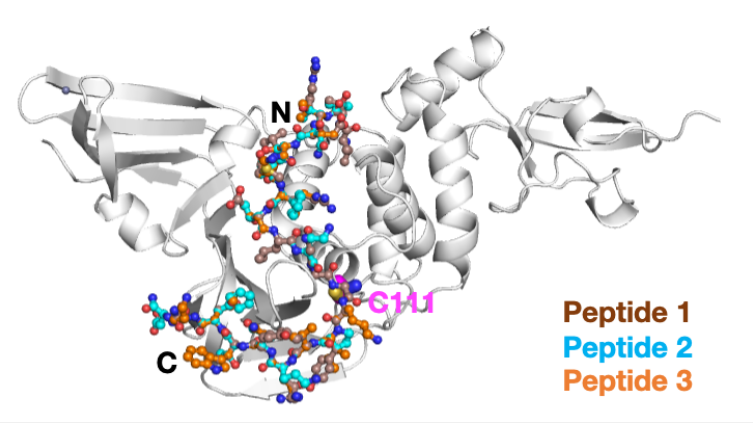


Figure S6.12: Starting conformations of PL^{pro} complexed with each of the three nsp oligopeptides **1**, **2**, and **3**, based on the d2_01 conformation.

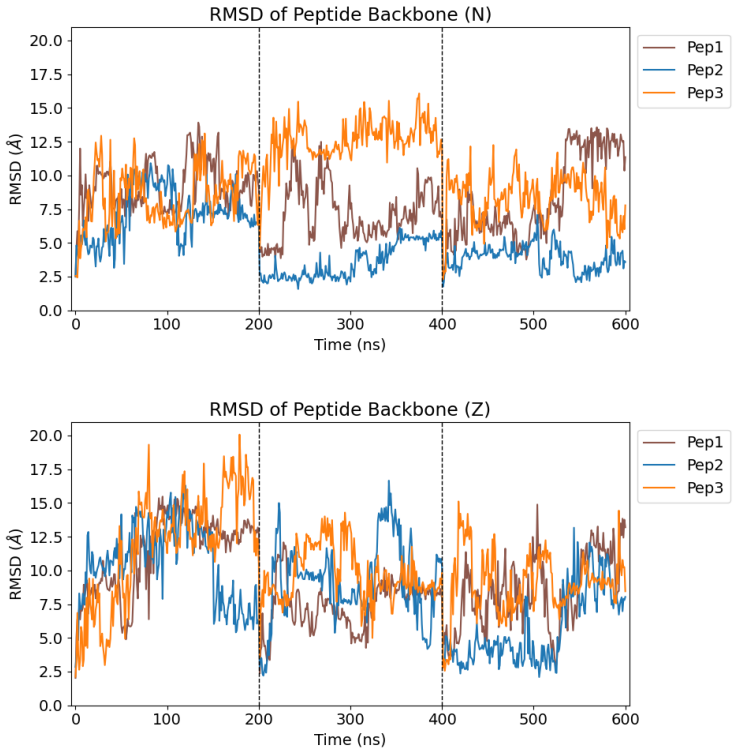


Figure S6.13: RMSD of peptide backbone atoms in the combined 3 × 200 ns MD trajectories (fitted based on PL^{pro} backbone) for the three nsp peptides initiated from the d2_01 conformation, in the (top) N and (bottom) Z states of Cys111-His272, relative to their corresponding starting conformations.

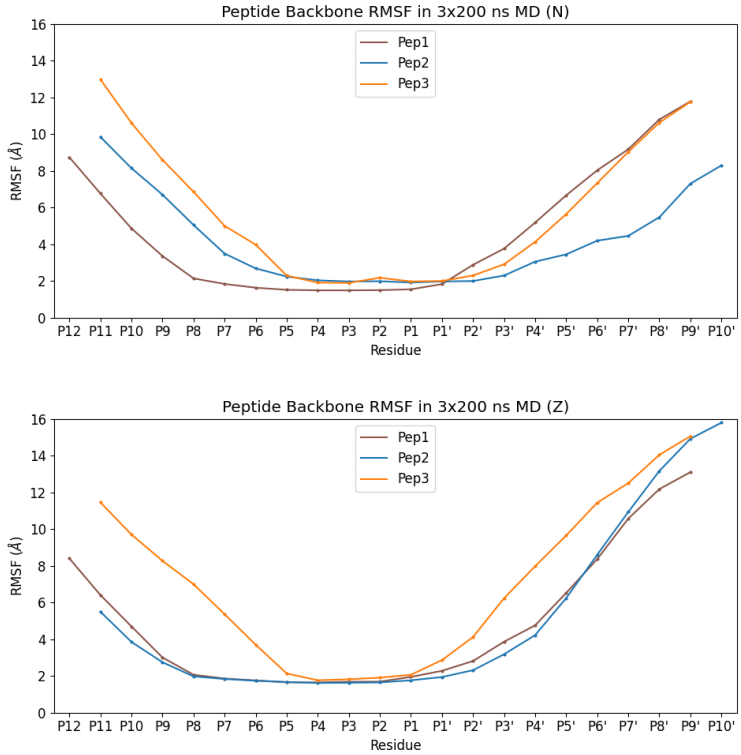


Figure S6.14: Per-residue RMSF of peptide backbone atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for the three nsp peptides initiated from the d2_01 conformation, in the (top) N and (bottom) Z states of Cys111-His272. Note that the C-terminal NH_2 group is treated as a separate residue.

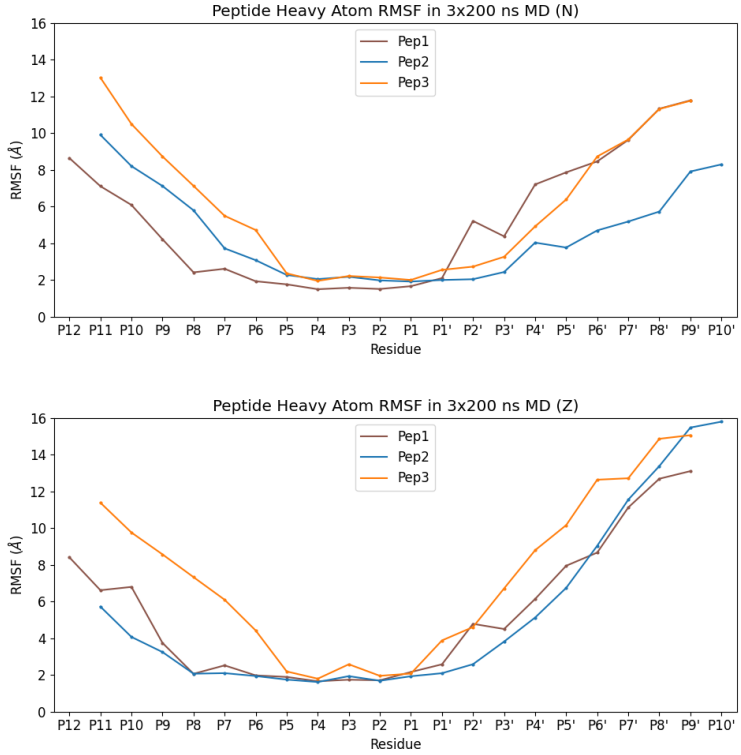


Figure S6.15: Per-residue RMSF of peptide non-hydrogen atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for the three nsp peptides initiated from the d2_01 conformation, in the (top) N and (bottom) Z states of Cys111-His272.

Table S6.2: PL^{pro}-peptide hydrogen bonds (HBs) of significance (occurrence ≥ 25% out of 600 frames analysed every ns) observed over the combined 3 × 200 ns MD for each of the three nsp peptides **1**, **2**, and **3**, all initiated from the d2_01 conformation, in the (left) N and (right) Z states of Cys111-His272. HB occurrences involving equivalent carboxylate oxygens (in Asp and Glu sidechains) are combined, while salt bridges involving carboxylate oxygens and Arg sidechain guanidino group are combined and halved to avoid double counting. Peptide residues are italicised, while PL^{pro} residues are not. Outside the core P5-P1′ region (see **Figure 6.3a**), N-terminal HBs are in blue and C-terminal HBs are in red.

Donor	Acceptor	%		Donor	Acceptor	%
Pep 1 (N)	P5-P1′=8-13			Pep 1 (Z)		
GLY12N	GLY271O	96		*GLU8N*	ASP164OD	97
GLY11N	GLY163O	95		*ASN10ND2*	TYR268O	96
GLU8N	ASP164OD	94		*GLY11N*	GLY163O	95
GLY163N	*GLY11O*	93		*LEU9N*	ASP164OD	83
ASN10ND2	TYR268O	91		GLY271N	*ASN10O*	80
LEU9N	ASP164OD	80		GLY163N	*GLY11O*	80
ARG166NH	*GLU8OE*	80		ARG166NH	*GLU8OE*	80
ASN10N	TYR264OH	74		*ASN10N*	TYR264OH	79
GLY271N	*ASN10O*	74		*GLY12N*	GLY271O	76
TRP106NE1	*GLY12O*	59		*ALA13N*	GLY271O	48
ALA13N	GLY271O	36		*THR15OG1*	ASP286OD	35
ARG3NH	GLU203OE	32		*LEU5N*	TYR171OH	28
ARG7NH	GLU161OE	25		*ARG3NH*	GLU203OE	25
Pep 2 (N)	P5-P1′=7-12			Pep 2 (Z)		
GLY11N	GLY271O	95		*THR7N*	ASP164OD	100
TRP106NE1	*GLY11O*	95		*GLY10N*	GLY163O	94
GLY163N	*GLY10O*	92		*LEU8N*	ASP164OD	86
GLY10N	GLY163O	92		GLY271N	*LYS9O*	80
THR7N	ASP164OD	70		GLY163N	*GLY10O*	74
GLY266N	*VAL16O*	69		*LYS9N*	TYR264OH	68
THR265OG1	*THR14O*	68		*GLY11N*	GLY271O	66
LEU8N	ASP164OD	64		*ASN4ND2*	GLN174OE1	58
GLY271N	*LYS9O*	56		*ASN4N*	TYR171OH	45
LYS9N	TYR264OH	44		HIP272ND1	*GLY11O*	42
GLN250N	*GLY19O*	41		*ALA12N*	GLY271O	39
ALA12N	GLY271O	36		ASN109ND2	*GLY11O*	35
ARG166NH	*THR7OG1*	33		GLN174NE2	*ASN4OD1*	33
GLN250NE2	*PHE18O*	28		ARG166NH	*THR7OG1*	29
ASN4ND2	GLN174OE1	26		GLN174NE2	*THR2O*	27
Pep 3 (N)	P5-P1′=7-12			Pep 3 (Z)		
TRP106NE1	*GLY11O*	96		*GLY10N*	GLY163O	88
GLY163N	*GLY10O*	94		*GLY11N*	GLY271O	72
GLY11N	GLY271O	94		GLY163N	*GLY10O*	58
GLY10N	GLY163O	92		GLY271N	*LYS9O*	38
GLY271N	*LYS9O*	60		*LYS12N*	GLY271O	34
LYS9N	TYR268O	35		*LYS9N*	TYR264OH	32
LEU8N	ASP164OD	28		*LYS9N*	TYR268O	28
ALA7N	ASP164OD	28				

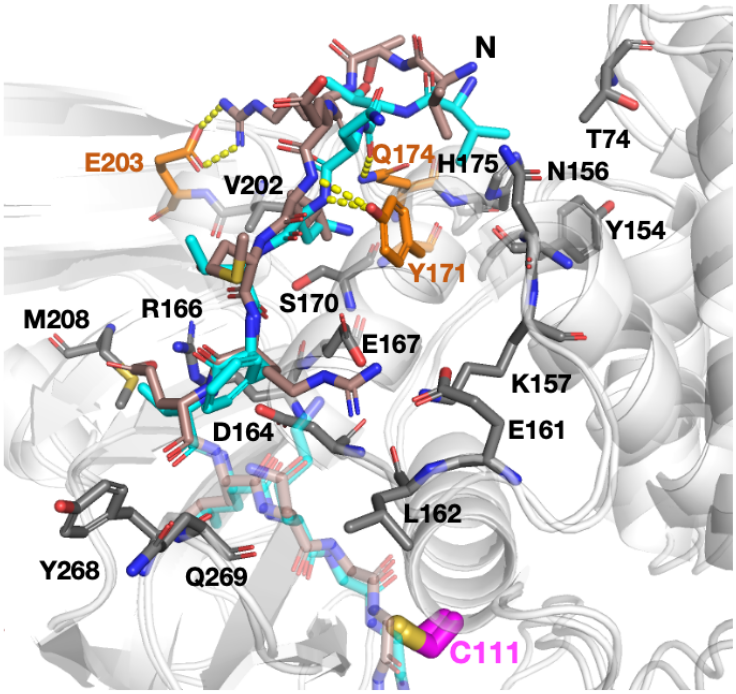


Figure S6.16: View of clustering-derived representative structures of Z-state PL^{pro} (white cartoon) complexed with peptides **1** (brown) or **2** (cyan), with a focus on the interactions in the N-terminal region outside the consensus P4-P1 sequence. PL^{pro} residues involved in HBs (yellow dotted lines; see **Table S6.2**; not all significant HBs are present in these snapshots) are shown as orange sticks. Other residues in close contact (within 4 Å) with the peptide residues N-terminal of P4 are in grey. The active site Cys111 is shown as magenta sticks. Clustering was performed using a 3 Å RMSD cut-off on the peptide non-hydrogen atoms using *gmx cluster* (gromos algorithm).[5, 6] Out of 601 frames extracted every ns from the combined 3 × 200 ns MD (including t = 0), the number of (multimembered) clusters obtained for peptides **1**, **2**, and **3** are respectively 291 (81), 233 (76), and 361 (77).

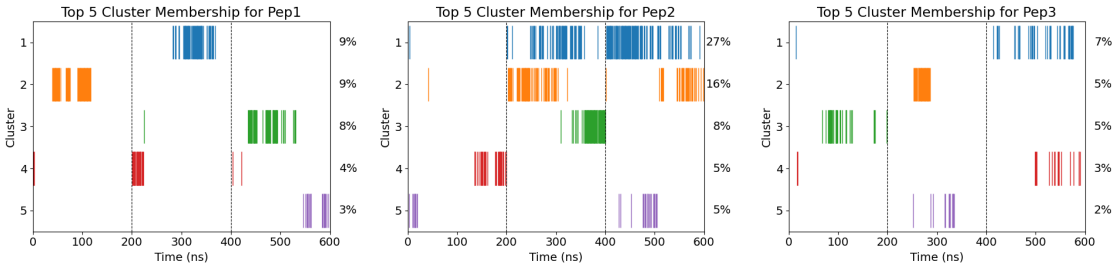


Figure S6.17: Cluster membership of frames extracted every ns from the combined 3 × 200 ns MD (fitted using PL^{pro} backbone) of each of the three PL^{pro}-peptide (N state) complexes, with clustering performed using a 3 Å RMSD cut-off on the peptide backbone atoms (N, Cα, C) using *gmx cluster* (gromos algorithm).[5, 6] Only the 5 clusters with the highest occupancy percentages are shown. Out of 601 frames, the number of (multimembered) clusters obtained for peptides **1**, **2**, and **3** are respectively 157 (73), 66 (41), and 201 (77).

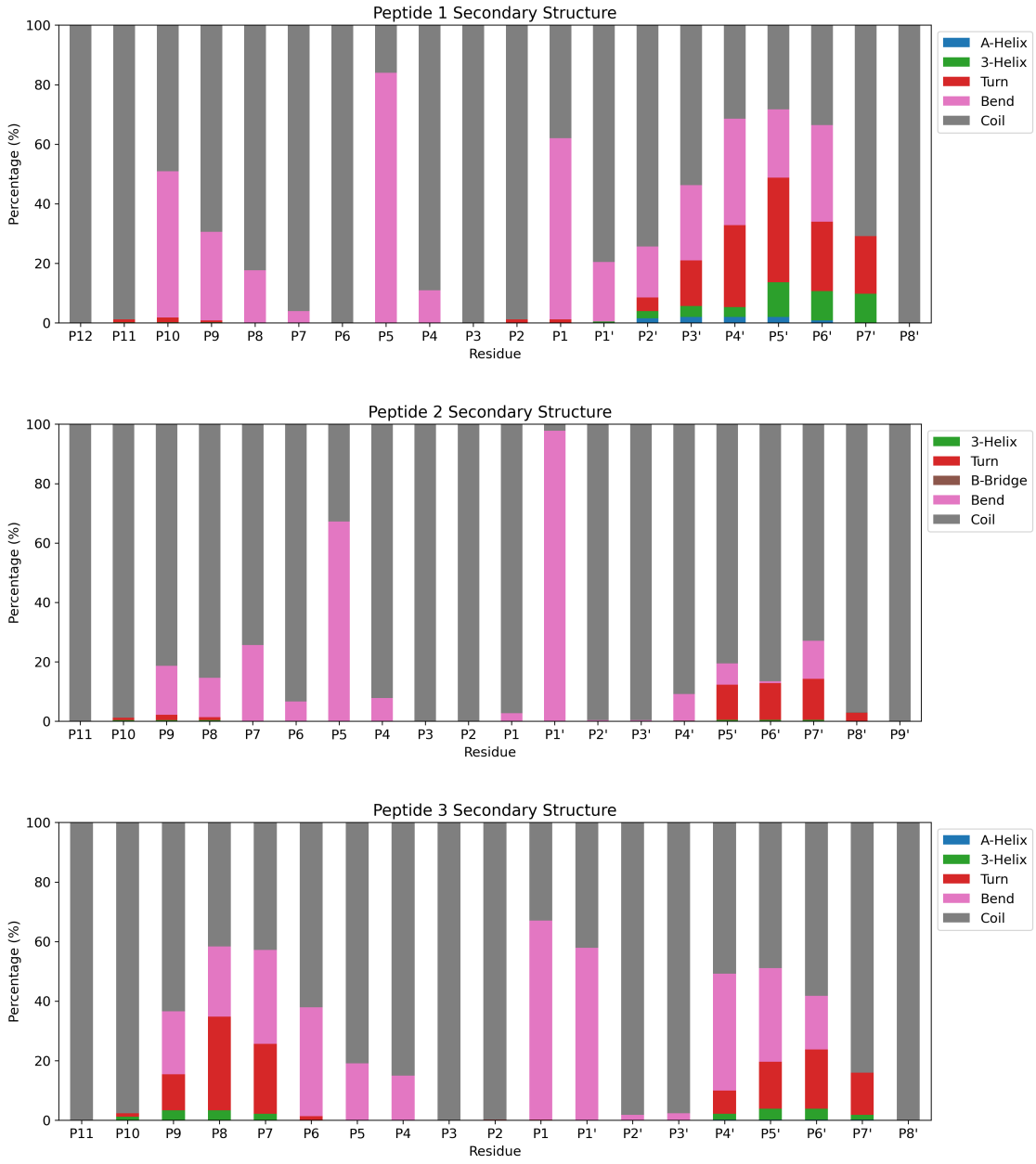


Figure S6.18: Proportion of secondary structure adopted by every peptide residue over the combined 3 × 200 ns MD of PL^{pro}-peptide (N state) complexes, assigned by DSSP (version 2.0.4).[7, 8] Frames were analysed every ns.

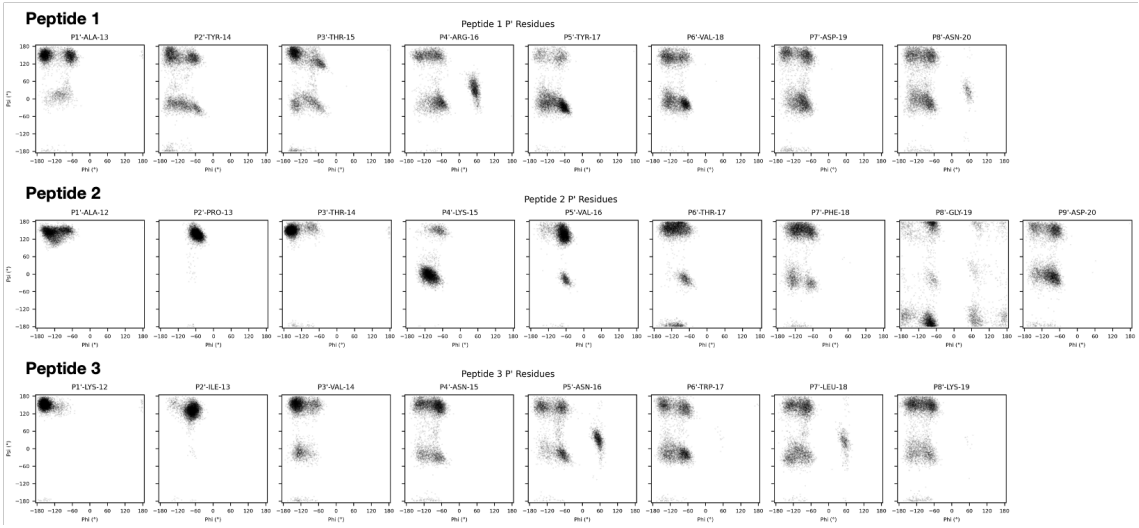


Figure S6.19: Ramachandran plots of every peptide P′ residue over the combined 3 × 200 ns MD of PL^{pro}-peptide (N state) complexes. Frames were analysed every 0.1 ns.

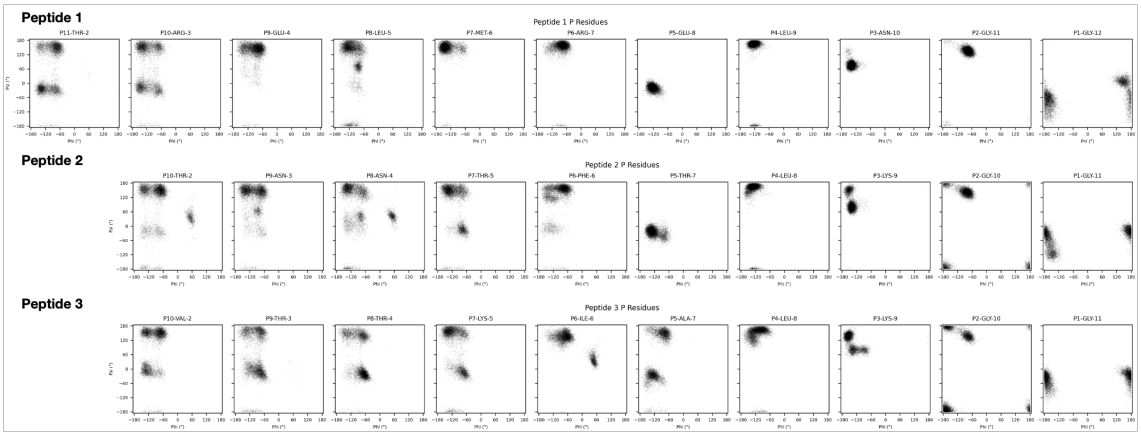


Figure S6.20: Ramachandran plots of every peptide P residue over the combined 3 × 200 ns MD of PL^{pro}-peptide (N state) complexes. Frames were analysed every 0.1 ns.

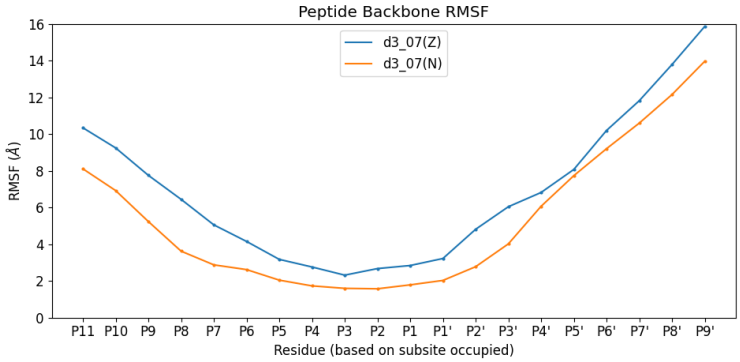


Figure S6.21: Per-residue RMSF of peptide backbone atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for peptide **3** initiated from the d3_07 conformation, in the N (orange) and Z (blue) states of Cys111-His272. Note that the C-terminal NH_2 cap is treated as a separate residue (P11).

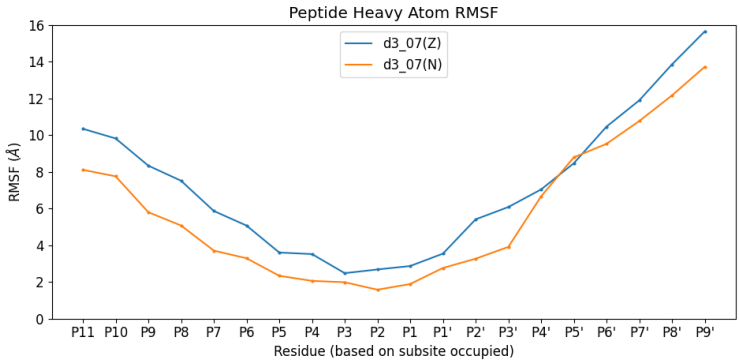


Figure S6.22: Per-residue RMSF of peptide non-hydrogen atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for peptide **3** initiated from the d3_07 conformation, in the N (orange) and Z (blue) states of Cys111-His272.

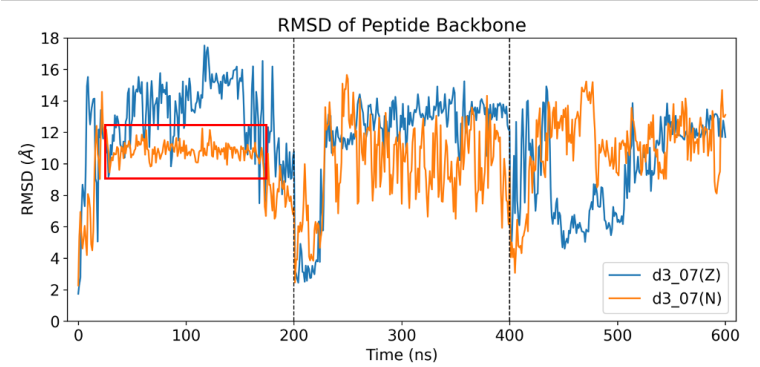


Figure S6.23: RMSD of peptide backbone atoms in the combined 3 × 200 ns MD trajectories (fitted based on PL^{pro} backbone) for peptide **3** initiated from the d3_07 conformation, in the N (orange) and Z (blue) states of Cys111-His272, relative to the starting conformation. The period of stable binding in the N state is indicated by a red box.

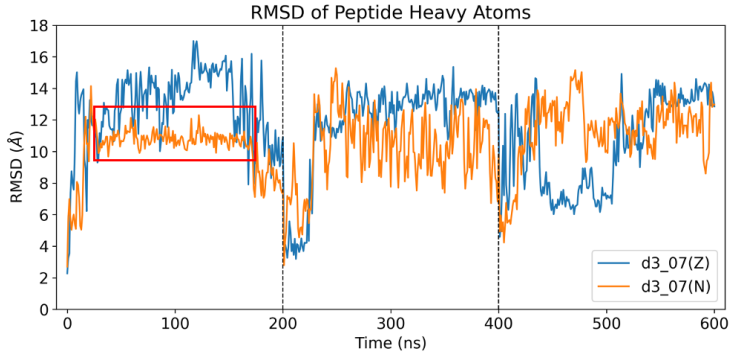


Figure S6.24: RMSD of peptide non-hydrogen atoms in the combined 3 × 200 ns MD trajectories (fitted based on PL^{pro} backbone) for peptide **3** initiated from the d3_07 conformation, in the N (orange) and Z (blue) states of Cys111-His272, relative to the starting conformation. The period of stable binding in the N state is indicated by a red box.

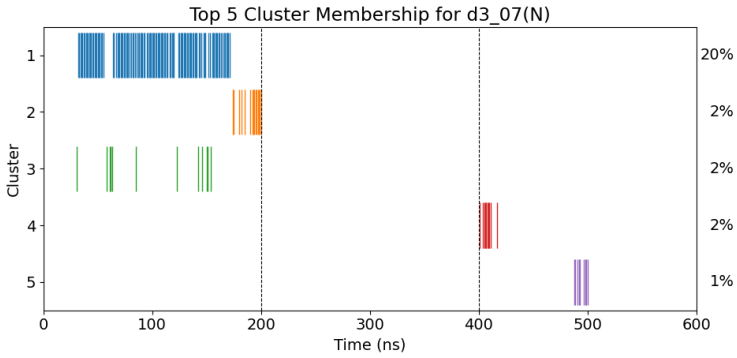


Figure S6.25: Cluster membership of frames extracted every ns from the combined 3 × 200 ns MD (fitted using PL^{pro} backbone) of PL^{pro} complexed with d3_07 (N state), with clustering performed using a 3 Å RMSD cut-off on the peptide non-hydrogen atoms using *gmx cluster* (gromos algorithm).[5, 6] Only the 5 clusters with the highest occupancy percentages are shown. Cluster 1, which has the highest population, corresponds to the period of stable binding shown in **Figures S6.23-24**. The complex structure at t = 159 ns is determined to be the representative middle structure for cluster 1, as defined by the smallest average RMSD to all other structures in the same cluster. Out of 601 frames, the number of (multimembered) clusters obtained is 344 (63).

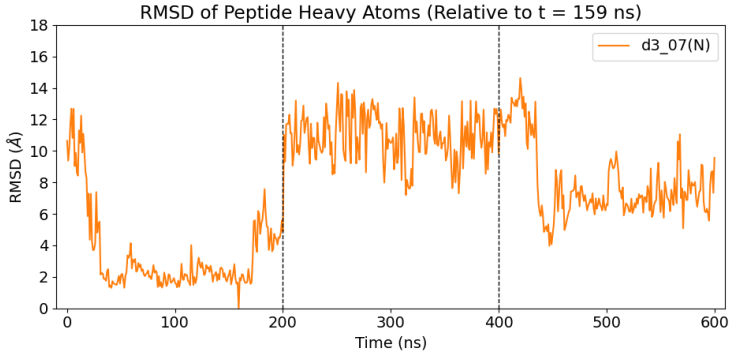


Figure S6.26: RMSD of peptide non-hydrogen atoms in the combined 3 × 200 ns MD trajectories (fitted based on PL^{pro} backbone) for peptide **3** initiated from the d3_07 conformation in the N state, relative to the t = 159 ns frame which is representative of cluster 1 (see **Figure S6.25**).

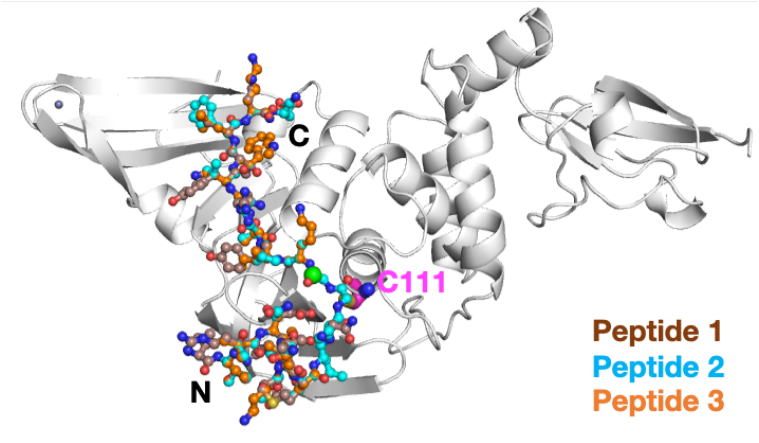


Figure S6.27: Starting conformations of PL^{pro} complexed with the three nsp oligopeptides **1**, **2**, and **3** in the reverse active site binding mode.

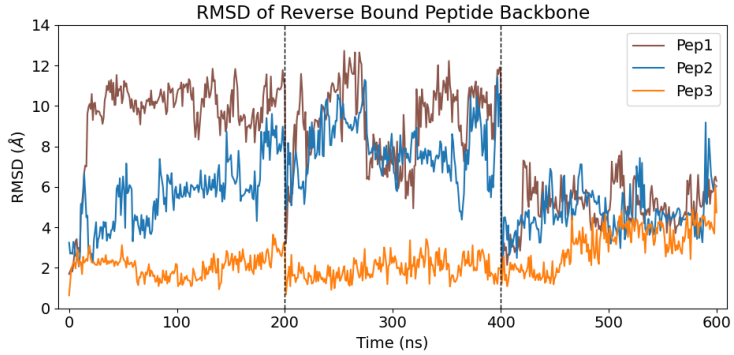


Figure S6.28: RMSD of peptide backbone atoms in the combined 3 × 200 ns MD trajectories (fitted based on PL^{pro} backbone) for the three reverse-bound nsp peptides, in the Z state of Cys111-His272, relative to their corresponding starting conformations.

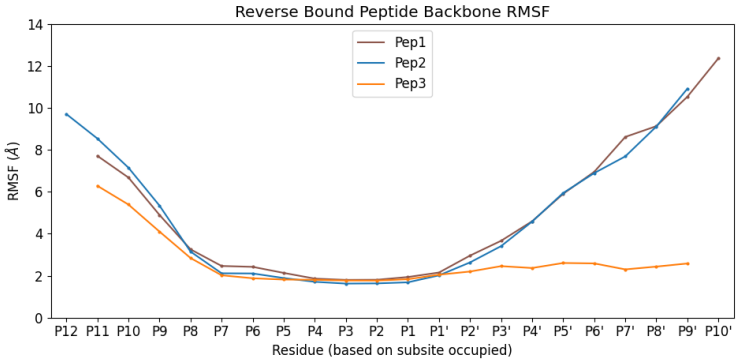


Figure S6.29: Per-residue RMSF of peptide backbone atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for the three reverse-bound nsp peptides, in the Z state of Cys111-His272. Note that the C-terminal NH_2 cap is treated as a separate residue (P12 for peptide **2**; P11 for peptides **1** and **3**).

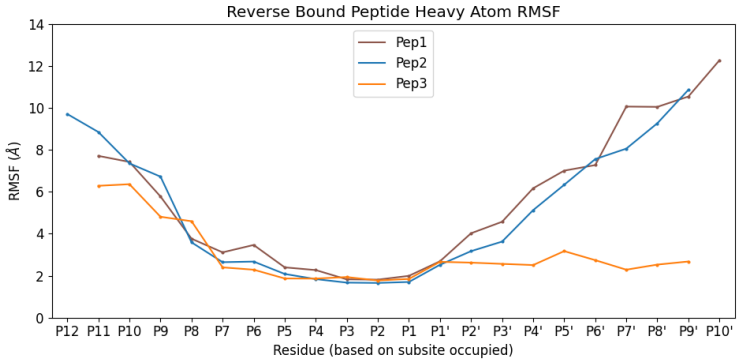


Figure S6.30: Per-residue RMSF of peptide non-hydrogen atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for the three reverse-bound nsp peptides, in the Z state of Cys111-His272.

Table S6.3: PL^{pro}-peptide HBs of significance (occurrence ≥ 25% out of 600 frames analysed every ns) observed over the combined 3 × 200 ns MD for each of the three reverse-bound nsp peptides, in the Z state of Cys111-His272. HB occurrences involving equivalent carboxylate oxygens (in Asp and Glu sidechains) are combined, while salt bridges involving carboxylate oxygens and Arg sidechain guanidino group are combined and halved to avoid double counting. Peptide residues are italicised, while PL^{pro} residues are not. HBs in the S and S′ subsites that are outside the S4-S1 subsites are in blue and red respectively.

Donor	Acceptor	%		Donor	Acceptor	%		Donor	Acceptor	%
Pep 1 (Z)	Res 10-13 in S1-S4			Pep 2 (Z)	Res 10-13 in S1-S4			Pep 3 (Z)	Res 10-13 in S1-S4	
ALA12N	GLY163O	94		*ALA12N*	GLY163O	85		ASN267ND2	*VAL1O*	95
GLY271N	*GLY11O*	87		ARG166NH	*LYS15O*	69		ASN267ND2	*THR3OG1*	94
TYR112N	*GLY10O*	58		*THR14OG1*	ASP164OD	68		*LYS12N*	GLY163O	93
GLY11N	GLY271O	52		TYR112N	*GLY10O*	51		GLY271N	*GLY11O*	91
ARG6NH	ASP286OD	42		*LYS15NZ*	GLU167OE	51		*THR3OG1*	GLN269OE1	81
TYR13N	GLN269O	36		*GLY10N*	GLY271O	35		*ILE13N*	GLN269O	71
ARG15NH	GLU167OE	35		ARG166NE	*THR14OG1*	34		TYR112N	*GLY10O*	66
THR14OG1	ASP164OD	32		TYR273OH	*ALA12O*	33		TYR273OH	*LYS12O*	48
TYR273OH	*ALA12O*	32		GLY271N	*GLY11O*	33		ARG166NH	*ASN15O*	46
HIP272ND1	*ASN9O*	30		*GLY11N*	GLY271O	33		*LYS12NZ*	ASP164OD	44
ARG166NE	*THR14OG1*	26		ARG166NH	*ASP20OD*	30		HIP272ND1	*LYS9O*	34
								ARG166NE	*ASN15O*	33
								GLY11N	GLY271O	30
								LYS19NZ	GLU203OE	30
								LYS12NZ	GLU167OE	29
								ASN109ND2	*LYS9O*	26

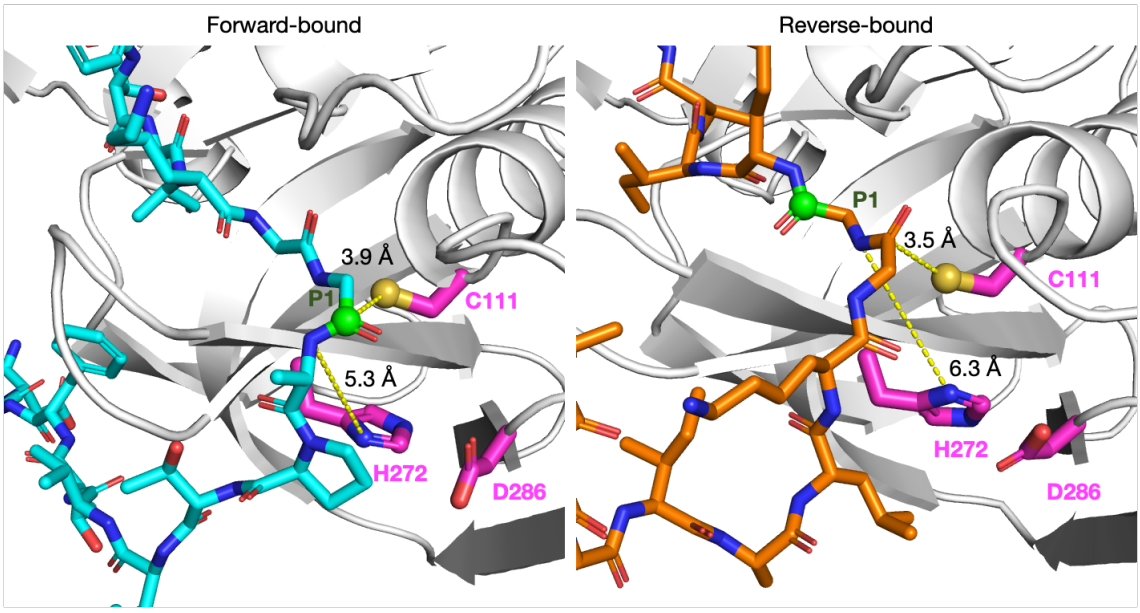


Figure S6.31: Comparison of the modelled forward and reverse oligopeptide binding modes near the active site, using the structures of PL^{pro} complexed with forward-bound peptide **2** (cyan) and reverse-bound peptide **3** (orange) prior to MD for illustration. The catalytic triad residue sidechains are shown as magenta sticks and are labelled. The original P1 residue backbone carbonyl carbon is shown as a green sphere. Note that in the reverse binding mode, while Cys111 is positioned to attack the P2 carbonyl carbon, in the subsequent step the substrate amide nitrogen atom would be too far from His272 to be protonated and act as the leaving group. Relevant calculated distances (yellow dashed lines) are given.

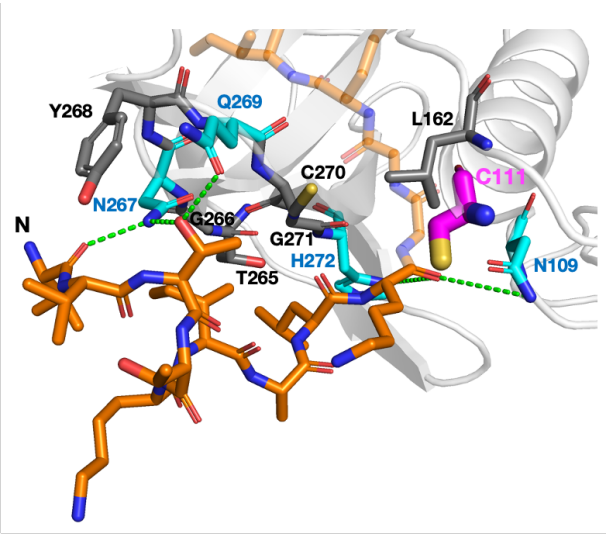


Figure S6.32: View of clustering-derived representative structure of Z-state PL^{pro} (white cartoon) complexed with reverse-bound peptide **3** (orange), with focus on the interactions in the S′ subsites. PL^{pro} residues involved in HBs (green dotted lines; see **Table S6.3**; not all significant HBs are present in these snapshots) are shown as cyan sticks. Other residues in close contact with the peptide N-terminal residues (within 4 Å) are shown in grey. The active site Cys111 is shown as magenta sticks. Clustering was performed using a 3 Å RMSD cut-off on the peptide non-hydrogen atoms using *gmx cluster* (gromos algorithm).[5, 6] Out of 601 frames extracted every ns from the combined 3 × 200 ns MD, the number of (multimembered) clusters obtained is 29 (21).

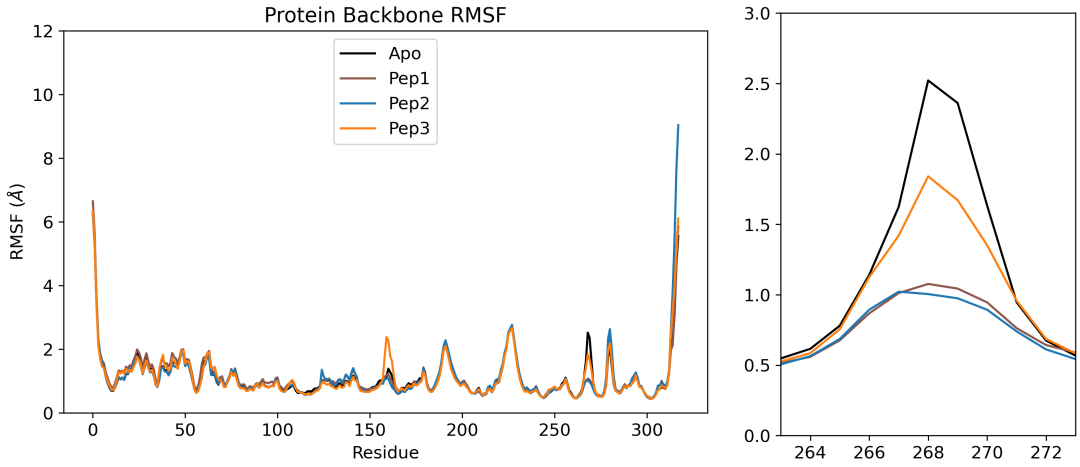


Figure S6.33: Per-residue RMSF of PL^{pro} backbone atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for PL^{pro} (Z state) when *apo* and when complexed with each of the three forward-bound nsp peptides, (left) in full and (right) focusing around the Tyr268-containing loop region.

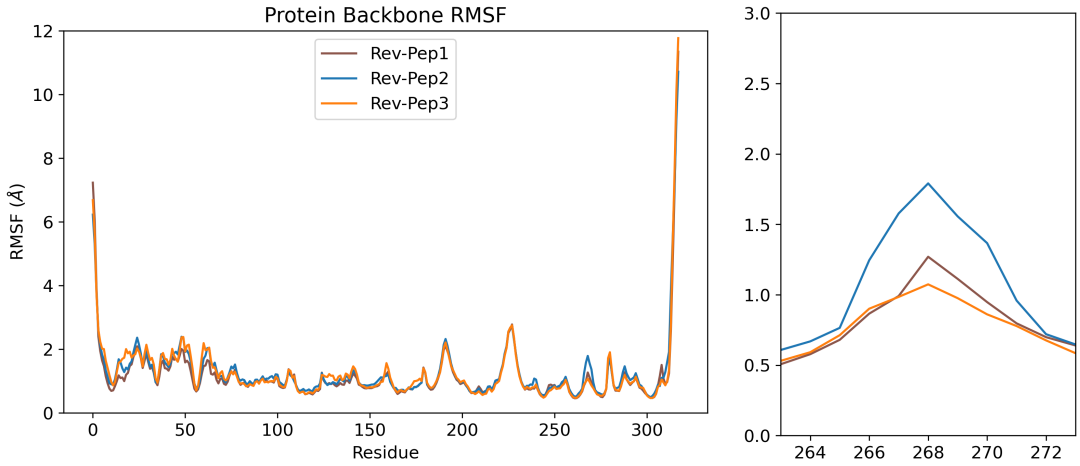


Figure S6.34: Per-residue RMSF of PL^{pro} backbone atoms in the combined 3 × 200 ns MD (fitted based on PL^{pro} backbone) for PL^{pro} (Z state) when complexed with each of the three reverse-bound nsp peptides, (left) in full and (right) focusing around the Tyr268-containing loop region.

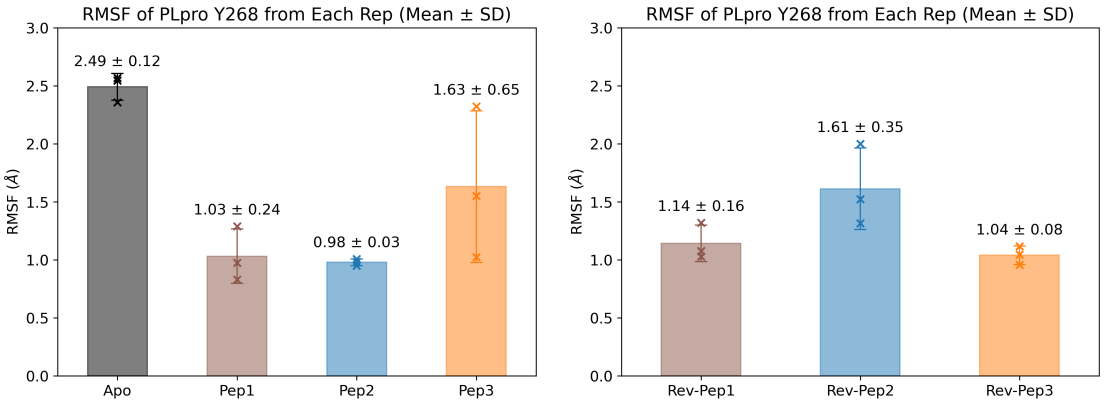


Figure S6.35: Per-residue RMSF of PL^{pro} Tyr268 backbone atoms in each replica of 200 ns MD (fitted based on PL^{pro} backbone) for PL^{pro} (Z state) (left) when *apo* and when complexed with forward-bound peptides, and (right) when complexed with reverse-bound peptides. The bar height is the arithmetic mean RMSF, and the error bar corresponds to the standard deviation (SD) across the three replicas.

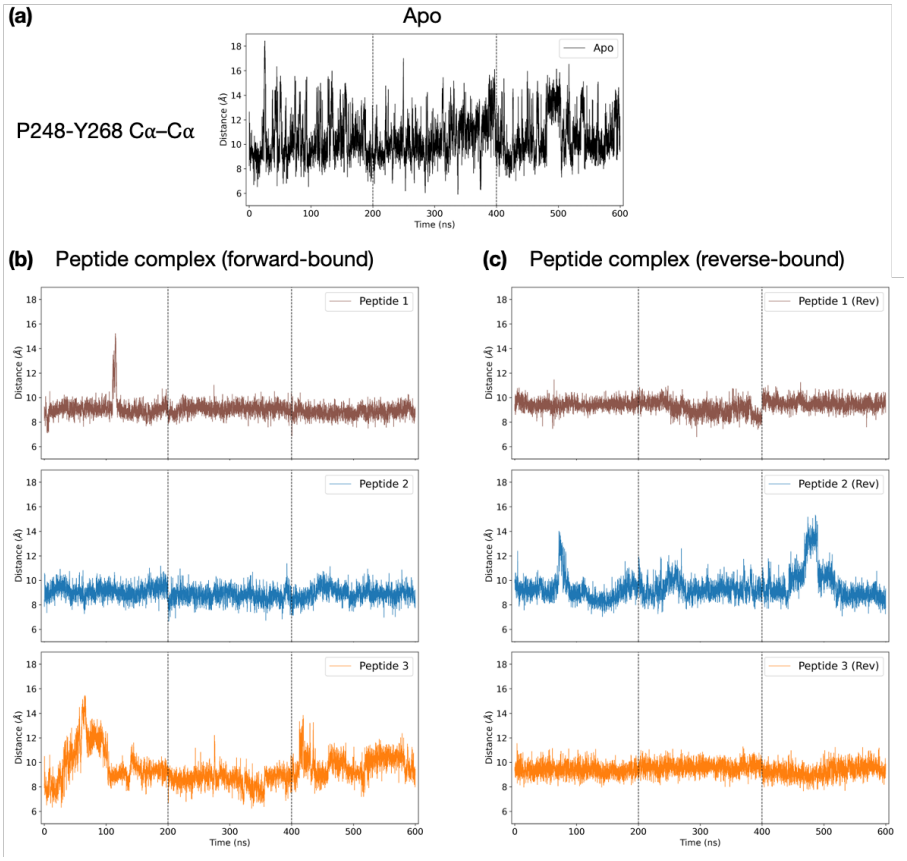


Figure S6.36: The evolution of the Pro248–Tyr268 Cα–Cα distance over combined 3 × 200 ns MD for PL^{pro} (a) in the *apo* state, and when complexed with the nsp peptides in the (b) forward and (c) reverse modes.

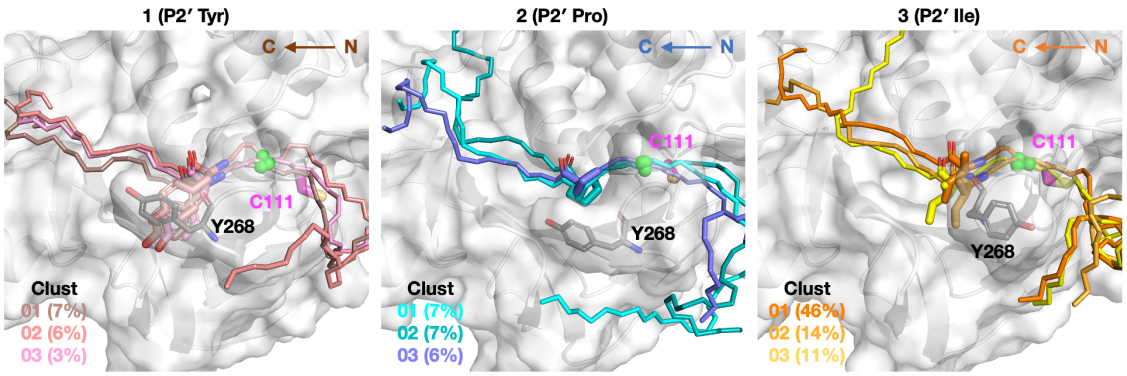


Figure S6.37: View of representative structures of the three most populated clusters of Z-state PL^{pro} (white cartoon and surface) complexed with reverse-bound peptides **1**, **2**, and **3**, with a focus on the S4-occupying residue (which is originally the P2′ residue). For simplification, for the rest of the peptides only the backbone atoms are shown in a uniform colour. Tyr268 is shown as grey sticks, Cys111 as magenta sticks, and the original P1-Gly amide carbonyl carbon atoms as green spheres. Clustering was performed using a 3 Å RMSD cut-off on the peptide non-hydrogen atoms using *gmx cluster* (gromos algorithm).[5,6] Out of 601 frames extracted every ns from the combined 3 × 200 ns MD, the number of (multimembered) clusters obtained for peptides **1**, **2**, and **3** are, respectively, 226 (83), 213 (68), and 29 (21).

Table S6.4: Deviations of P4-P1 Cα atom positions (compared to the VIR251 in PDB 6WX4)[3] and RMSD values for the 100 highest ranked solutions from ADCP[1] docking of peptides **8**, **9**, and **10**, derived from the human proteins IRF3, PROS1, and ULK1, respectively. Solutions that pass the filter are in green. Other solutions which have Gly-Gly in S2-S1 are in orange.

Peptide 8	IRF3						*Peptide 9*	PROS1						*Peptide 10*	ULK1				
Ranking	**Deviation (Å)**				**RMSD (Å)**		**Ranking**	**Deviation (Å)**				**RMSD (Å)**		**Ranking**	**Deviation (Å)**				**RMSD (Å)**
	P4	*P3*	*P2*	*P1*				*P4*	*P3*	*P2*	*P1*				*P4*	*P3*	*P2*	*P1*	
1	20.691	14.988	11.329	12.316	15.271		1	21.076	19.259	17.103	16.880	18.658		1	12.663	7.514	3.548	4.217	7.861
2	26.166	27.233	23.456	21.644	24.723		2	18.251	12.861	12.110	10.890	13.818		2	26.802	29.800	29.013	26.499	28.064
3	30.575	32.051	27.490	22.512	28.393		3	11.785	7.754	1.270	6.462	7.784		3	30.919	31.341	28.685	25.911	29.294
4	9.673	6.512	6.460	7.313	7.602		4	31.728	28.276	23.500	19.529	26.172		4	11.646	7.326	0.722	6.814	7.685
5	17.590	12.066	9.581	5.071	11.964		5	30.648	29.071	23.579	20.387	26.249		5	25.403	27.730	22.492	22.203	24.562
6	20.291	16.738	14.997	15.492	17.006		6	0.981	0.741	0.702	0.694	0.788		6	12.550	8.660	3.674	4.020	8.096
7	30.030	31.418	27.298	22.797	28.079		7	18.226	17.470	17.154	17.163	17.509		7	45.458	41.300	38.057	33.415	39.803
8	11.887	11.666	13.353	12.782	12.441		8	34.235	30.474	30.091	28.322	30.855		8	11.376	6.749	0.269	7.174	7.525
9	10.673	10.953	9.500	11.421	10.661		9	17.485	17.538	20.884	19.271	18.847		9	6.453	0.641	0.536	0.555	3.265
10	21.082	18.648	17.556	17.096	18.659		10	28.704	27.161	24.419	23.441	26.017		10	20.371	17.614	14.836	11.094	16.343
11	19.702	19.001	14.174	15.563	17.265		11	12.565	8.398	2.960	4.667	8.046		11	11.715	6.787	0.492	6.859	7.592
12	11.265	7.065	0.615	6.908	7.499		12	33.369	31.689	29.804	26.067	30.354		12	12.975	8.051	3.630	4.065	8.107
13	12.320	11.779	12.456	12.425	12.248		13	14.189	15.573	11.775	10.268	13.114		13	27.052	26.956	24.305	20.514	24.850
14	18.997	17.115	15.134	10.545	15.764		14	17.633	15.702	13.248	9.508	14.347		14	21.116	18.839	13.351	12.353	16.820
15	17.135	17.100	15.180	10.824	15.277		15	21.699	15.364	17.451	17.363	18.117		15	31.001	26.026	24.437	19.253	25.526
16	19.973	16.753	15.648	16.511	17.299		16	37.978	32.768	31.810	27.811	32.792		16	30.274	25.702	24.355	19.348	25.222
17	13.107	11.471	10.480	12.410	11.908		17	15.371	12.243	14.213	10.322	13.179		17	20.039	19.637	14.544	10.271	16.615
18	22.687	19.700	15.029	10.059	17.535		18	21.918	19.298	16.352	13.399	18.026		18	5.990	0.314	0.204	0.326	3.006
19	20.821	18.225	16.819	16.613	18.197		19	29.433	29.209	23.557	22.084	26.278		19	31.486	25.165	24.560	19.813	25.595
20	17.576	11.120	9.732	5.296	11.783		20	40.358	35.592	33.287	26.988	34.395		20	19.005	18.689	15.009	16.053	17.273
21	12.096	11.960	12.863	12.971	12.481		21	1.135	0.429	0.547	0.825	0.783		21	14.141	13.780	13.020	13.568	13.633
22	19.244	15.705	10.521	12.630	14.893		22	14.277	9.035	3.562	4.178	8.883		22	12.411	7.737	3.733	4.061	7.815
23	17.168	15.298	17.201	15.271	16.262		23	39.317	34.383	32.064	28.067	33.705		23	6.370	0.480	0.407	0.334	3.205
24	14.089	13.870	12.312	14.472	13.710		24	30.028	23.453	20.499	12.864	22.569		24	13.952	8.293	3.821	3.846	8.556
25	19.352	17.359	14.816	12.363	16.188		25	12.449	8.372	3.093	4.471	7.979		25	20.832	18.536	14.731	11.662	16.812
26	12.480	11.524	11.101	13.938	12.309		26	21.488	18.817	14.730	9.978	16.825		26	30.233	27.101	24.454	19.728	25.669
27	17.630	11.413	9.271	5.143	11.763		27	22.502	20.023	16.919	15.477	18.928		27	22.096	19.487	14.444	15.792	18.207
28	20.641	22.132	26.152	28.800	24.644		28	26.184	25.970	29.012	29.035	27.590		28	13.144	15.459	15.164	15.139	14.755
29	25.775	25.880	30.175	32.643	28.767		29	30.986	26.733	23.596	20.537	25.755		29	38.179	32.881	30.668	25.908	32.213
30	20.205	16.639	14.992	15.520	16.961		30	31.806	28.913	25.012	19.249	26.663		30	30.727	29.031	30.082	30.484	30.088
31	24.233	22.299	26.627	28.817	25.612		31	21.046	17.326	12.489	8.450	15.576		31	32.089	27.556	27.318	22.498	27.575
32	23.044	18.123	16.737	12.799	18.052		32	19.615	14.246	12.180	9.108	14.309		32	24.393	19.035	18.574	13.644	19.291
33	19.841	18.325	14.756	15.791	17.296		33	19.701	14.034	13.578	8.600	14.521		33	20.053	16.283	15.602	10.838	16.033
34	29.485	27.341	20.815	19.940	24.737		34	20.643	18.835	16.820	16.580	18.294		34	31.005	29.692	30.207	30.504	30.356
35	22.886	19.103	18.205	18.458	19.753		35	17.205	14.089	13.719	10.832	14.142		35	38.870	34.856	33.079	27.674	33.859
36	26.523	27.270	24.076	21.748	24.999		36	13.374	8.961	3.363	4.314	8.501		36	14.647	9.749	11.869	13.805	12.659
37	26.606	22.046	17.865	14.534	20.762		37	22.831	16.179	15.177	9.894	16.668		37	19.449	18.656	13.305	9.808	15.808
38	7.044	6.165	9.760	10.641	8.604		38	17.916	15.767	11.780	8.513	13.972		38	13.992	13.712	15.522	16.683	15.025
39	23.949	20.824	17.191	15.842	19.708		39	14.400	15.122	15.782	15.866	15.304		39	38.435	36.025	32.879	26.802	33.817
40	14.874	12.558	13.756	16.505	14.496		40	16.689	17.024	17.090	17.777	17.150		40	40.739	41.688	39.097	34.607	39.127
41	19.069	18.892	14.883	11.338	16.360		41	19.004	16.033	14.334	10.930	15.355		41	15.809	10.564	11.097	9.719	12.032
42	12.268	6.742	0.644	6.856	7.800		42	40.782	35.525	32.894	28.495	34.710		42	39.401	35.295	30.683	26.048	33.234
43	13.340	16.318	17.258	21.746	17.428		43	22.383	20.613	16.273	13.634	18.551		43	28.203	25.487	24.805	25.407	26.009
44	29.522	26.607	20.674	20.107	24.552		44	17.027	12.239	9.223	7.938	12.122		44	5.488	0.778	0.376	0.405	2.785
45	11.696	11.449	13.140	12.376	12.183		45	17.695	13.742	11.542	11.610	13.874		45	30.508	29.834	29.577	29.551	29.870
46	30.424	27.337	24.964	17.480	25.503		46	26.188	21.323	22.136	21.173	22.797		46	30.065	24.869	23.670	18.878	24.693
47	18.992	19.500	15.180	16.004	17.518		47	16.113	12.168	14.415	10.869	13.543		47	12.991	17.066	22.337	28.731	21.121
48	19.040	17.955	14.451	12.287	16.161		48	27.159	26.221	24.256	20.438	24.654		48	20.242	21.853	23.716	24.397	22.610
49	12.246	6.726	0.692	6.839	7.785		49	30.480	30.550	25.052	24.015	27.689		49	21.073	18.474	15.922	11.910	17.181
50	21.462	17.141	13.699	13.579	16.782		50	14.814	12.838	14.369	12.699	13.711		50	17.967	13.439	11.692	7.416	13.182
51	27.908	28.925	25.734	24.096	26.732		51	27.023	26.952	24.537	20.713	24.939		51	37.171	32.765	31.211	26.662	32.172
52	25.663	22.495	24.807	24.688	24.441		52	39.351	32.574	30.144	27.115	32.609		52	12.889	8.350	3.605	4.194	8.161
53	29.671	27.910	23.621	18.261	25.252		53	8.913	2.015	7.871	13.392	9.011		53	18.799	16.341	20.513	20.978	19.244
54	18.734	17.404	20.405	21.843	19.668		54	27.557	26.375	24.011	19.885	24.632		54	35.488	28.976	24.772	20.153	27.923
55	18.370	12.512	9.910	5.155	12.438		55	26.560	26.679	29.708	30.054	28.298		55	13.577	12.056	11.402	13.414	12.645
56	27.075	26.458	23.556	19.451	24.322		56	19.402	13.226	14.128	9.822	14.555		56	27.469	29.756	26.799	25.072	27.326
57	21.372	18.220	13.268	14.190	17.074		57	37.838	32.868	33.611	29.328	33.548		57	7.145	0.376	0.309	0.420	3.587
58	27.416	22.950	21.197	16.348	22.332		58	0.814	0.227	0.354	0.693	0.575		58	28.243	24.220	24.823	20.538	24.608
59	11.046	7.784	7.255	9.063	8.908		59	13.419	13.331	13.502	13.993	13.564		59	18.359	14.162	16.956	17.214	16.744
60	18.556	11.710	9.933	5.081	12.308		60	15.034	10.638	10.345	8.994	11.479		60	24.948	21.134	20.163	16.328	20.870
61	30.378	32.123	27.672	23.238	28.550		61	28.684	26.185	22.582	22.923	25.218		61	13.157	10.947	12.003	11.043	11.821
62	29.465	25.386	20.854	13.210	23.033		62	32.653	29.050	23.693	18.989	26.609		62	15.153	11.757	10.722	12.871	12.732
63	18.813	14.783	14.520	15.039	15.886		63	21.482	18.976	17.415	17.097	18.822		63	30.944	30.191	27.836	29.789	29.712
64	19.235	17.356	20.577	20.606	19.489		64	21.759	17.088	13.132	13.918	16.820		64	31.708	27.685	23.867	16.933	25.633
65	16.665	11.720	11.730	8.710	12.535		65	22.324	22.765	18.469	14.002	19.709		65	12.798	13.036	11.485	9.454	11.779
66	28.216	26.191	20.952	18.130	23.716		66	42.391	38.717	37.043	33.314	38.007		66	33.496	35.322	33.082	30.927	33.243
67	10.723	7.143	7.167	7.800	8.340		67	18.926	21.189	15.219	12.485	17.282		67	44.035	43.143	39.384	34.631	40.469
68	13.189	7.837	8.228	9.332	9.876		68	12.576	14.023	15.133	14.911	14.196		68	6.268	0.348	0.677	1.028	3.199
69	29.453	27.523	29.247	28.515	28.695		69	32.491	29.851	25.878	19.487	27.369		69	45.462	41.439	37.403	33.058	39.610
70	21.829	21.695	23.691	26.053	23.384		70	32.214	27.082	26.563	22.656	27.341		70	24.552	27.003	22.592	21.379	23.976
71	12.551	12.164	13.986	12.562	12.835		71	11.665	8.976	12.068	12.049	11.263		71	26.778	25.074	24.809	21.665	24.651
72	19.477	16.851	20.195	21.865	19.680		72	21.107	19.615	18.674	13.991	18.539		72	5.391	0.830	0.368	0.686	2.755
73	17.329	12.440	10.789	10.510	13.056		73	30.662	28.185	29.179	27.507	28.907		73	30.960	29.574	30.164	30.027	30.185
74	16.904	16.583	15.971	16.218	16.423		74	31.275	29.465	26.133	20.521	27.159		74	30.289	29.590	29.719	28.374	29.501
75	34.709	36.127	34.818	33.596	34.824		75	21.169	19.693	24.833	22.877	22.226		75	12.836	11.994	14.555	14.711	13.572
76	33.779	36.609	35.487	30.978	34.279		76	31.331	28.389	23.830	19.075	26.074		76	28.680	28.638	22.228	19.308	25.048

77	15.351	11.047	10.719	10.344	12.037		77	21.110	14.497	14.503	10.332	15.596		77	11.266	6.657	0.383	7.314	7.498
78	11.793	10.642	10.062	13.661	11.621		78	30.036	25.670	21.575	21.574	24.960		78	25.757	21.168	18.450	14.354	20.359
79	26.638	28.374	22.533	21.123	24.842		79	36.648	35.685	33.309	28.223	33.625		79	12.338	8.272	3.351	4.398	7.925
80	10.620	10.878	10.094	11.854	10.880		80	13.411	8.909	3.122	4.651	8.524		80	20.202	16.316	15.598	11.222	16.152
81	13.312	13.470	10.736	14.199	12.995		81	28.468	26.260	23.207	22.892	25.311		81	30.268	30.852	29.007	24.920	28.855
82	34.840	34.727	30.609	29.112	32.420		82	38.657	33.586	33.706	29.521	34.022		82	63.529	57.479	53.332	47.791	55.830
83	23.462	21.572	15.908	12.209	18.828		83	27.877	26.274	24.054	19.675	24.663		83	31.400	26.767	26.052	21.216	26.605
84	19.064	14.861	14.321	14.746	15.865		84	16.768	17.975	17.085	17.719	17.393		84	4.044	3.996	10.493	16.966	10.372
85	28.433	27.112	20.999	19.472	24.308		85	14.685	12.910	14.588	12.465	13.698		85	14.347	11.485	11.522	12.470	12.510
86	11.606	7.171	0.733	6.645	7.597		86	30.389	26.999	25.251	22.592	26.459		86	5.613	0.951	0.891	0.924	2.918
87	15.770	10.738	10.010	7.412	11.392		87	18.461	15.648	13.876	15.514	15.960		87	22.803	20.597	14.633	12.274	18.090
88	12.523	13.524	11.707	15.679	13.441		88	14.463	16.910	13.376	13.006	14.519		88	14.643	12.631	12.550	10.327	12.630
89	45.239	45.557	48.074	46.112	46.259		89	38.390	32.932	31.981	27.958	33.026		89	24.102	19.276	18.604	14.628	19.446
90	22.863	18.192	16.886	12.580	18.007		90	23.658	22.885	21.198	19.149	21.791		90	33.137	27.723	25.252	21.145	27.163
91	20.126	16.732	17.682	14.132	17.301		91	36.404	30.811	28.744	24.192	30.356		91	22.105	23.275	21.697	19.275	21.637
92	24.487	25.176	27.819	30.362	27.061		92	12.172	7.650	1.364	6.525	7.923		92	11.144	6.626	0.681	7.370	7.465
93	17.701	11.093	9.375	5.372	11.760		93	18.423	13.003	12.023	10.899	13.891		93	13.714	12.225	12.306	8.841	11.907
94	30.556	29.966	32.072	32.570	31.309		94	29.586	29.605	31.566	31.064	30.468		94	11.985	11.423	11.155	10.385	11.252
95	20.159	20.271	15.490	16.705	18.278		95	24.271	24.022	23.634	24.265	24.049		95	16.431	14.010	17.500	18.366	16.657
96	17.734	15.158	17.133	15.121	16.328		96	17.338	10.679	9.691	8.998	12.140		96	11.805	7.187	0.603	6.736	7.693
97	27.092	24.839	22.227	20.378	23.771		97	1.332	0.508	0.178	0.382	0.743		97	19.636	12.804	11.247	7.828	13.577
98	19.454	20.097	18.269	19.247	19.278		98	27.880	22.269	17.417	10.810	20.576		98	32.158	30.987	31.349	31.800	31.576
99	32.660	31.484	32.540	31.449	32.038		99	34.652	32.575	31.146	27.991	31.684		99	23.225	24.207	22.031	20.372	22.504
100	12.424	12.057	12.723	10.933	12.054		100	40.157	33.291	28.952	26.947	32.731		100	12.447	7.368	3.541	4.324	7.753

Table S6.5: Deviations of P4-P1 Cα atom positions (compared to the VIR251 in PDB 6WX4)[3] and RMSD values for the 100 highest ranked solutions from ADCP[1] docking of peptide **11** derived from the human protein ATG7. None of the solutions passed the filter or had a Gly-Gly in S2-S1. Solutions where the peptide passes through S2-S1 are in yellow.

Peptide 11	ATG7				
Ranking	**Deviation(Å)**				**RMSD(Å)**
	P4	*P3*	*P2*	*P1*	
1	24.146	18.466	14.408	8.272	17.321
2	22.661	23.382	21.177	15.339	20.880
3	27.039	30.261	27.297	26.712	27.863
4	21.213	21.798	20.288	15.273	19.812
5	28.858	26.516	23.809	24.188	25.922
6	25.972	25.956	25.364	22.815	25.060
7	26.087	26.573	23.289	21.794	24.515
8	29.679	31.882	29.182	27.231	29.540
9	29.074	24.228	18.254	12.004	21.850
10	26.997	21.082	21.437	18.325	22.185
11	43.782	37.931	34.472	27.697	36.439
12	27.916	28.063	25.421	21.954	25.957
13	23.682	23.583	21.601	15.514	21.356
14	34.681	34.114	32.082	30.547	32.897
15	23.121	21.828	19.609	12.732	19.734
16	13.513	7.986	10.109	9.339	10.438
17	39.861	38.628	40.778	41.826	40.291
18	20.335	17.590	11.596	6.524	15.000
19	26.799	22.570	23.739	18.063	23.008
20	9.246	3.158	3.166	7.696	6.417
21	17.448	22.418	26.469	29.720	24.448
22	24.168	24.618	25.501	23.925	24.560
23	23.340	17.549	13.661	8.147	16.626
24	29.742	32.834	30.845	26.788	30.132
25	16.359	11.543	9.929	8.919	12.031
26	21.760	20.369	14.298	9.299	17.170
27	25.744	25.200	21.251	19.254	23.022
28	35.985	34.227	31.426	25.205	31.974
29	25.468	20.450	17.549	11.722	19.444
30	18.149	17.654	15.596	11.627	15.965
31	23.469	22.271	21.732	15.577	20.987
32	32.757	28.374	23.792	22.195	27.096
33	12.383	12.076	9.240	7.307	10.464
34	23.294	21.501	18.033	14.224	19.573
35	26.815	22.678	20.954	15.843	21.929
36	21.106	20.956	15.086	13.041	17.904
37	17.799	13.363	7.549	4.815	11.995
38	11.963	8.214	12.195	15.410	12.214
39	24.486	24.894	25.494	22.736	24.424
40	35.218	35.235	33.692	27.884	33.145
41	16.111	16.971	12.972	11.562	14.573
42	19.689	13.857	10.849	9.782	14.080
43	25.753	20.952	15.913	14.413	19.769
44	29.451	28.255	23.957	21.827	26.058
45	27.441	24.773	25.094	22.394	24.990
46	36.855	35.511	33.371	34.126	34.991
47	23.834	24.400	24.154	22.173	23.656
48	8.088	4.003	4.028	7.231	6.123
49	32.119	27.936	25.673	20.879	26.959
50	34.957	30.435	27.744	21.984	29.161
51	27.168	30.062	27.429	25.690	27.632
52	16.750	15.689	10.854	8.123	13.328
53	16.082	16.522	17.695	19.788	17.580
54	29.480	28.185	22.300	20.323	25.366
55	27.406	21.161	20.718	17.509	21.992
56	20.743	18.213	14.725	13.384	17.014
57	28.452	25.228	23.794	22.963	25.196
58	44.517	38.651	34.920	29.808	37.362
59	43.926	38.541	35.424	28.655	37.050
60	28.481	31.029	26.524	26.383	28.167
61	28.568	25.866	24.765	19.437	24.881
62	26.914	20.908	21.625	17.729	22.042
63	26.867	27.681	27.292	30.578	28.142
64	30.946	33.258	29.226	27.193	30.238
65	44.062	37.789	36.976	32.794	38.119
66	20.063	14.382	11.671	10.054	14.549
67	35.626	32.205	28.949	23.406	30.382
68	3.947	8.819	13.124	14.870	11.031
69	15.958	15.415	10.769	7.879	12.945
70	22.895	16.141	12.849	8.445	15.978
71	25.323	19.264	18.567	12.420	19.438
72	17.197	22.859	26.480	29.292	24.380
73	32.720	30.270	24.796	22.179	27.810
74	28.921	26.225	23.383	23.665	25.646
75	19.772	18.516	19.601	15.553	18.438
76	31.739	25.947	22.864	18.456	25.219

77	17.466	22.216	27.582	32.940	25.712
78	18.950	19.698	14.874	13.083	16.878
79	44.167	38.874	35.751	30.657	37.683
80	31.887	29.713	26.523	25.003	28.409
81	26.758	30.275	27.049	26.545	27.699
82	33.434	30.317	25.931	18.718	27.657
83	21.738	20.458	21.000	17.991	20.345
84	27.989	27.519	22.246	20.844	24.849
85	26.843	22.182	19.922	14.634	21.352
86	23.842	21.479	18.132	13.534	19.632
87	24.162	22.162	16.957	13.548	19.660
88	19.540	13.959	11.139	9.690	14.095
89	18.141	14.707	10.166	5.392	13.018
90	17.780	22.471	25.685	28.218	23.859
91	27.179	25.918	26.832	24.900	26.222
92	23.165	17.632	13.196	9.198	16.630
93	19.612	16.053	11.628	6.053	14.267
94	16.920	22.975	26.466	28.589	24.145
95	3.878	4.159	8.390	12.619	8.093
96	11.413	12.372	14.710	11.489	12.567
97	22.907	17.155	13.205	8.290	16.295
98	25.420	26.473	23.376	21.716	24.316
99	27.590	27.642	24.082	23.276	25.725
100	27.255	23.527	23.373	17.906	23.256

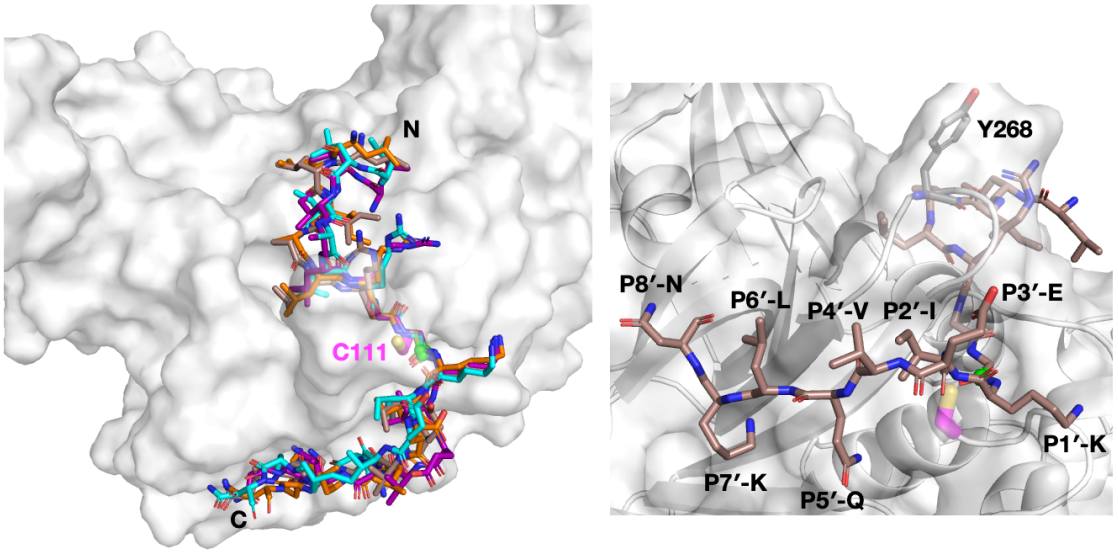


Figure S6.38: Docked conformations of PROS1-derived peptide **9** that passed the applied filter, *i.e.*, the four Cα atoms of P4-P1 have to be within 2 Å of the corresponding Cα atoms of VIR251[3]. The poses include ranked solutions 06 (brown), 21 (cyan), 58 (orange), and 97 (purple), with the general positioning of the terminals labelled. The P1 amide carbonyl carbon is shown as a green sphere. A magnified view of the P′ residues in solution 06 is shown, along with Tyr268 (grey sticks) in the BL2 loop.

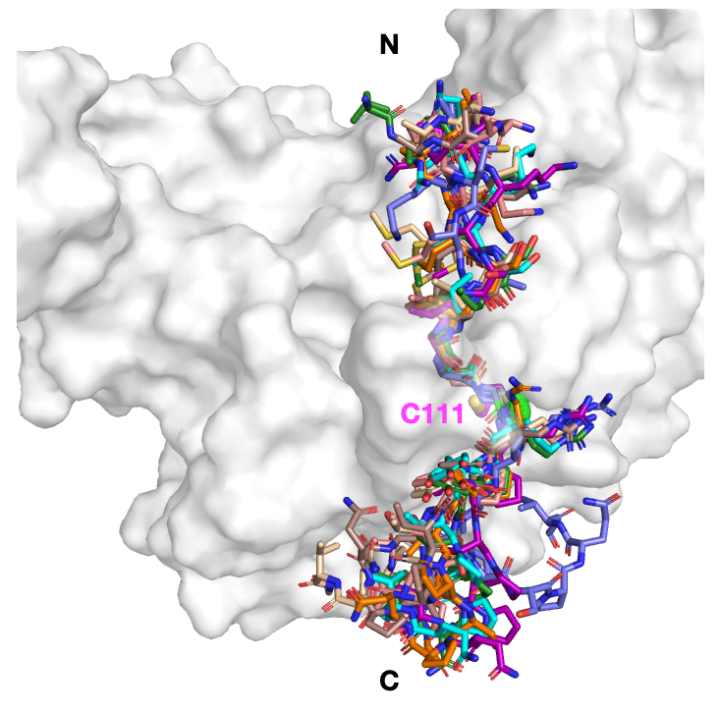


Figure S6.39: Docked conformations of ULK1-derived peptide **10** where P3-P1 Cα deviations were within 2 Å, but not P4. The poses include ranked solutions 09 (brown), 18 (cyan), 23 (orange), 44 (purple), 57 (dark green), 68 (slate), 72 (wheat), and 86 (salmon), with the general positioning of the terminals labelled. The P1 amide carbonyl carbon is shown as a green sphere.

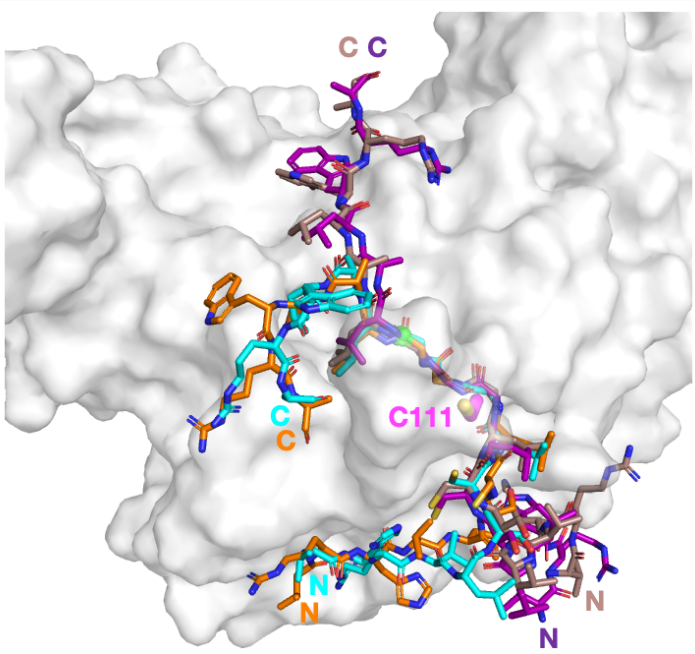


Figure S6.40: Docked conformations of IRF3-derived peptide **8** that are bound in reverse with the P2-Gly in S2 and P3-Gly in S1. The poses include ranked solutions 12 (brown), 42 (cyan), 49 (orange), and 86 (purple), with individual N- and C-terminals labelled. The P1 amide carbonyl carbon is shown as a green sphere.

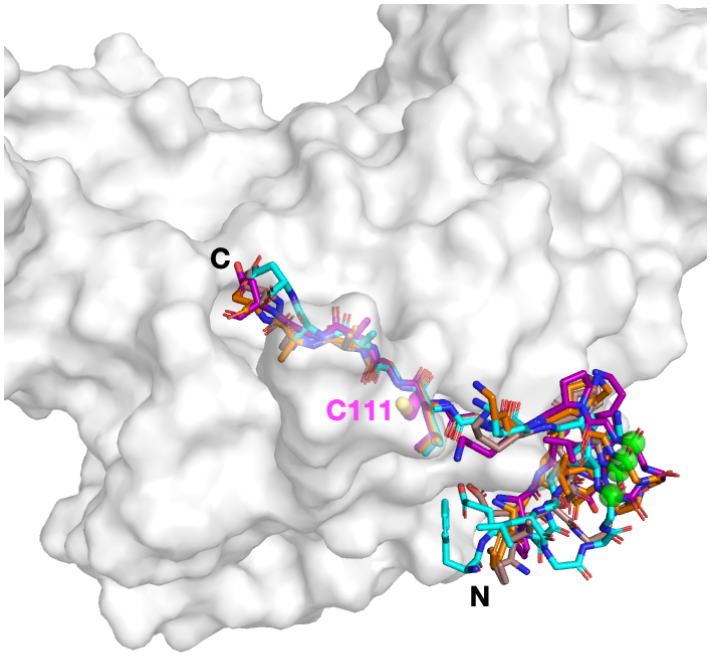


Figure S6.41: Docked conformations of ATG7-derived peptide **11** that are bound in reverse with the P5′-P8′ LAAA residues in the S1-S4 subsites, respectively. The poses include ranked solutions 02 (brown), 4 (cyan), 13 (orange), and 31 (purple), with the general positioning of the terminals labelled. The P1 amide carbonyl carbon is shown as a green sphere.

Table S6.6: Assignment of PL^{pro} histidine protonation states from the structure of PDB 6WX4.[3] Neutral N^{δ}-protonated, neutral N^{ε}-protonated, and doubly protonated histidines are named HID, HIE, and HIP respectively.

HIS	Assignment	Reason
17	HIP	HE2 hydrogen bond (HB) donor to the Glu67 carboxylate (N-O 2.8 Å). Solvent exposed on ND1 side. pK_a calculations (H++ and PROPKA)[9, 10] consistently suggest double protonation at pH 7, with negatively charged residues Asp12 and Glu67 nearby, but no positively charged residues nearby.
47	HIE	ND1 HB acceptor for Ser49 sidechain OH (N-O 3.0 Å) or Ser49 backbone NH (N-N 3.1 Å).
50	HIE	ND1 HB acceptor for Tyr27 sidechain OH (N-O 2.9 Å) or His47 backbone NH (N-N 3.2 Å).
73	HIE	No clear HB partners nearby, assuming HIE.
89	HID	NE2 solvent exposed. ND1/HD1 HB donors to Ser85 backbone O (N-O 3.4 Å).
175	HID	ND1/HD1 HB donors to Tyr171 backbone O (N-O 3.0 Å).
255	HIE	NE2/HE2 HB donors to Lys279 backbone O (N-O 2.6 Å).
272	HIE/HIP	Part of the catalytic triad. HE2 HB donor to Asp286 (N-O 3.0 Å). ND1 side solvent exposed; protonation state likely depends on the charge state of Cys111. Subsequent QM/MM-US calculations (**Figures S6.7-11**) suggested HIP is favoured.
275	HID	NE2 HB acceptor for Gln122 NE2 (N-N 2.8 Å), which in turn is a HB donor to backbone O of Leu118 (N-O 3.0 Å) and HB acceptor, via its OE1, for T277 sidechain OH (O-O 2.6 Å).

References

1. Zhang, Y.; Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019,** *35*, 5121.
2. Brewitz, L.; Kamps, J.; Lukacik, P.; Strain-Damerell, C.; Zhao, Y.*, et al.* Mass spectrometric assays reveal discrepancies in inhibition profiles for the SARS-CoV-2 papain-like protease. *ChemMedChem* **2022,** *17*, e202200016.
3. Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D.*, et al.* Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **2020,** *6*, eabd4596.
4. Grossfield, A. WHAM: an implementation of the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham/ (accessed 2020-04-17).
5. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
6. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.*, et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
7. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.*, et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015,** *43*, D364.
8. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577.
9. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012,** *40*, W537.
10. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525.

Appendix for Chapter 7

(Supplementary Information to Manuscript V)

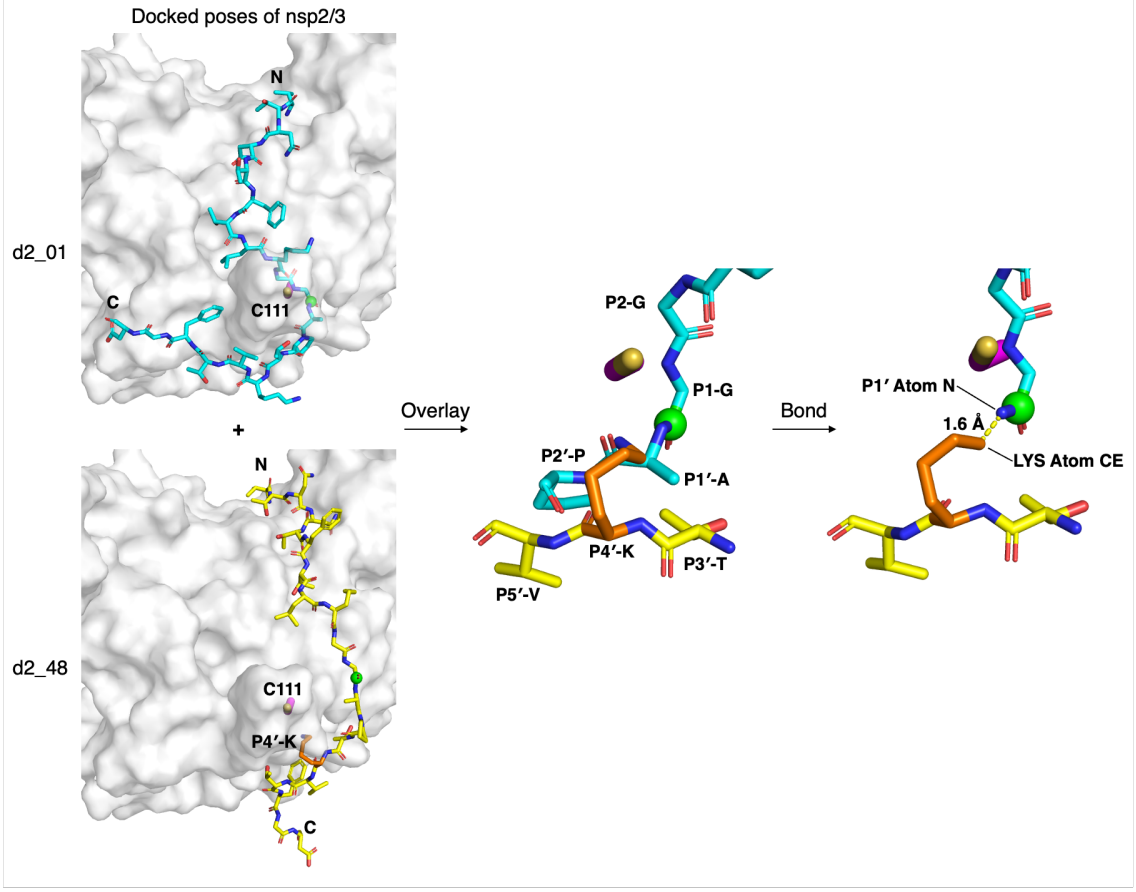


Figure S7.1: AutoDock CrankPep (ADCP)[1]-predicted conformations of the pp1a/1ab-derived nsp2/3$_{808\text{-}827}$ peptide **2** (VTNNTFTLKGG↓APTKVTFGD)[2] in complex with SARS-CoV-2 PL^{pro} (PDB 6WX4).[3] The 1st and 48th ranked poses are in cyan and yellow, respectively. The N- and C-termini of **2** are labelled, with the P1 scissile amide carbon as a green sphere, and the P4′ lysine of the 48th ranked pose in orange. The active site Cys111 is in magenta. These poses provided a basis for modelling the branched peptide sequences.

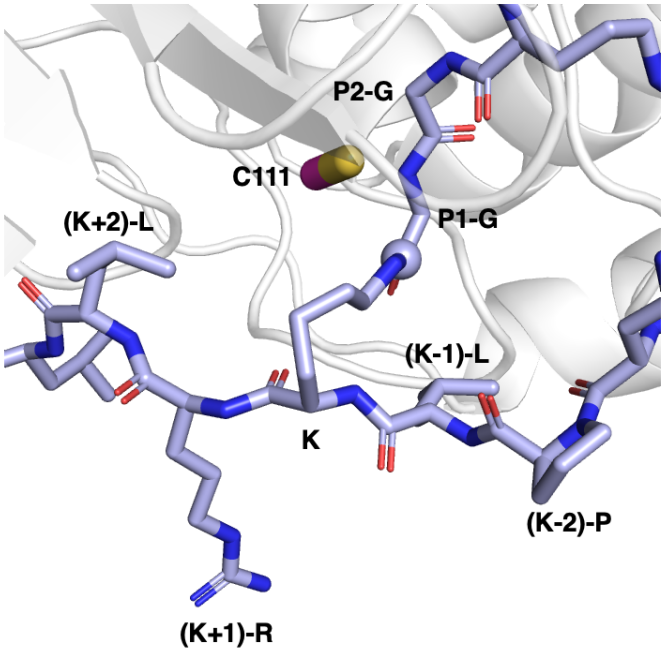


Figure S7.2: View of the modelled active site of the PL^{pro}-$IRF3_{189-197}$-ISG15 (**4**) complex. The N^{ε}-lysine amide linkage was constructed as described in the **Material and Methods** section.

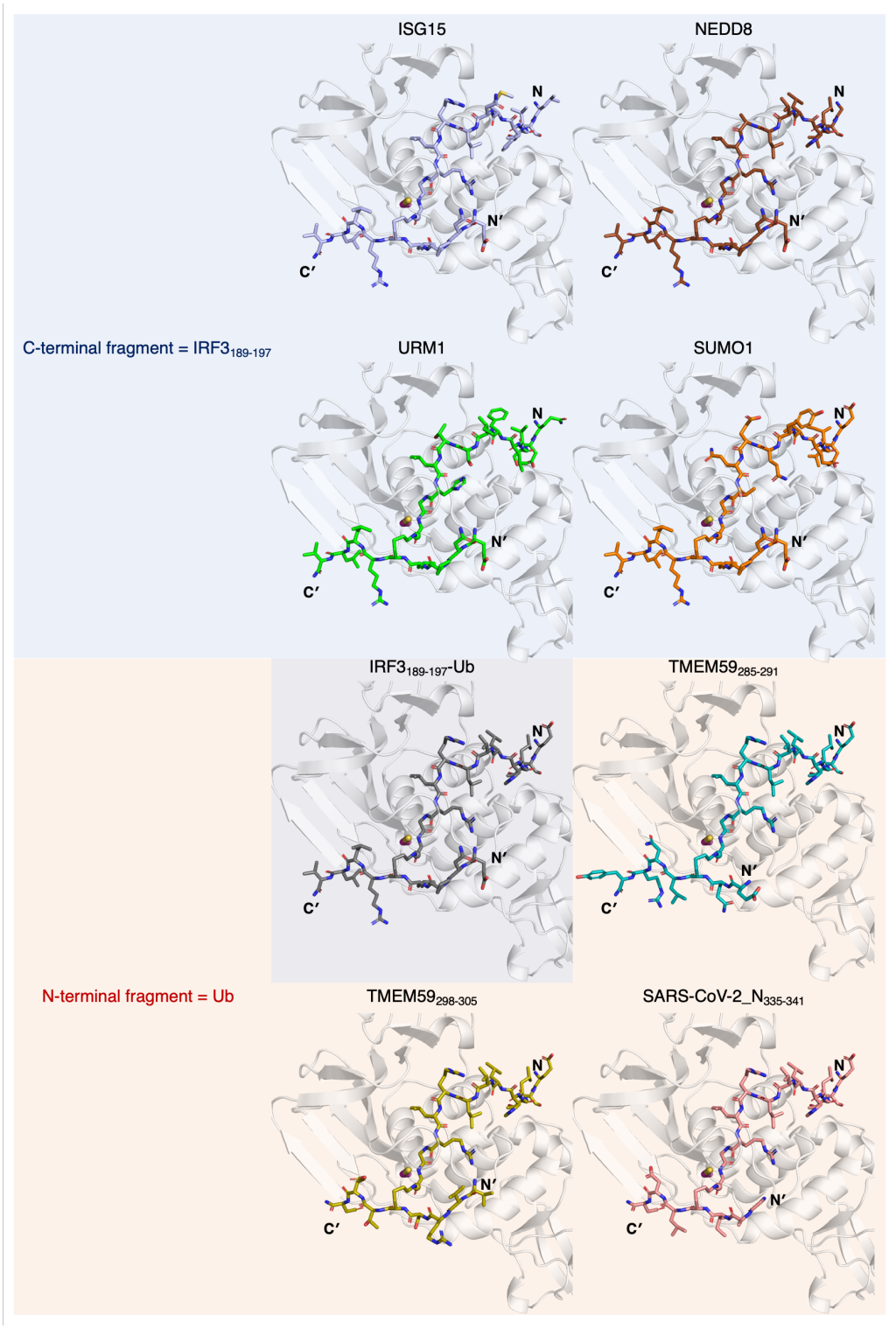


Figure S7.3: View of the modelled PL^{pro} complexes with the N^{ε}-lysine-branched oligopeptides $IRF3_{189\text{-}197}$-(ISG15/Ub/NEDD8/URM1/SUMO1) and ($TMEM59_{285\text{-}291}$/$TMEM59_{298\text{-}305}$/SARS-CoV-2_$N_{335\text{-}341}$)-Ub. The N-terminus of the UBL-derived fragment and the N- and C-termini of the S′ binding fragment are labelled N, N′, and C′, respectively. The P1 scissile amide carbons are shown as a sphere, and Cys111 is in magenta. For molecular dynamics (MD) simulations of the URM1 derived peptide, both the neutral (^{N}) and positive charged ($^{+}$) states of the P3 histidine imidazole group were considered.

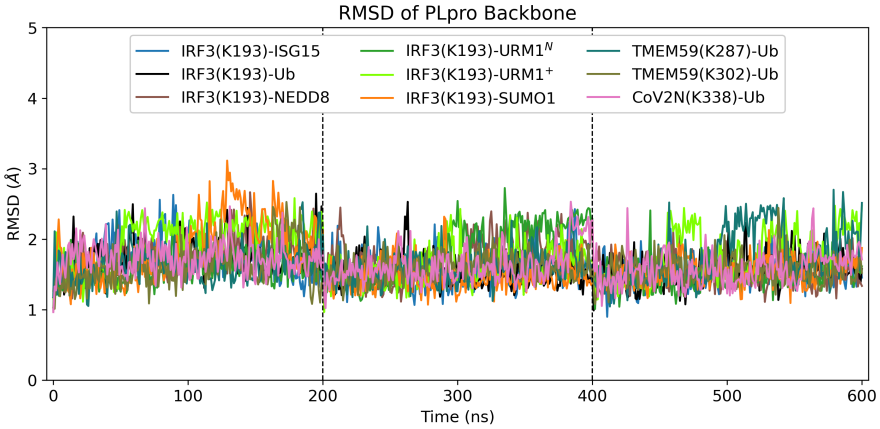


Figure S7.4: Root mean square deviation (RMSD) of PL^{pro} backbone atoms (N, Cα, C) in the combined 3 × 200 ns MD trajectories (fitted based on the PL^{pro} backbone), relative to the PL^{pro} crystal structure (PDB 6WX4).[3]

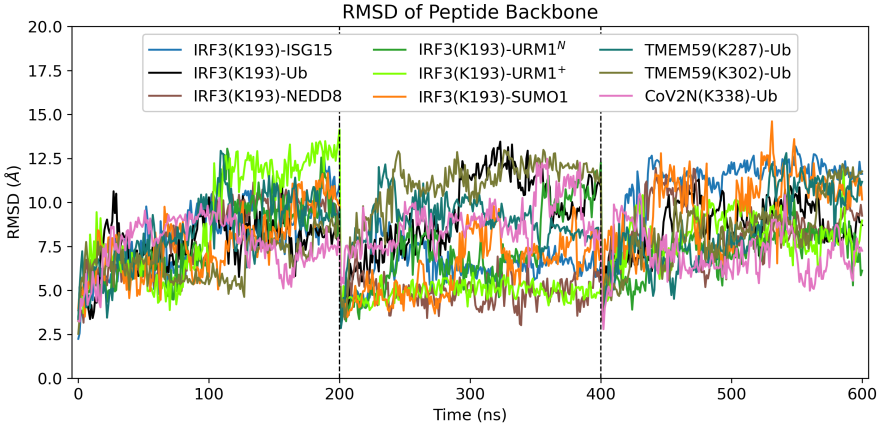


Figure S7.5: RMSD of peptide backbone atoms (N, Cα, C in both fragments) in the combined 3 × 200 ns MD trajectories (fitted based on the PL^{pro} backbone), relative to the starting peptide poses.

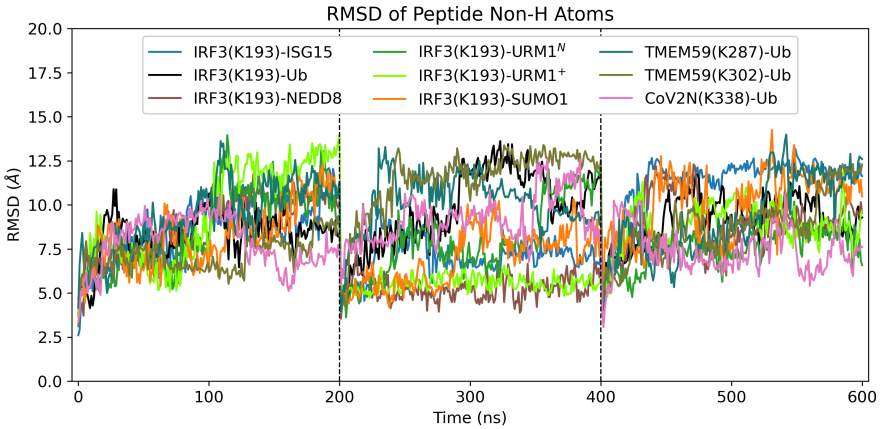


Figure S7.6: RMSD of peptide non-hydrogen atoms in the combined 3 × 200 ns MD trajectories (fitted based on the PL^{pro} backbone), relative to the starting peptide poses.

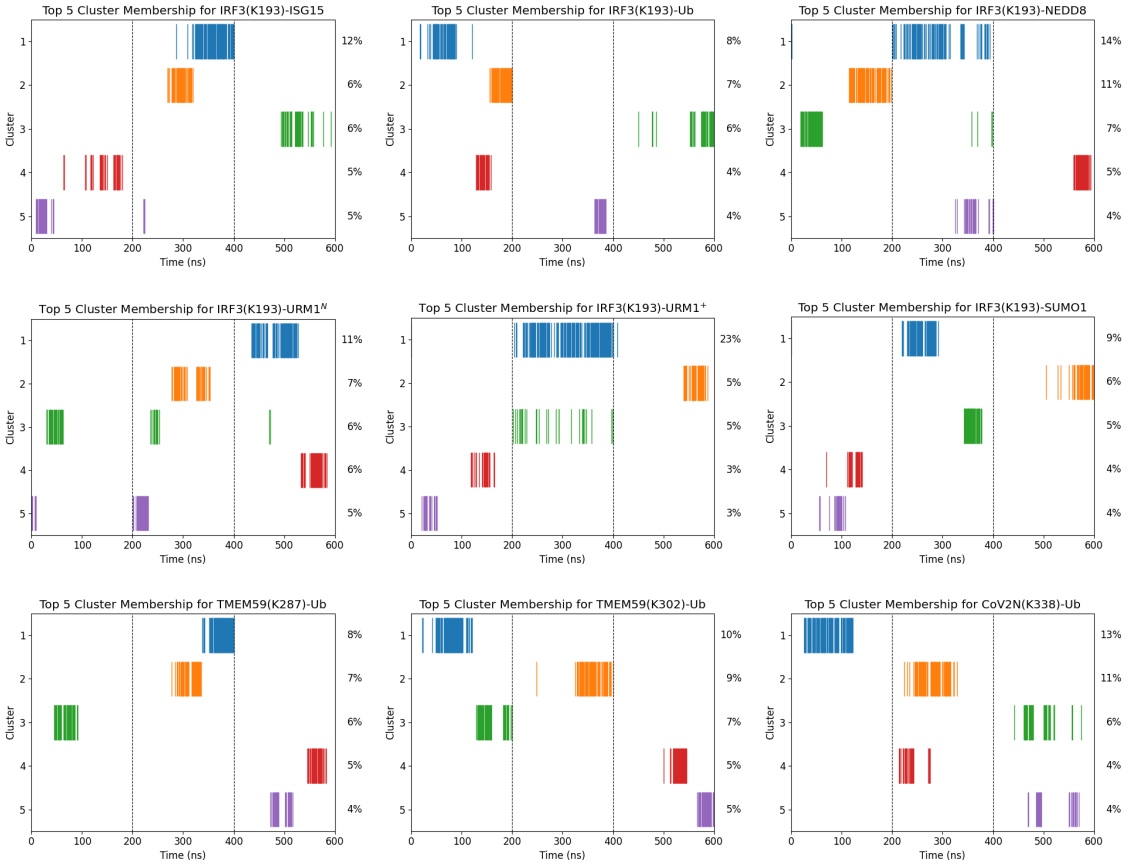


Figure S7.7: Cluster membership of frames extracted every ns from the combined 3 × 200 ns MD (fitted using PL^{pro} backbone) of each PL^{pro}-oligopeptide complex, with clustering performed using a 3 Å RMSD cut-off on the peptide backbone atoms using *gmx cluster* (gromos algorithm).[4, 5] Only the 5 clusters with the highest occupancy percentages are shown. Out of 601 frames, the number of (multimembered) clusters obtained for peptides $IRF3_{189\text{-}197}$-(ISG15/Ub/NEDD8/$URM1^{N}$/$URM1^{+}$/SUMO1) and ($TMEM59_{285\text{-}291}$/$TMEM59_{298\text{-}305}$/SARS-CoV-2_$N_{335\text{-}341}$)-Ub are 128 (57), 181 (78), 106 (58), 150 (63), 160 (63), 159 (74), 138 (76), 104 (61), and 123 (62), respectively.

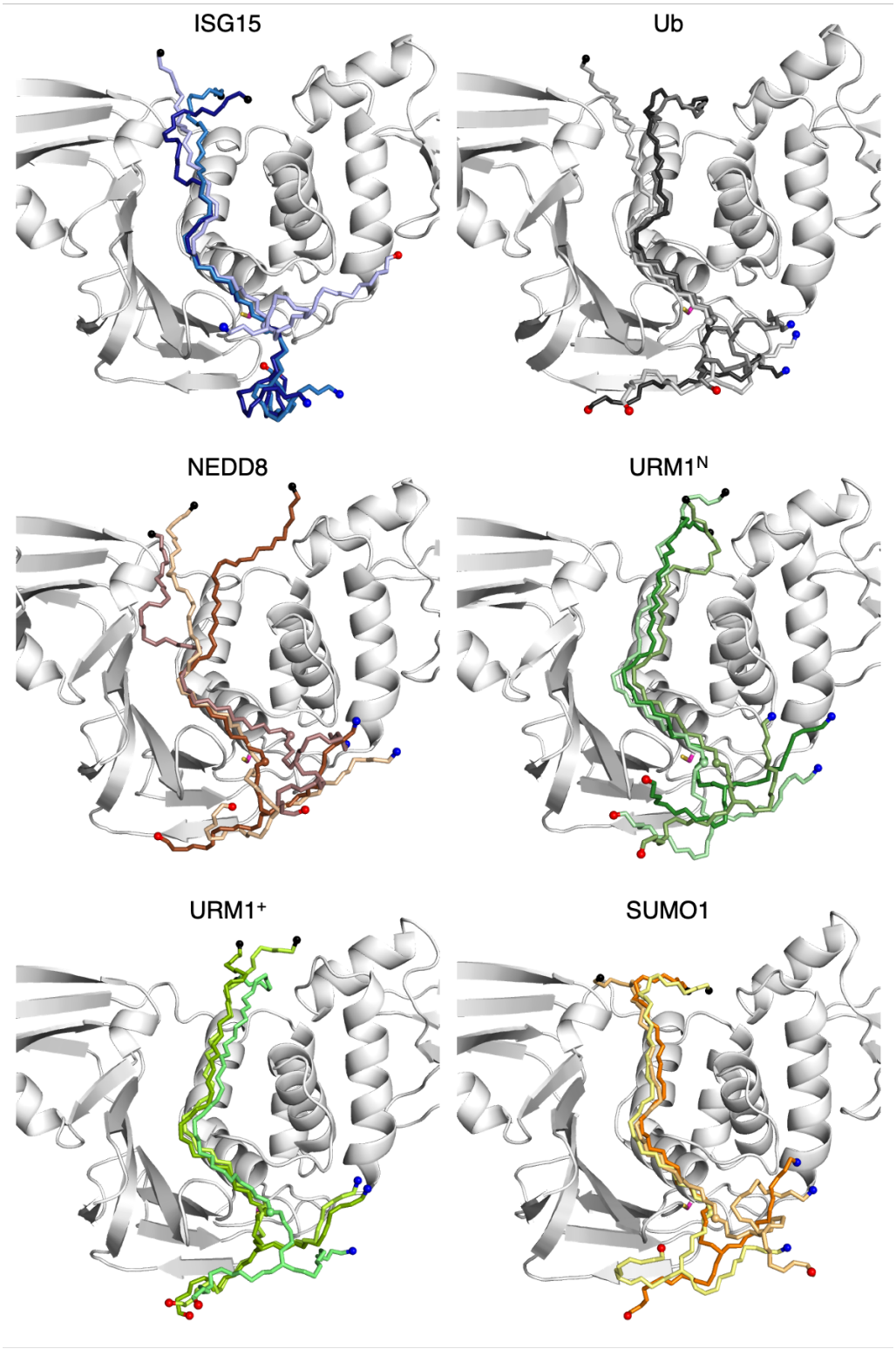


Figure S7.8: View of representative oligopeptide conformers from the PL^{pro}-$IRF3_{189-197}$-UBL (*i.e.,* **4**, **15**, **16**, **17**, **18**) complexes of the three most populated clusters, obtained from the combined 3 × 200 ns MD of each of the PL^{pro} complexes. The PL^{pro} structure is from the most populated cluster, Cys111 is in magenta. The peptide backbones are shown in uniform colours with darker colours corresponding to more populated clusters (P1 scissile amide carbons are shown as sphere), with the N-terminus of the UBL fragment, and the N- and C-termini of the $IRF3_{189-197}$ fragment shown as black, blue, and red spheres, respectively.

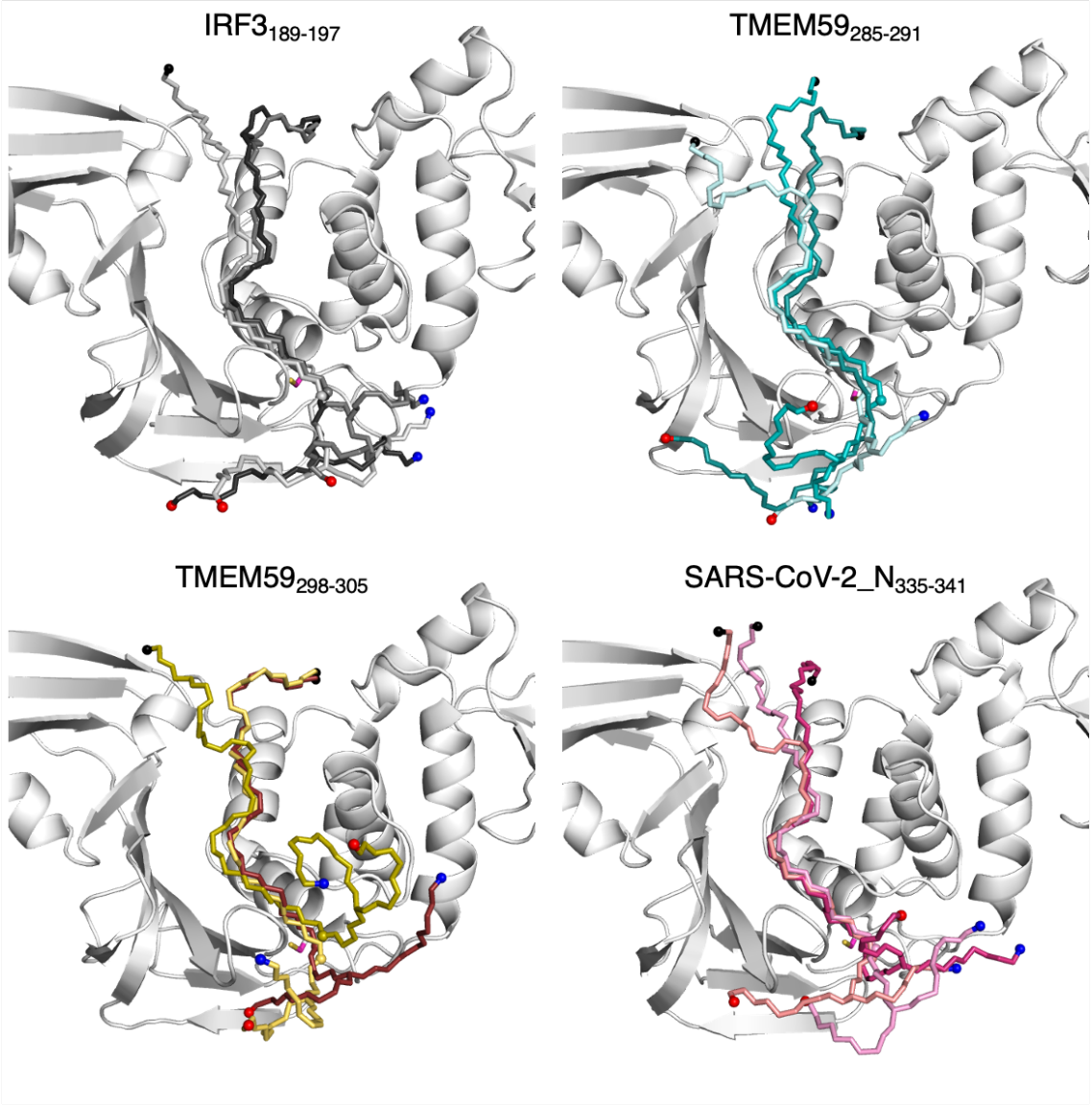


Figure S7.9: View of representative oligopeptide conformers from the PL^{pro}-oligopeptide-Ub (*i.e.,* **15**, **21**, **22**, **23**) complexes of the three most populated clusters, obtained from the combined 3 × 200 ns MD of each of the PL^{pro} complexes. The PL^{pro} structure is from the most populated cluster, Cys111 is in magenta. The peptide backbones are shown in uniform colours with darker colours corresponding to more populated clusters (P1 scissile amide carbons are shown as sphere), with the N-terminus of the UBL fragment, and the N- and C-termini of the IRF3$_{189-197}$ fragment shown as black, blue, and red spheres, respectively.

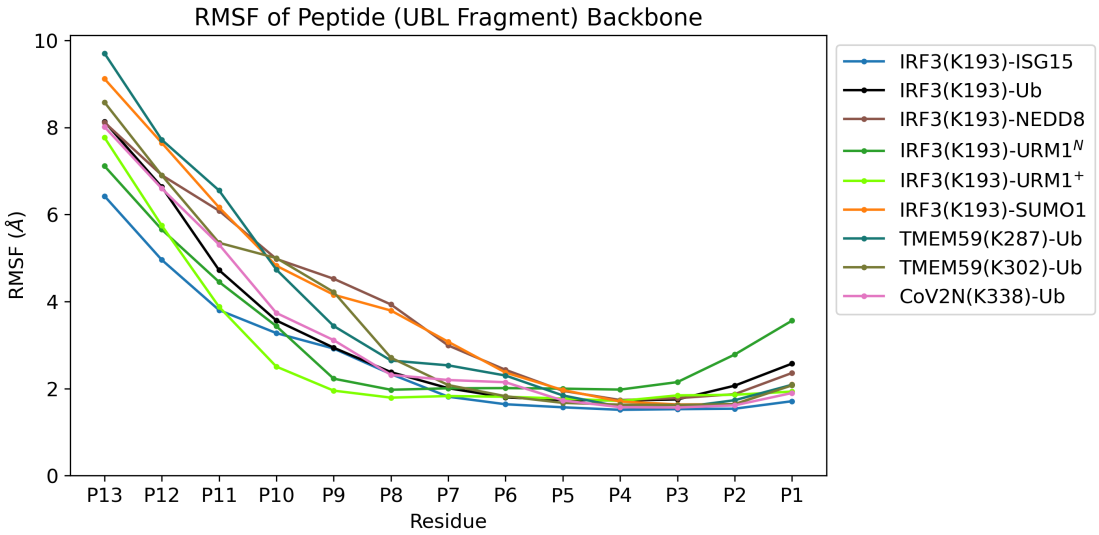


Figure S7.10: Root mean square fluctuation (RMSF) for each residue backbone atoms in the UBL fragment, in the combined 3 × 200 ns MD (fitted based on the PL^{pro} backbone).

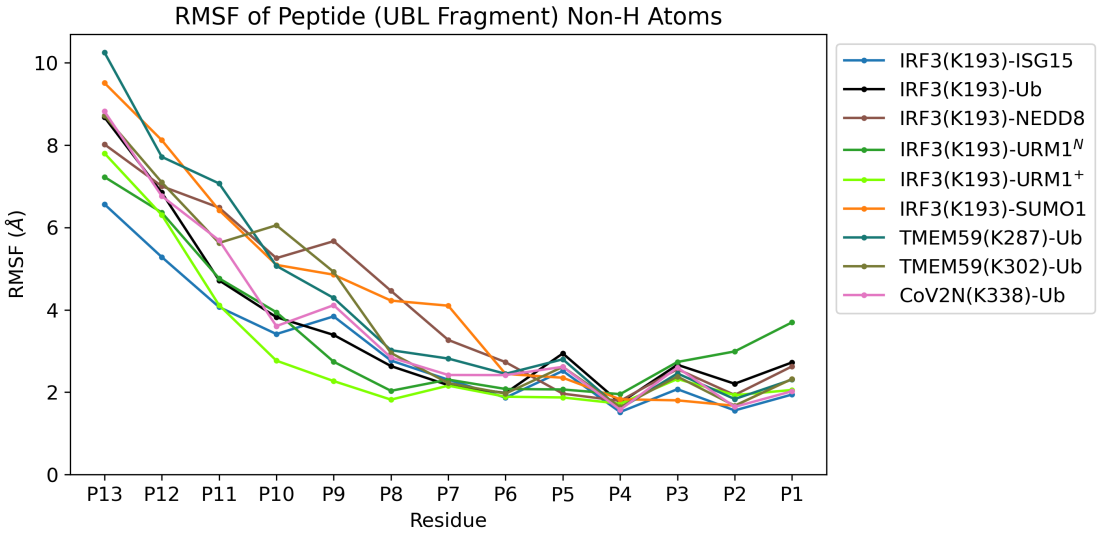


Figure S7.11: RMSF of peptide non-hydrogen atoms in the UBL fragment, in the combined 3 × 200 ns MD (fitted based on the PL^{pro} backbone).

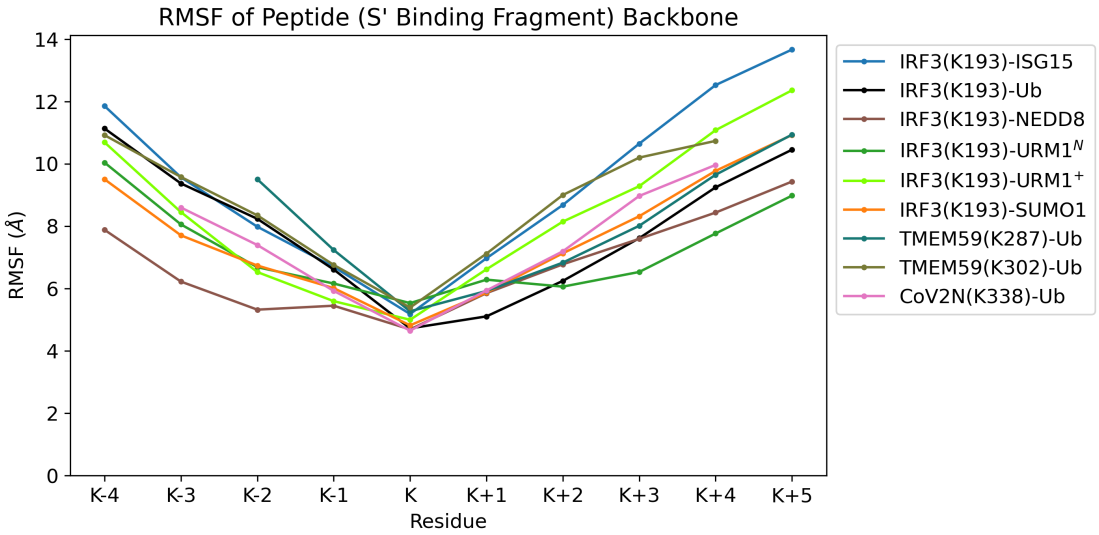


Figure S7.12: RMSF for each residue backbone atoms of the oligopeptide fragment binding to the PL^{pro} S′ sites, in the combined 3 × 200 ns MD (fitted based on the PL^{pro} backbone). Note that the C-terminal NH_2 group is treated as a separate residue.

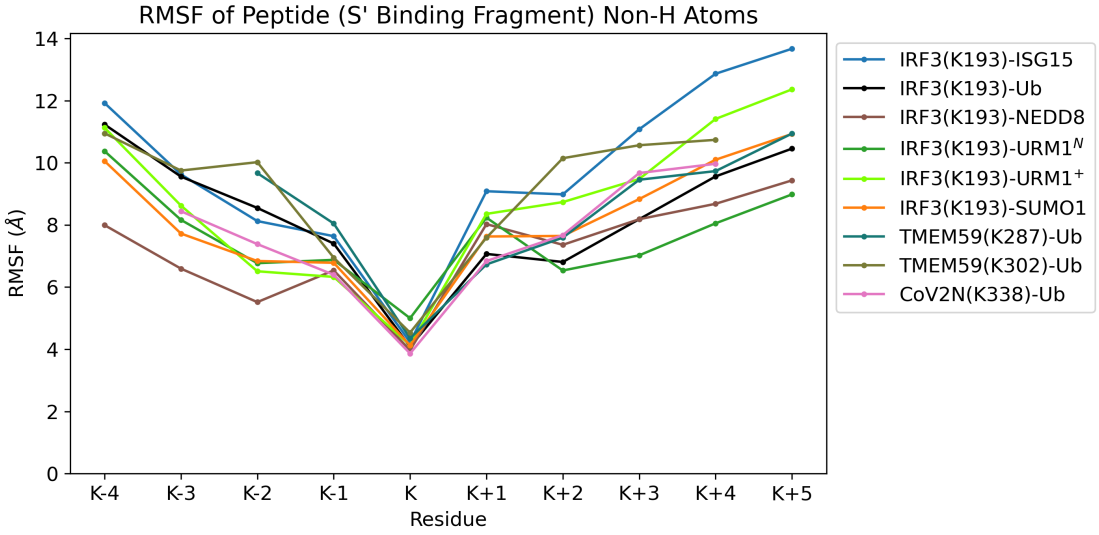


Figure S7.13: RMSF of peptide non-hydrogen atoms of the oligopeptide fragment binding to the PL^{pro} S′ sites, in the combined 3 × 200 ns MD (fitted based on the PL^{pro} backbone). Note that the C-terminal NH_2 group is treated as a separate residue.

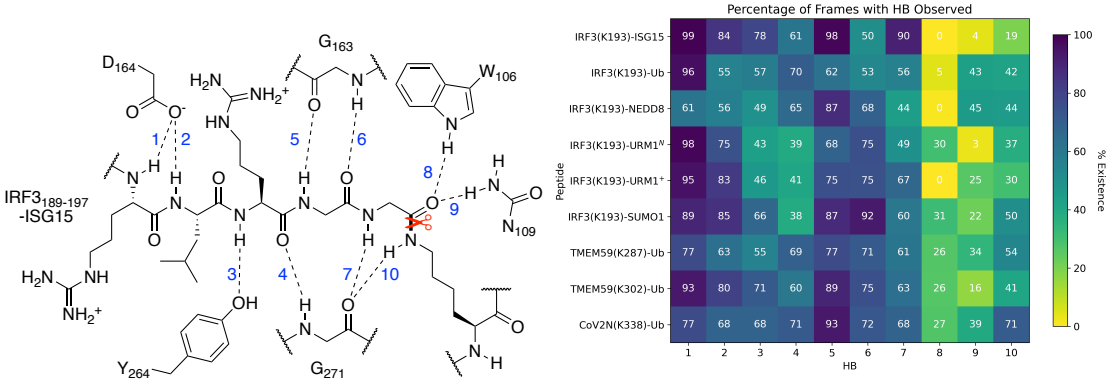


Figure S7.14: Conserved hydrogen bonding interactions in the P5-P1′ region and their percentage occurrences (occurrence ≥ 25% for at least five out of nine peptides), observed in 3 × 200 ns MD. Frames were analysed every ns.

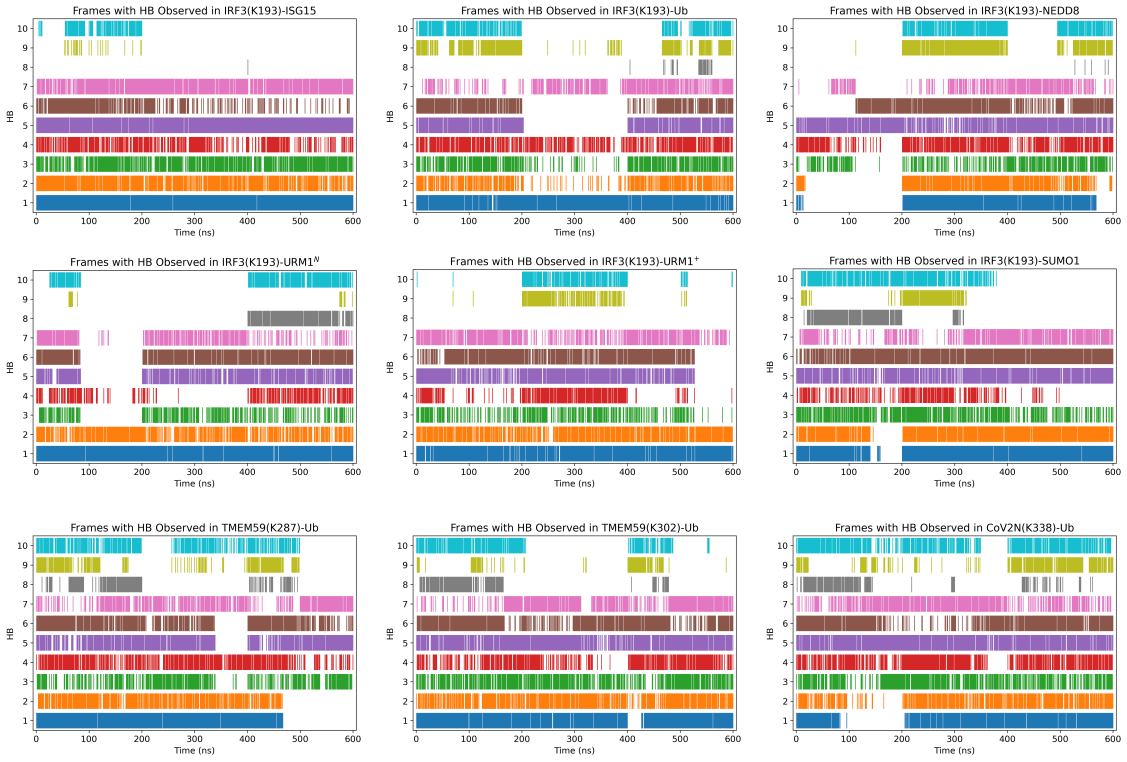


Figure S7.15: Event plots showing the occurrence of the conserved hydrogen bonding interactions in 3 × 200 ns MD.

Table S7.1: Hydrogen bonding interactions between PL^{pro} and N^{ε}-lysine-branched oligopeptides in the modelled PL^{pro}-oligopeptide complexes that were consistently observed (occurrence ≥25% out of 600 frames analysed every ns) over the combined 3 × 200 ns MD for each peptide, other than those already identified to be conserved in the core P5-P1′ region (**Figure S7.14**). Hydrogen bonding interactions involving equivalent carboxylate oxygens (in aspartate and glutamate sidechains) are combined, while salt bridges involving carboxylate oxygens and arginine sidechain guanidino groups are combined and halved to avoid double counting. Peptide residues are italicised, while PL^{pro} residues are not. Hydrogen bonding interactions in the core P5-P1′ region that are not shown in **Figure S7.14** are in black. Outside this region, N-terminal HBs are in blue and C-terminal HBs are in red.

	Donor	Acceptor	%		Donor	Acceptor	%		Donor	Acceptor	%
ISG15	*LYC5Nζ*	ASN109Oδ1	67	IRF3-Ub	*LEU4N*	GLN174Oε1	48	NEDD8	GLU203N	*GLY1O*	34
	VAL4N	GLN174Oε1	55		*ARG9Nη*	ASP164Oδ	41		*ASN2Nδ2*	GLY160O	31
	ARG11Nη	TYR268O	40		*ARG9Nε*	ASP164Oδ	32		ASN109Nδ2	*PRO3O*	29
	GLN174Nε2	*VAL4O*	39		*THR3N*	GLN174Oε1	25				
					THR75Oγ1	*GLU1Oε*	25				
URM1N	*PHE6N*	TYR171Oη	72	URM1+	*PHE6N*	TYR171Oη	72	SUMO1	*GLN8Nε2*	LEU162O	66
	ARG166Nη	*THR9Oγ1*	61		*SER8Oγ*	GLU167Oε	71		ARG166Nη	*GLU9Oε*	64
	SER8Oγ	GLU167Oε	60		ARG166Nη	*THR9Oγ1*	54		*ILE4N*	GLN174Oε1	47
	ASN109Nδ2	*PRO3O*	40		GLN174Nε2	*VAL4O*	51		LYS157Nζ	*GLN8Oε1*	32
	GLN174Nε2	*VAL4O*	38		GLY271N	*GLY13O*	43		ASN109Nδ2	*PRO3O*	29
	LYC5N	ASN109Oδ1	37		ASN109Nδ2	*PRO3O*	28		*VAL3N*	GLN174Oε1	25
	THR9Oγ1	ASP164Oδ	30		*LYC5Nζ*	ASN109Oδ1	27		*GLN10Nε2*	ASP164Oδ	25
	GLY271N	*GLY13O*	27								
	THR75N	*ASP2Oδ*	26								
TMEM59(K287)	*LEU4N*	GLN174Oε1	34	TMEM59(K302)	ARG166Nη	*VAL7O*	47	CoV2N	ARG166Nη	*VAL7O*	54
	GLN174Nε2	*LEU4O*	30		*ARG11Nη*	TYR268O	38				
					LEU4N	GLN174Oε1	30				
					LEU6N	SER170Oγ	28				
					LEU162N	*ARG3O*	26				

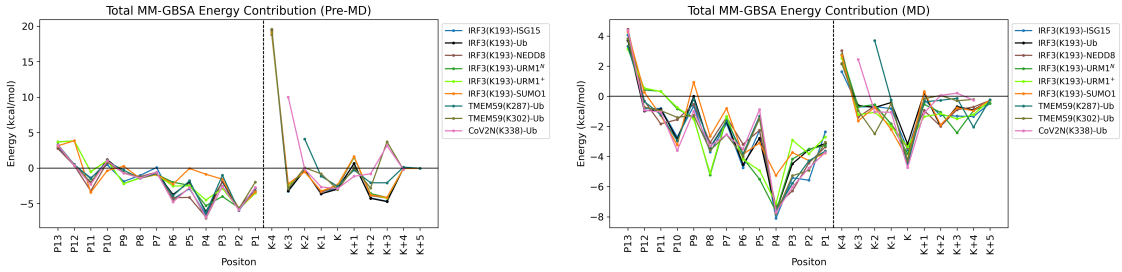


Figure S7.16: Contribution of each peptide residue to the PL^{pro}-oligopeptide molecular mechanics/generalised Born surface area (MM/GBSA) binding energy, from complex structures prior to MD and obtained from MD, where frames were analysed every 5 ns.

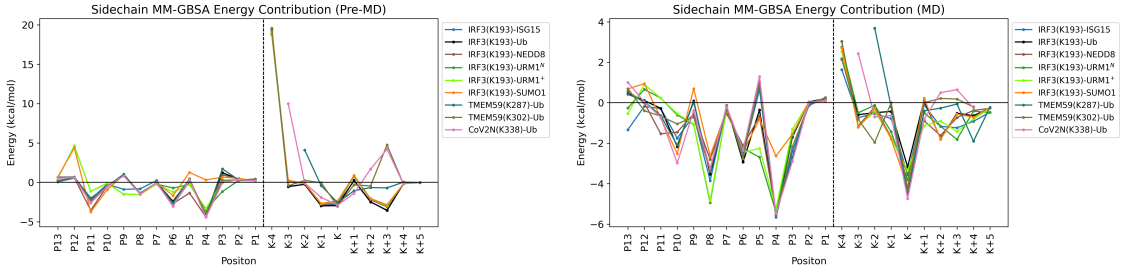


Figure S7.17: Contribution of the sidechain of every peptide residue to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD.

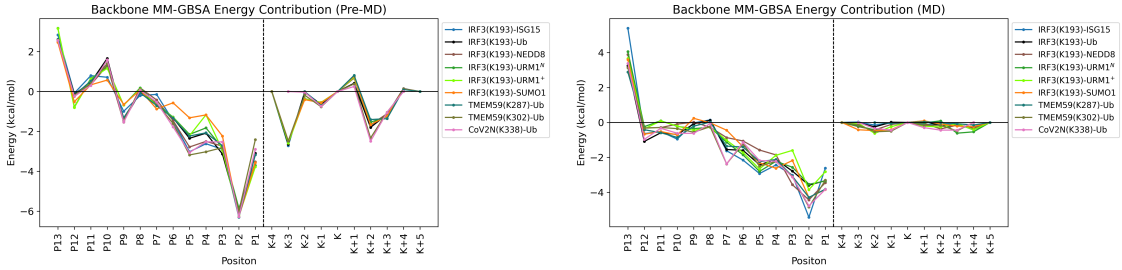


Figure S7.18: Contribution of the backbone of every peptide residue to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD.

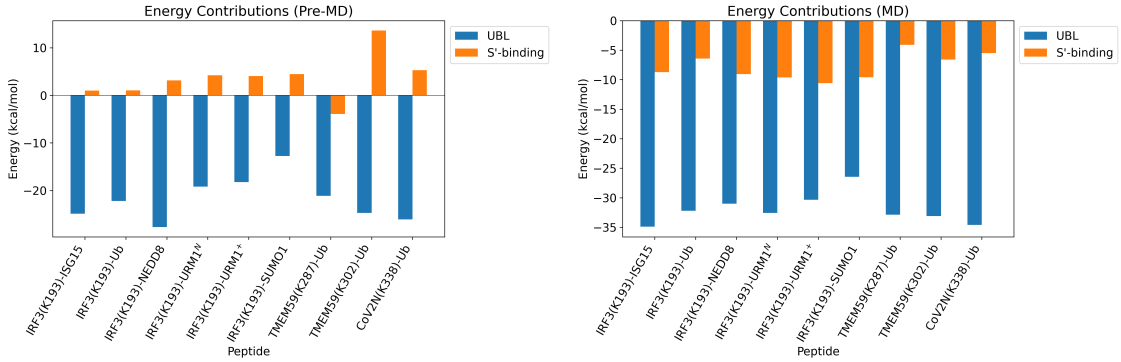


Figure S7.19: Contribution of the UBL-derived peptide fragment and the peptide fragment binding to the S′ sites to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD.

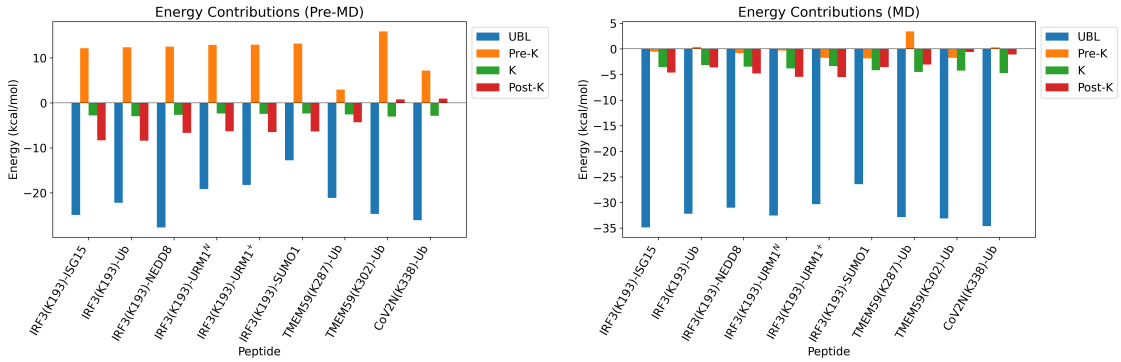


Figure S7.20: Contribution of the UBL-derived peptide fragment and the peptide fragment binding to the S′ sites (split into the individual contributions by the N^{ε}-branched lysine residue, and by the residues N- and C-terminal of this N^{ε}-branched lysine residue) to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD.

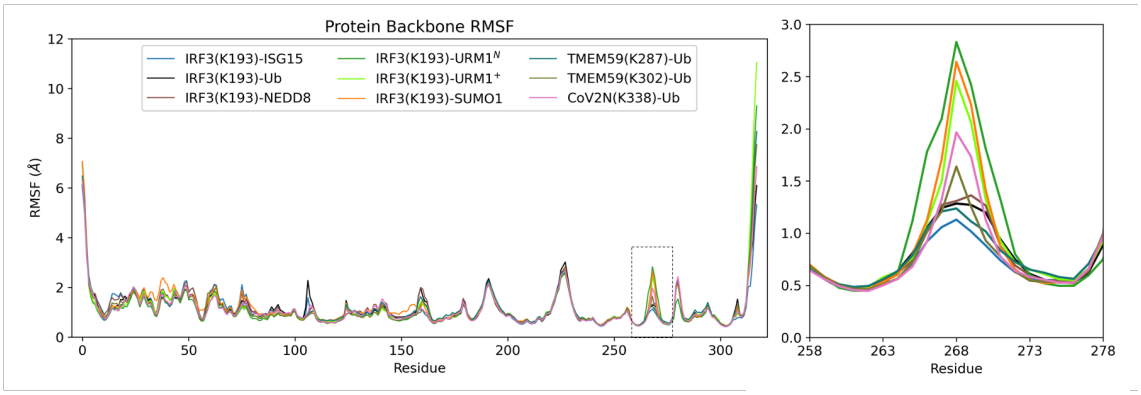


Figure S7.21: RMSF of the backbone atoms of PL^{pro} residues in the PL^{pro}-oligopeptide complexes in the combined 3 × 200 ns MD (left), and (right) focusing around the Tyr268-containing blocking loop 2 (BL2) region.

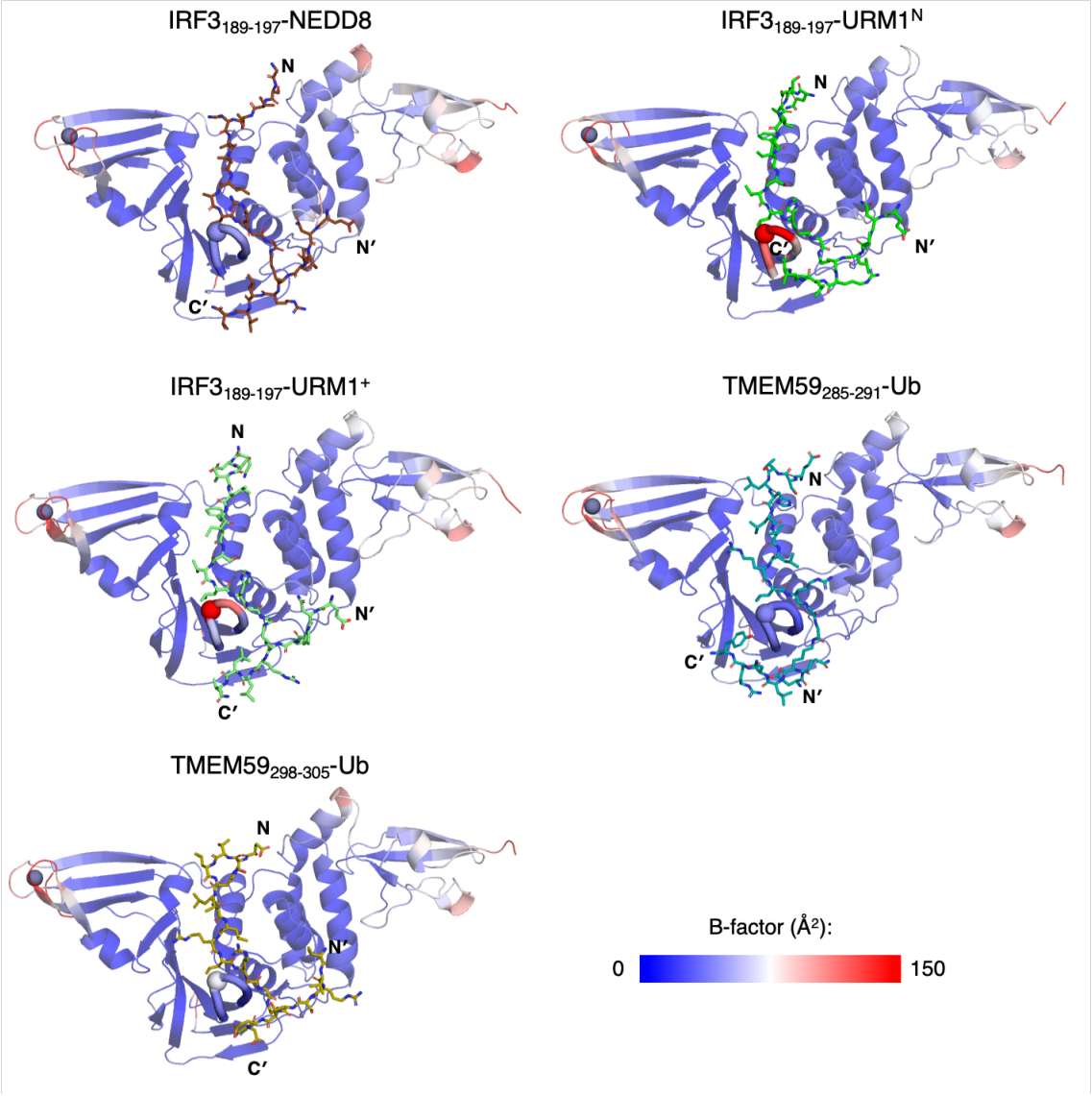


Figure S7.22: Views of the PL^{pro}-oligopeptide complexes representative of the most populated cluster (**Figure S7.7**) with PL^{pro} residues coloured based on MD-derived backbone B-factors, when complexed with IRF3$_{189-197}$-(NEDD8/URM1^{N}/URM1^{+}) and (TMEM59$_{285-291}$/TMEM59$_{298-305}$)-Ub. BL2 (Gly266-Gly271) is shown as a thickened loop, with Tyr268 Cα shown as a sphere. The N-terminus of the UBL-derived fragment and the N- and C-termini of the S′ binding fragment are labelled N, N′, and C′, respectively.

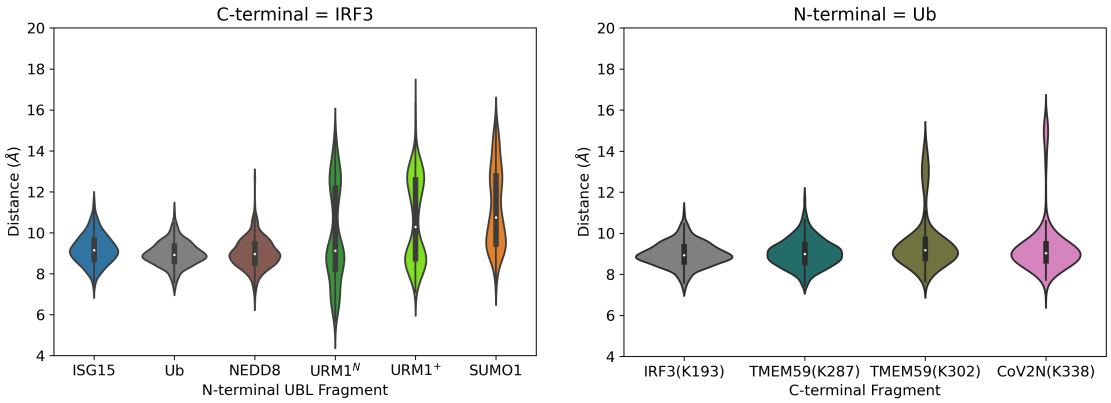


Figure S7.23: The distribution of the PL^{pro} Pro248–Tyr268 Cα–Cα distance over combined 3 × 200 ns MD for each PL^{pro}-oligopeptide complex.

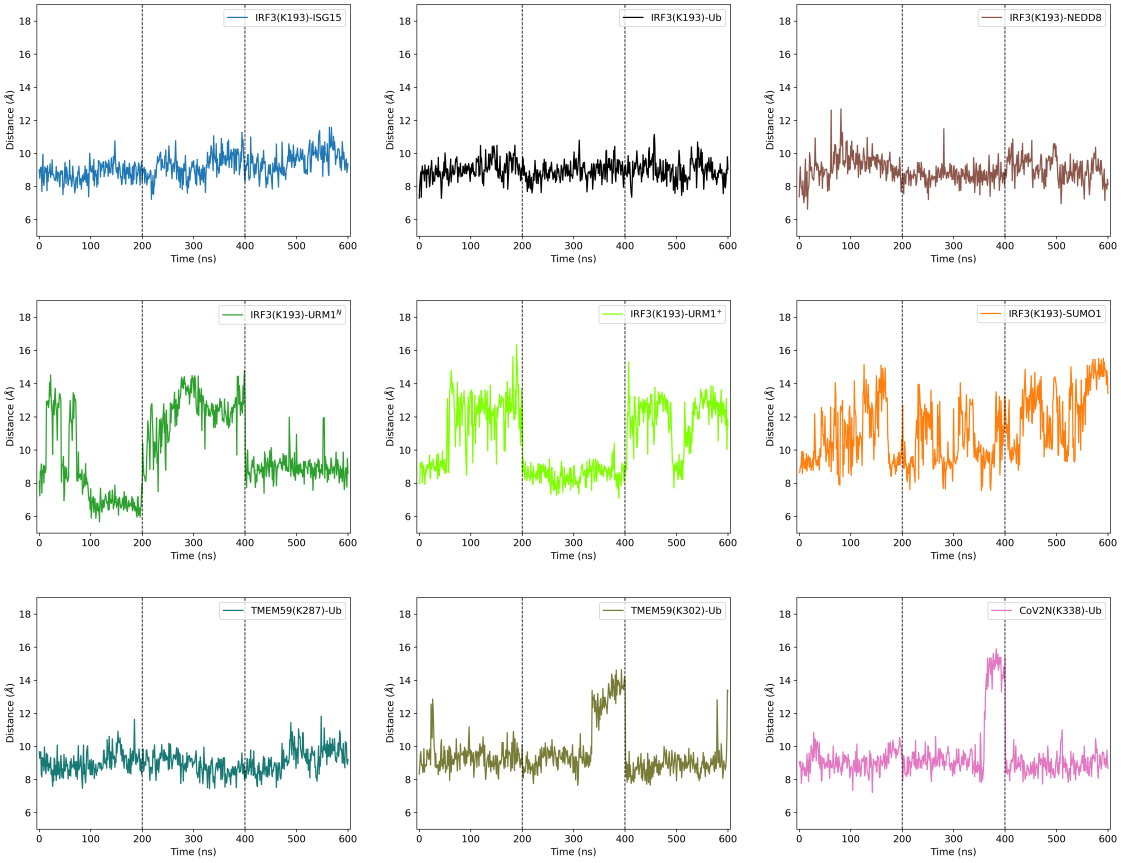


Figure S7.24: The evolution of the PL^{pro} Pro248–Tyr268 Cα–Cα distance over combined 3 × 200 ns MD for each PL^{pro}-oligopeptide complex.

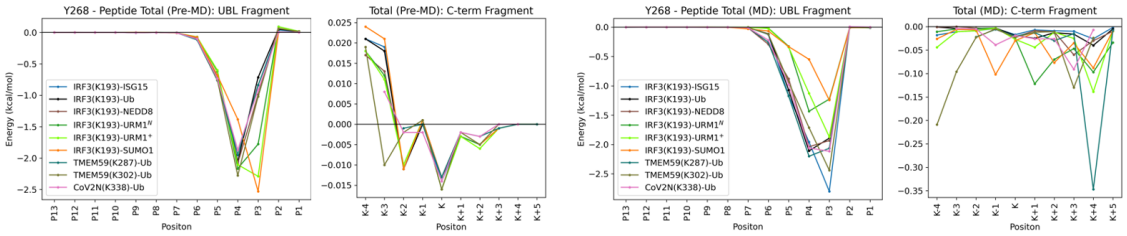


Figure S7.25: Pairwise contribution between Tyr268 and every peptide residue to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD, where frames were analysed every 5 ns.

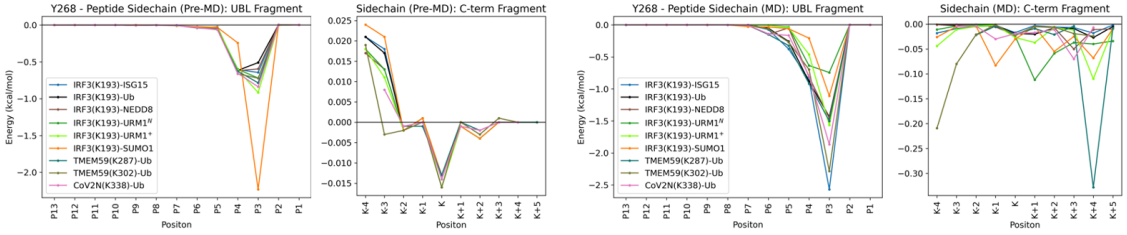


Figure S7.26: Pairwise contribution between Tyr268 and the sidechain of every peptide residue to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD.

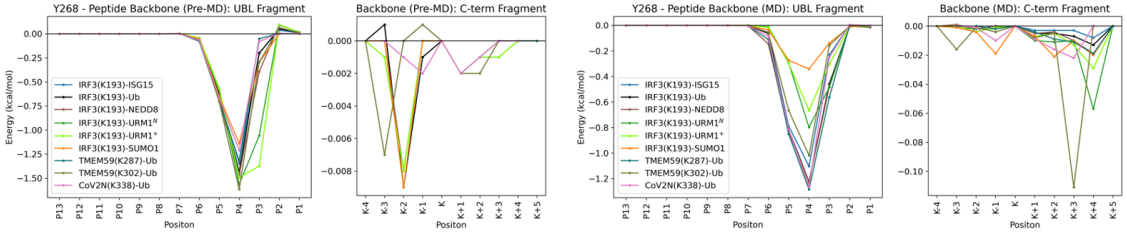
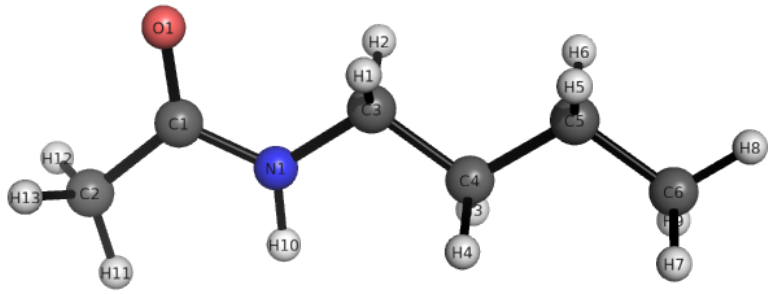


Figure S7.27: Pairwise contribution between Tyr268 and the backbone of every peptide residue to the PL^{pro}-oligopeptide MM/GBSA binding energy, from complex structures prior to MD and obtained from MD.

Table S7.2: Assignment of ionic and protonation states of titratable residues of PL^{pro} (PDB 6WX4)[3] analysed in this study. The setup is identical to that previously reported.[2] Neutral histidine residues that are N^{δ}-protonated and N^{ε}-protonated are referred to as "HID" and "HIE", respectively, according to the AMBER force field nomenclature.[6] Doubly protonated, positively charged histidine residues are referred to as "HIP". The N- and C-termini of PL^{pro} were capped as *N*-acetyl and *N*-methyl amides, respectively. In the peptides, the N-termini were uncapped, while the C-terminus of the PL^{pro} S′ site binding fragment was capped with an NH_2 group. Peptide residues in the S′ binding fragment are labelled with a prime symbol (′), and for ease of setup and subsequent analysis, are numbered according to their positions relative to the N^{ε}-branched lysine, with the lysine always being residue 5′ regardless of the length of the S′ binding fragment.

Residue	LYS (+1)	ARG (+1)	ASP (-1)	GLU (-1)	CYM(-1)	HID (0)	HIE (0)	HIP (+1)	NT (+1)	ZN(+2)	**TOTAL**
PL^{pro}	6	3	12	1	111	89	47	17		1	
	43	65	22	51	189	175	50	272			
	45	82	37	67	192	275	73				
	53	138	40	70	224		255				
	91	140	61	124	226						
	92	166	62	143							
	94	183	76	161							
	105		108	167							
	126		134	203							
	157		164	214							
	182		179	238							
	190		286	252							
	200		302	263							
	217			280							
	218			295							
	228			307							
	232										
	254										
	274										
	279										
	292										
	297										
	306										
	315										
Charge	**+24**	**+7**	**-13**	**-16**	**-5**	**0**	**0**	**+2**	**0**	**+2**	**+1**
Peptides											
ISG15		9, 11, 6′		1′					1, 1′		
Charge		**+3**		**-1**					**+2**		**+4**
$IRF3_{189\text{-}197}$-Ub		9, 11, 6′		1, 1′			5		1, 1′		
Charge		**+3**		**-2**					**+2**		**+3**
NEDD8		11, 6′		1′			5		1, 1′		
Charge		**+2**		**-1**					**+2**		**+3**
$URM1^{N}$		6′	2	1′			11		1, 1′		
Charge		**+1**	**-1**	**-1**					**+2**		**+1**
$URM1^{+}$		6′	2	1′				11	1, 1′		
Charge		**+1**	**-1**	**-1**				**+1**	**+2**		**+2**
SUMO1		6′	2	1, 5, 9, 1′					1, 1′		
Charge		**+1**	**-1**	**-4**					**+2**		**-2**
$TMEM59_{285\text{-}291}$		9, 11, 8′		1, 3′			5		1, 1′		
Charge		**+3**		**-2**					**+2**		**+3**
$TMEM59_{298\text{-}305}$		9, 11, 3′	8′	1, 7′			5		1, 1′		
Charge		**+3**	**-1**	**-2**					**+2**		**+2**
$CoV2N_{335\text{-}341}$		9, 11	7′, 8′	1			5		1, 1′		
Charge		**+2**	**-2**	**-1**					**+2**		**+1**




```
@<TRIPOS>MOLECULE
LYC
   21    20     1     0     0
SMALL
resp


@<TRIPOS>ATOM
      1 C1           2.2290     0.0980    -0.0000 C          1 LYC       0.811218
      2 O1           2.4440     1.3200     0.0000 O          1 LYC      -0.641673
      3 N1           0.9840    -0.4010    -0.0000 N          1 LYC      -0.509681
      4 C2           3.3530    -0.9010     0.0000 CT         1 LYC      -0.601831
      5 C3          -0.1910     0.4500    -0.0000 CT         1 LYC      -0.136733
      6 C4          -1.4590    -0.3840    -0.0000 CT         1 LYC       0.002717
      7 H1          -0.1660     1.1040    -0.8810 H1         1 LYC       0.104686
      8 H2          -0.1660     1.1040     0.8810 H1         1 LYC       0.104686
      9 C5          -2.7140     0.4790     0.0000 CT         1 LYC       0.121031
     10 H3          -1.4630    -1.0390     0.8820 HC         1 LYC       0.025164
     11 H4          -1.4630    -1.0390    -0.8820 HC         1 LYC       0.025164
     12 C6          -3.9900    -0.3510     0.0000 CT         1 LYC      -0.239027
     13 H5          -2.7000     1.1380    -0.8790 HC         1 LYC      -0.004123
     14 H6          -2.7000     1.1380     0.8790 HC         1 LYC      -0.004123
     15 H7          -4.0400    -0.9960    -0.8850 HC         1 LYC       0.055081
     16 H8          -4.8810     0.2860     0.0000 HC         1 LYC       0.055081
     17 H9          -4.0400    -0.9960     0.8850 HC         1 LYC       0.055081
     18 H10          0.8530    -1.4030    -0.0000 H          1 LYC       0.288578
     19 H11          3.0070    -1.9380     0.0000 HC         1 LYC       0.162901
     20 H12          3.9780    -0.7320     0.8830 HC         1 LYC       0.162901
     21 H13          3.9780    -0.7320    -0.8830 HC         1 LYC       0.162901
@<TRIPOS>BOND
     1     1     2 2
     2     1     3 1
     3     1     4 1
     4     3     5 1
     5     3    18 1
     6     4    19 1
     7     4    20 1
     8     4    21 1
     9     5     6 1
    10     5     7 1
    11     5     8 1
    12     6     9 1
    13     6    10 1
    14     6    11 1
    15     9    12 1
    16     9    13 1
    17     9    14 1
    18    12    15 1
    19    12    16 1
    20    12    17 1
@<TRIPOS>SUBSTRUCTURE
     1 LYC         1 TEMP              0 ****  ****    0 ROOT
```

Figure S7.28: The structure of *N*-butyl acetamide was used as a model for parametrising the N^{ε}-branched lysine residue (LYC). The mol2 file displays the determined AMBER atom types[7] and RESP charges.[8, 9]

```
[ LYC ]
 [ atoms ]
     N    N           -0.41570     1
     H    H            0.27190     2
    CA    CT          -0.07206     3
    HA    H1           0.09940     4
    CB    CT          -0.04845     5
   HB1    HC           0.03400     6
   HB2    HC           0.03400     7
    CG    CT           0.06612     8
   HG1    HC           0.01041     9
   HG2    HC           0.01041    10
    CD    CT           0.00272    11
   HD1    HC           0.02516    12
   HD2    HC           0.02516    13
    CE    CT          -0.13673    14
   HE1    H1           0.10469    15
   HE2    H1           0.10469    16
    NZ    N           -0.43370    17
   HZ1    H            0.28858    18
     C    C            0.59730    19
     O    O           -0.56790    20
 [ bonds ]
     N     H
     N    CA
    CA    HA
    CA    CB
    CA     C
    CB   HB1
    CB   HB2
    CB    CG
    CG   HG1
    CG   HG2
    CG    CD
    CD   HD1
    CD   HD2
    CD    CE
    CE   HE1
    CE   HE2
    CE    NZ
    NZ   HZ1
     C     O
    -C     N
 [ impropers ]
    -C    CA     N     H
    CA    +N     C     O
```

Figure S7.29: The added N^{ε}-branched lysine residue (LYC) in GROMACS topology.[4] Standard charges from a neutral lysine residue (LYN) in the AMBER99FFSB-ILDN forcefield[6] were retained for atoms in the backbone and sidechain up to CG/HG1/HG2. Remaining atom charges were assigned based on RESP calculations, with residual charge correction applied on the NZ atom. No new atom types or bonded parameters were added. Two amide improper dihedrals (centred at GLY_C and LYC_NZ) involving this bond were added to the topology.

References

1. Zhang, Y.; Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019,** *35*, 5121.
2. Chan, H. T. H.; Brewitz, L.; Strain-Damerell, C.; Lukacik, P.; Walsh, M. A.*, et al.* Studies on the selectivity of the SARS-CoV-2 papain-like protease reveal the importance of the P2' proline of the viral polyprotein. *bioRxiv* **2023**, 2023.07.11.548309.
3. Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D.*, et al.* Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **2020,** *6*, eabd4596.
4. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.*, et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.
5. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.*, et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236.
6. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950.
7. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.*, et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179.
8. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269.
9. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993,** *115*, 9620.