Studies of Antibody Conformational Ensembles Towards Better Understanding of Structure And Function



Bora Guloglu

Merton College

University of Oxford

Submitted in partial completion of the

Doctor of Philosophy

Michaelmas 2024

Acknowledgements

I was always told that a DPhil is a solitary endeavour, but it turns out that it takes a village. I am deeply grateful to my supervisors, Matt and Charlotte, both fantastic mentors. From the day I came to Charlotte with my project proposal, she has pushed me to do better science, to explore new approaches, and to keep going even when a project seems hopeless - I am all the richer for having been a part of her group, OPIG. Matt allowed me to do the project I wanted to do, even if it meant that there was a random antibody person wandering around the Higgins group. He has taught me to chase the interesting stories, to trust my instincts, and to be critical, always. I cannot thank either of you enough.

My sincerest thanks also go to all my wonderful colleagues and friends, both in OPIG and the Higgins group. Especially to Sam, Alex, Hannah, Stephan and Brendan - thank you for letting me be a nuisance whenever I was bored in the office. Maybe I won't miss the terrible sandwiches, but I'll miss lunch with you all, at 12 pm sharp. Thank you for the impromptu pub trips, the podcast discussions, and for adopting this DPhil student into the ranks of the cool post-docs. Thank you also to Alissa, Sarah, Eve, and everyone else in OPIG whom I dragged out for walks around Uni park, cornered with conversations about flying trapeze, and who made me feel welcome even if for a year it was through private messages during Zoom calls.

So many friends, old and new, have supported me throughout this DPhil. Patrick, Angus, and Anna, thank you for making a terrible 2020 into one of the best summers I have had, with picnics, cosmos, and lots of focaccia. Leonie, Helena, and all my wonderful housemates, thank you for making Oxford feel like home, introducing me to climbing, and eating my failed croissants (one day...). Molly, Maria, Jess, and Immy, thank you for the dinners, Cowley Retreat trips, and for keeping me up-to-date on all the St. Hilda's gossip. Ozge, Kerem, Defne, Ella, Elif, Cem, and Joel, thank you for the multi-hour calls, making the distance feel non-existent. I miss you all so much. Alice, I couldn't have done this without you. Thank you for the wine and cheese evenings with trashy television and sorry about the late dinners.

To my wonderful family, thank you for all your support and giving me perspective. Barbara, Gaby, Rudi, Dania, and Marcel, I am so glad that I can finally come back home for Christmas again after the lockdowns. Dids, Dave, Lucy, and Conlan, I feel so lucky to call you my family. Melisa, Ela, and Hasan, thank you for supporting

me across continents. I love you all. Thank you to my parents for all the sacrifices you have made for me. I would not be here without your love. Finally, thank you James, my best friend, for everything. You were right, it all turns out okay.

Work Ownership

The use of the first person plural in this thesis is a stylistic choice. All work that is presented in this thesis was conducted by me, expect where stated. The work conducted by collaborators has been described, and the individuals are named. Appendix D contains additional work not included in the thesis. In these, author contribution statements within each publication clearly state the work which I conducted.

Abstract

In this thesis, we examine relationships between structural, biophysical and dynamic properties of antibodies. First, we investigate determinants of the binding rates of antibodies, showing that the flexibility of the third complementarity-determining region of the heavy chain (CDR-H3) and the epitope are predictors of this, while analysis of static structures does not yield information on binding rates. We also show that consideration of the conformational ensemble of an antibody gives an enriched view of the surface properties that contribute heavily to developability of an antibody therapeutic. Next, we investigate the sources of CDR-H3 flexibility and structural diversity. We provide evidence in support of the long-standing hypothesis that CDR-H3 structure and flexibility are fine-tuned by the rest of the antibody and propose a model for a steric cage that other CDR loops form around CDR-H3, thereby modulating its conformational space. We also demonstrate the effects of different genetically encoded structural attributes on the flexibility and structural variability of CDR-H3, showing that certain genetic lineages of antibodies have preferences for CDR-H3 loops of differing levels of flexibility. Lastly, we train deep learning models to improve predicted models of antibody structures, demonstrating that errors made by state-of-the-art structure prediction algorithms can effectively be learned.

Contents

List of Figures **xii**

List of Tables **xv**

List of Abbreviations **xvi**

1 Introduction **1**

1.1 Chapter abstract 1
1.2 Antibodies 2
1.2.1 Antibodies as part of adaptive immunity 2
1.2.2 Affinity maturation 3
1.2.3 Antibody organisation and structure 4
1.2.3.1 Organisation of antibody domains 4
1.2.3.2 Genetic lineages and sequence diversity of antibodies 6
1.2.3.3 Antibody structure 7
1.2.4 Antibody numbering schemes and regions 10
1.3 Alternate forms of antibodies 10
1.4 Antibody structural modelling 13
1.4.1 Homology-based and *ab initio* modelling 13
1.4.2 Deep-learning focused approaches 14
1.5 Antibody dynamics and conformational ensembles 16
1.5.1 Antibodies as conformational ensembles 17
1.5.2 The effect of affinity maturation on antibody dynamics . . . 19
1.6 The antibody-antigen interaction 19
1.6.1 Antibody binding in a dynamic context 20
1.7 Therapeutic antibodies 21
1.7.1 Discovery of therapeutic antibodies 22
1.7.2 Developability liabilities 23
1.8 Thesis outline 24

2 Theory **27**
2.1 Chapter abstract . . . 27
2.2 Molecular dynamics simulations . . . 28
2.2.1 Molecular mechanics . . . 28
2.2.1.1 Force fields . . . 29
2.2.1.2 Energy minimisation . . . 31
2.2.2 Molecular dynamics . . . 31
2.2.2.1 Integration algorithms . . . 32
2.2.2.2 Temperature and pressure coupling . . . 33
2.2.2.3 Periodic boundary conditions . . . 34
2.2.3 Well-tempered metadynamics simulations . . . 35
2.2.3.1 Procedure . . . 36
2.2.3.2 Recovery of equilibrium ensembles . . . 37
2.3 Deep learning and graph neural networks . . . 38
2.3.1 Machine learning . . . 38
2.3.2 Training neural networks . . . 39
2.3.3 Graph neural networks . . . 41

3 Contributions of conformational ensembles to biophysical properties relevant to vaccinology and therapeutic antibody design **45**
3.1 Chapter abstract . . . 46
3.2 Introduction . . . 47
3.3 Methods . . . 51
3.3.1 Prediction of antibody binding rates . . . 51
3.3.1.1 Kinetic data set curation . . . 51
3.3.1.2 Static analyses of epitope and paratope structures . 52
3.3.1.3 Prediction of flexibility using molecular dynamics simulations . . . 54
3.3.1.4 Conformational ensemble analysis using Sphinx and MODELLER . . . 56
3.3.2 Analysis of therapeutic surface profiles . . . 58
3.3.2.1 Therapeutic antibody data set curation . . . 58
3.3.2.2 Generation of conformational ensembles using molecular dynamics simulations . . . 59
3.3.2.3 Profiling of surface properties . . . 60
3.4 Results . . . 60
3.4.1 A diverse set of antibody structures shows that antibody binding and dissociation rates are independent of each other 60
3.4.2 Physicochemical descriptors of static structures are poor predictors of binding rates in antibodies . . . 61

3.4.3 CDR-H3 length predicts binding rate constants of antibodies 66
3.4.4 CDR-H3 flexibility correlates with binding rates and can be probed using computationally inexpensive methods 68
3.4.5 Sphinx efficiently identifies differences in CDR-H3 flexibility 72
3.4.6 Secondary structure and surface flexibility of antigens influence antibody binding rates 76
3.4.7 Limited exploration of antibody conformational ensembles enriches profiling of therapeutic properties 77
3.5 Discussion and conclusions 85

4 The structure and structural variability of CDR-H3 loops are influenced by specific attributes of the V_L domain 88
4.1 Chapter abstract 89
4.2 Introduction 90
4.3 Methods 92
4.3.1 Comparison of paired and single-domain antibody CDR-H3 loops 92
4.3.1.1 Data set curation 92
4.3.1.2 Analysis of CDR-H3 loop orientations 93
4.3.2 Analysis of V_L domain influences on CDR-H3 conformations 96
4.3.2.1 Crystal data set curation 96
4.3.2.2 Region definitions 96
4.3.2.3 Crystal radial distribution functions 96
4.3.2.4 Molecular dynamics simulations 97
4.3.2.5 Dynamic time warping analysis 99
4.3.2.6 V_H-V_L pairing analysis 100
4.4 Results 101
4.4.1 Single-domain antibody CDR-H3 loops show different conformations from paired antibody CDR-H3 loops 101
4.4.2 Antibody V_H and V_L CDR loops are characterised by non-equivalent inter-loop contacts in crystal structures 105
4.4.3 Conformational rearrangements in CDR-H3 do not require major conformational changes in minor loops 107
4.4.4 Minor CDR loops create steric boundaries within which CDR-H3 samples conformational space 111
4.4.5 Antibody subtypes, genes, and CDR canonical forms show preference for CDR-H3 conformations with different levels of diversity 114
4.4.6 Identical V_H sequences show differences in CDR-H3 conformation when paired with different V_L sequences 119
4.5 Discussion and conclusions 121

5 Refinement of Predicted Antibody Structures Using Graph Neural Networks 125
5.1 Chapter abstract . . . 126
5.2 Introduction . . . 126
5.3 Methods . . . 129
5.3.1 Quality assessment of structures predicted by GNNRefine . . 129
5.3.2 Training, validation, and test set curation . . . 130
5.3.3 Graph neural network architecture . . . 131
5.3.3.1 Input Features . . . 132
5.3.3.2 Architecture . . . 134
5.3.4 Training schemes . . . 136
5.3.5 Calculation of predicted distances and comparison to ABodyBuilder2 . . . 137
5.3.6 Refinement of structures using OpenMM . . . 138
5.4 Results . . . 138
5.4.1 GNNRefine is unable to consistently improve antibody models generated by ABodyBuilder2 . . . 138
5.4.2 An expanded data set of full-length Fvs can be generated from the PDB by modelling terminal residues . . . 142
5.4.3 Graph neural networks trained on antibody structures predict improved inter-Cβ distances . . . 144
5.4.4 Trained graph neural networks improve structures predicted to be of lower quality . . . 147
5.4.5 Energy minimisation using modified molecular mechanics force fields is unable to translate predicted distances to structural modifications . . . 149
5.4.6 Inter-Cβ distances predicted by our models obey the triangle inequality . . . 151
5.5 Discussion and conclusions . . . 153

6 Discussion, Conclusions and Future Work 155
6.1 Discussion and conclusions . . . 155
6.2 Future work . . . 158
6.3 Concluding remarks . . . 159

Appendices

A Supplementary Material for Chapter 3 163
A.1 Kinetic data set . . . 163
A.2 Flags used to assess antibody-antigen interactions using Arpeggio . 170
A.3 Flags used to model CDR-H3 loops using Sphinx . . . 170
A.4 Supplementary figures . . . 170

B Supplementary Material for Chapter 4 **180**
B.1 Example of dynamic time warping algorithm 180
B.2 Supplementary figures 182

C Supplementary Material for Chapter 5 **184**
C.1 Lists of antibodies used to train, test, and validate models 184
C.2 Algorithmic details for message passing layers 214
C.3 Supplementary figures 215

D Additional Publications **220**

References **241**

List of Figures

1.1 Diagram of an IgG antibody 5
1.2 Gene boundaries in an F_V region 7
1.3 Structure of an F_V region 8
1.4 Comparison of the IMGT and Chothia numbering schemes 11
1.5 Domain architectures and variable domain structures of IgNARs and camelid heavy chain antibodies 12

2.1 Diagram of periodic boundary conditions 35
2.2 Diagram of well-tempered metadynamics simulations 37
2.3 Diagram of message passing layers in graph neural networks 43

3.1 Coverage of kinetic data in the kinetic data set. 62
3.2 Coverage of genetic attributes in the kinetic data set. 63
3.3 Correlations between binding rates and contacts made between antigens and antibodies 65
3.4 Correlations between CDR loop lengths and binding rates. 67
3.5 Correlations between antibody flexibility and binding rates. 71
3.6 Comparisons of Sphinx and MD CDR-H3 conformational ensembles. 72
3.7 Performance of Sphinx in predicting unstructured loops. 73
3.8 Performance of Sphinx in prediciting effects of affinity maturation on CDR-H3 flexibility 75
3.9 Differences in binding kinetics based on epitope secondary structure. 78
3.10 Changes in surface hydrophobicity in of antibodies during simulation 81
3.11 Changes in surface positive patches in of antibodies during simulation 82
3.12 Changes in surface negative patches and asymmetry in the net heavy- and light-chain surface charges in of antibodies during simulation . 83

4.1 Depiction of how CDR-H3 loop orientation and reach is described using spherical coordinates. 95
4.2 Distributions of spherical coordinate descriptors of CDR-H3 conformations. 102
4.3 Relationships between spherical coordinate descriptors of CDR-H3 conformations. 103

4.4 Radial distribution functions showing CDR loop packing in crystal structures. 106
4.5 Flexibility of CDR loops in metadynamics simulations 108
4.6 Conformations explored by CDR-H3 through CDR loop contacts . . 112
4.7 Changes in loop heavy atom packing patterns in simulation and crystal structure data sets 113
4.8 Conformations of CDR loops in metadynamics simulations 115
4.9 Effects of genetically encoded antibody attributes on CDR-H3 conformational diversity 116
4.10 Correlations between V_L sequence identity and V_H RMSD. 120
4.11 Dependence of V_L sequence identity influence on V_H structure on V_L canonical forms. 121
4.12 Case study demonstrating the caged CDR-H3 model 122

5.1 GNNRefine architecture 133
5.2 Details of the GNNRefine message passing layers 135
5.3 Performance of unmodified GNNRefine on ABB2 models 140
5.4 Performance of modified GNNRefine versions on ABB2 models . . . 141
5.5 Distribution of first residues resolved in F_V crystal structures 143
5.6 Cumulative distributions of structures with missing terminal residues 144
5.7 Performance of our models across different regions of antibodies . . 145
5.8 Changes in MAE of inter-Cβ distances compared to ABB2 146
5.9 Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 error 148
5.10 Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 predicted error 150
5.11 Changes in RMSD of refined structures compared to ABB2 152

A.1 Distribution of species in the kinetic data set 170
A.2 Background distribution of genetic attributes across SAbDab. . . . 171
A.3 Distribution of CDR loop lengths in the kinetic data set 171
A.4 Correlations between predicted flexibility and CDR-H3 loop length 172
A.5 Correlations between antibody flexibility and unbinding rates. . . . 173
A.6 RMSDs of assessed CSTs from the starting structure over the course of MD simulations 175
A.7 Changes in surface hydrophobic patches of antibodies during simulation for all tested CSTs 176
A.8 Changes in surface positive patches of antibodies during simulation for all tested CSTs 177

A.9 Changes in surface negative patches of antibodies during simulation for all tested CSTs 178
A.10 Changes in asymmetry in the net heavy-and light-chain surface charges (SFvCSP) scores in of antibodies during simulation for all tested CSTs 179

B.1 Example of dynamic time warping distance calculation 181
B.2 Radial distribution functions of CDR loops of crystal structures separated by light chain type 182
B.3 Radial distribution functions of CDR loops in MD simulations . . . 183

C.1 Further performance of modified GNNRefine versions on ABB2 models 215
C.2 Changes in MAE of inter-Cβ distances compared to ABB2 216
C.3 Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 error 217
C.4 Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 predicted error 218
C.5 Changes in RMSD of refined structures compared to ABB2 219

List of Tables

3.1 Residue groupings used for SASA calculations 52
3.2 Equilibration protocol for MD simulations 55
3.3 Correlations between binding rates and buried surface area 64
3.4 Summary of naive and affinity matured antibody CDR-H3 loops used in validation 74

A.1 Rate constants of the kinetic data set 163
A.2 List of CSTs used for assessment of TAP properties in conformational ensembles 174
A.3 Values used as thresholds for TAP flags 174

List of Abbreviations

Ab . . . . . antibody

ABB2 . . . . . ABodyBuilder2

AF2 . . . . . AlphaFold2

AFM . . . . . AlphaFold-Multimer

Ag . . . . . antigen

BCR . . . . . B-cell receptor

BSA . . . . . buried surface area

CDR . . . . . complementarity-determining region

C_H . . . . . constant domain of the heavy chain

C_L . . . . . constant domain of the light chain

CST . . . . . clinical-stage therapeutic

CV . . . . . collective variable

DTW . . . . . dynamic time warping

EGNN . . . . . E(n) equivariant graph neural network

Fab . . . . . antigen-binding fragment

F_C . . . . . crystallisable fragment

F_V . . . . . variable fragment

FW . . . . . framework

GC . . . . . germinal center

GNN . . . . . graph neural network

Ig . . . . . immunoglobulin

IgNAR . . . . immunoglobulin new antigen receptors

K_D . . . . . equilibrium dissociation constant

k_{off} . . . . . dissociation rate constant

k_{on} . . . . . binding rate constant

LSTM . . . . . long short-term memory

mAb . . . . . monoclonal antibody

MAE . . . . . mean absolute error

MC . . . . . Monte-Carlo

MD . . . . . molecular dynamics

ML . . . . . machine learning

MM . . . . . molecular mechanics

MPL . . . . . message-passing layer

MSA . . . . . multiple sequence alignment

MSE . . . . . mean squared error

NPT . . . . . constant number of particles, pressure, and temperature

OLL . . . . . ordinal log loss

PDB . . . . . Protein Data Bank

PNC . . . . . patches of negative charge

PPC . . . . . patches of positive charge

PSH . . . . . patches of surface hydrophobicity

RDF . . . . . radial distribution function

ReLu . . . . . rectified linear unit

RMSD . . . . . root mean squared deaviation

RMSF . . . . . root mean squared fluctuation

RMSPE . . . . root mean squared predicted error

SAbDab . . . . Structural Antibody Database

SASA . . . . . solvent-accessible surface area

sdAb . . . . . single-domain antibody

SFvCSP . . . . Structural F_v Charge Symmetry Parameter

TAP . . . . . Therapeutic Antibody Profiler

TheraSAbDab Therapeutic Structural Antibody Database

vdW . . . . . van der Waals

$\mathbf{V_H}$ . . . . . variable domain of the heavy chain

$\mathbf{V_HH}$ . . . . . variable domain of single-domain antibodies

$\mathbf{V_L}$ . . . . . variable domain of the light chain

VNAR . . . . variable domain of immunoglobulin new antigen receptors

WHO . . . . . World Health Organisation

XRC . . . . . X-ray crystallography

1 Introduction

Contents

1.1 Chapter abstract 1
1.2 Antibodies 2
1.2.1 Antibodies as part of adaptive immunity 2
1.2.2 Affinity maturation 3
1.2.3 Antibody organisation and structure 4
1.2.4 Antibody numbering schemes and regions 10
1.3 Alternate forms of antibodies 10
1.4 Antibody structural modelling 13
1.4.1 Homology-based and *ab initio* modelling 13
1.4.2 Deep-learning focused approaches 14
1.5 Antibody dynamics and conformational ensembles . . 16
1.5.1 Antibodies as conformational ensembles 17
1.5.2 The effect of affinity maturation on antibody dynamics 19
1.6 The antibody-antigen interaction 19
1.6.1 Antibody binding in a dynamic context 20
1.7 Therapeutic antibodies 21
1.7.1 Discovery of therapeutic antibodies 22
1.7.2 Developability liabilities 23
1.8 Thesis outline 24

1.1 Chapter abstract

In this chapter, we describe the necessary background information that forms the basis of the research detailed in the rest of the thesis. We begin by describing

antibodies, or immunoglobulins, in terms of their function, domain organisation, structure with emphasis on the ability of the immune system to generate a vast space of antibodies starting from a small number of genes (Section 1.2). We then briefly outline alternate forms of antibodies, where the antigen-binding fragments are markedly different from human antibodies (Section 1.3). Next, we provide an account of how antibody structures are predicted using structural modelling (Section 1.4). We begin by describing homology-based approaches and follow this up by discussing current state-of-the-art methods, which heavily rely on artificial intelligence. Having established the structure and function of antibodies, we then move to a dynamic picture (Section 1.5), and conceptualise antibodies as conformational ensembles discussing how this approach allows for a more complete explanation of the structural basis of antibody function, including the binding event between antibodies and their antigens (Section 1.6). Lastly, we provide a brief account of how antibodies can be used as therapeutics (Section 1.7), focusing on liabilities that may hinder the development of a candidate antibody.

1.2 Antibodies

1.2.1 Antibodies as part of adaptive immunity

The adaptive immune system in jawed vertebrates features antibodies, also known as immunoglobulins. These globular proteins circulate in the bloodstream, binding to foreign entities termed antigens. An antibody attaches to a specific region on an antigen, known as an epitope, and each antigen can present multiple epitopes. The parts of an antibody that interact with with an antigen are called paratopes.

B-cells, a type of blood-circulating cell, produce antibodies. Initially, they generate "naive" antibodies with low affinity for antigens and present these antibodies on their surface as B-cell receptors (BCR). During an immune response, the antibodies' affinity for their antigens increases through a process called affinity maturation (Alberts et al. 2022). Concurrently, the number of B-cells producing a particular antibody also rises through clonal expansion (Adams et al. 2020), resulting in a large quantity of "mature" antibodies that drive the immune response. When an

antibody binds to its specific epitope, it blocks the surface of the antigen, preventing harmful interactions of the antigen with other host systems, and triggers the immune system to recruit cells that help destroy the foreign body (Lu et al. 2017).

After undergoing affinity maturation and clonal selection, mature B-cells evolve into either plasma cells or memory B-cells. Plasma cells mainly operate from the bone marrow and secrete antibodies in large amounts to target and eliminate current antigenic threats. Memory B-cells, in contrast, are long-lived and circulate throughout the body. They rapidly produce antibodies without needing clonal expansion or affinity maturation upon re-encounter with an antigen, thus accelerating the immune response (Alberts et al. 2022).

1.2.2 Affinity maturation

It is thought that the diversity of possible epitopes is vastly higher than the number of possible naive antibody sequences (the origins of antibody sequence diversity will be outlined in Section 1.2.3.2), with a theoretical upper limit placed at 10^{10} - 10^{11} (Greiff et al. 2017; Briney et al. 2018). The number of B-cells in the human body has recently been placed at 3×10^{11} (Sender et al. 2023), although this number includes mature B-cells, suggesting that the actual number of naive antibody sequences in a repertoire is lower than this theoretical upper limit. Consequently, a single naive antibody must be able to target multiple potential epitopes (polyspecific binding) for complete coverage of possible antigen space. Indeed, recent work has shown that naive antibodies exhibit weak but polyspecific binding (Willis et al. 2013; Shehata et al. 2019), and that mature antibodies tend to lose this property in favour of increased affinity for the target antigen (Shehata et al. 2019).

During the initial antigenic insult, multiple clonal lineages are usually activated (Chernyshev et al. 2023; Kotagiri et al. 2021), binding the antigen with relatively weak affinity (Alberts et al. 2022). The B-cells expressing these binding-competent BCRs then migrate to microanatomical structures in the lymphoid organs, the germinal centers (GC) (Berek et al. 1991), which are the site of affinity maturation. Within the GC, naive antibodies first undergo rounds of cellular division, followed

by somatic hypermutation biased to the antigen-binding regions of BCRs by way of activation-induced cytidine deaminase (Muramatsu et al. 2000). This enzyme introduces mutations to the genes encoding the BCR, which in turn are repaired using error-prone mechanisms, thereby resulting in diversification of the naive BCR sequence. Within the GC, B-cells are also placed in a state of programmed cell death (apoptosis) (Martinez-Valdez et al. 1996). Other cells within the GC produce chemical signals that inhibit the apoptotic pathway, thereby resulting in survival of the mutated B-cells (Fischer et al. 1998). However, whether a mutated B-cell gains access to these survival signals depends on its affinity for the antigen (Gitlin et al. 2014). As such, high affinity B-cells survive at higher rates than lower-affinity, driving the establishment of a cell population with high affinity for the antigen via this Darwinian mechanism. This cycle is repeated multiple times, with each generation resulting in approximately one mutation to the BCR (Rajewsky et al. 1987). At some point, a final signal causes the now matured B-cells to exit the GC and differentiate into the aforementioned mature plasma and memory B-cells (Inoue et al. 2020). The affinity maturation process takes place on the scale of several days to several weeks (Alberts et al. 2022). Therefore, the presence of memory B-cells provides a significant advantage in the case of a repeated infection with the same antigen or one with a sufficiently similar epitope.

1.2.3 Antibody organisation and structure

1.2.3.1 Organisation of antibody domains

Human antibodies all share the ability to bind antigens and can be found in one of five classes (IgG, IgD, IgE, and IgM in humans), with IgG being the dominant class, making up roughly $\frac{3}{4}$ of circulating human antibodies. The different classes of antibodies differ only in their constant regions which do not actively contribute to antigen recognition, but have different downstream effects[1]. A typical antibody is a "Y"-shaped heterotetrameric protein (Fig 1.1), made up of three broad regions and two identical heavy chains and two identical light chains. The two arms of the

[1]From hereon, we use the IgG class to discuss the structure of an example antibody.

"Y" are identical in sequence and structure, with the stalk of the antibody differing from these. Each arm of the "Y" is composed of one entire copy of the light chain and encompasses roughly half of the length of one heavy chain, whereas the stalk of the "Y" is made up of two identical halves of the heavy chain.

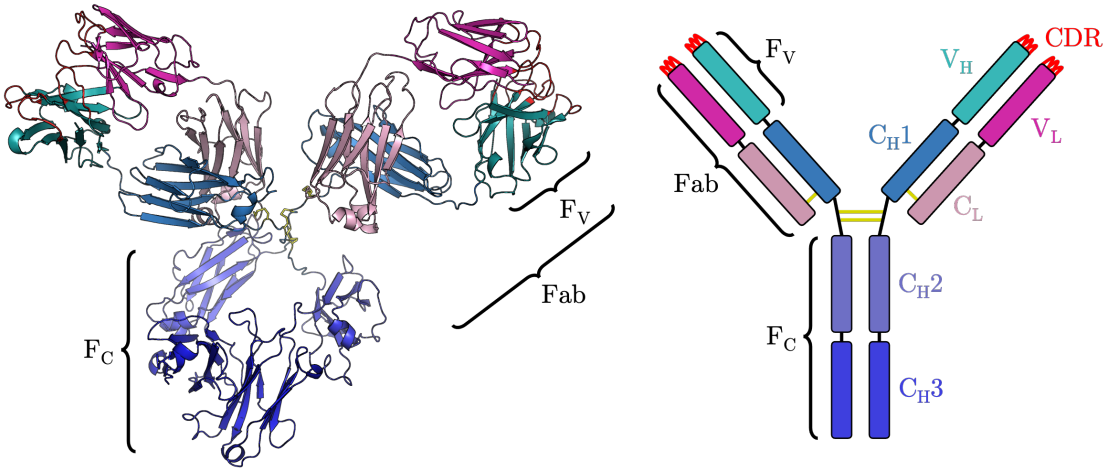


Figure 1.1: Diagram of an IgG antibody. On the left is the crystal structure of an antibody in cartoon representation (Protein Data Bank (PDB) accession code: 1IGY). The different sub-regions of the antibody are annotated, CDR loops are coloured red, and disulphide linkages are shown in yellow using stick representation. On the right, a diagrammatic representation of the structure on the left, with individual domains labelled and their connectivity shown.

Each of the heavy chains in an antibody is composed of four domains named the variable heavy (V_H) and the three constant heavy (C_H1-3). The light chains of an antibody contain the N-terminal variable light (V_L) domain followed by a single constant light (C_L) domain. The domains all present a characteristic immunoglobulin fold, consisting of a two-layer beta sandwich consisting of roughly eight β-strands, with the two layers being held together by a canonical disulphide linkage (Bork et al. 1994). Each arm of the antibody is made up of the V_H and V_L domain at the tip, the combination of which is referred to as the variable fragment (F_V) region, followed by the C_H1 and C_L domains, which are connected to one another by a disulphide linkage. Together, the four domains are referred to as the antigen-binding fragment (Fab). The domains in the F_V region exhibit the highest amount of sequence variability and contain three complementarity-determining region (CDR) loops on each domain (CDRH-1-3 and CDRL-1-3 on the V_H and V_L

domains, respectively), with the CDRs providing the major modality for antigen recognition and binding (Alberts et al. 2022; Akbar et al. 2021). The stalk of the antibody in turn is made up of two copies each of the C_H2 and C_H3 domains. Together, these four domains are referred to as the crystallisable fragment (F_C). The C_H1 and C_H2 domains are connected by a linker, and the linkers of the two heavy chains are connected to one another by further disulphide linkages.

1.2.3.2 Genetic lineages and sequence diversity of antibodies

As stated in Section 1.2.2, the size of the naive antibody repertoire is theoretically 10^{10} - 10^{11} (Greiff et al. 2017; Briney et al. 2018), which far exceeds the size of the number of protein-coding genes in the human genome (Salzberg 2018). The V_H domain is genetically encoded by three genes, the V-, D-, and J-genes. Similarly, the V_L domain is encoded by the V- and J-genes. In a haploid human genome, there exist 123-129 V-genes, 27 D-genes, and 9 J-genes in the IGH locus of chromosome 14 (Lefranc et al. 2001), yet each full V_H domain contains only one gene from each category. During the differentiation of a naive B-cell, a somatic recombination process called V(D)J recombination (Roth 2014) randomly[2] excises DNA between different D- and J-genes, joining the ends of the remaining DNA. Next, the excision is repeated between the D-gene that remained and a random V-gene. As a result, the genome of the B-cell now only contains one gene per category, and the expression of these forms a functional V_H domain (which is expressed together with constant domains as part of a functional BCR). Similarly, V_L domains are formed by this process, though the only junction is the V-J junction, resulting in less combinatorial diversity when compared to the V_H domain (Fig 1.2). The V- and J-genes that make up the V_L domain are found on two loci: the IGK locus on chromosome 2 (76 and 5 V- and J-genes, respectively) and the IGL locus on chromosome 22 (73-74 and 7-11 V- and J-genes, respectively) (Lefranc et al. 2001). As such, there exist two subtypes of light chains, termed κ and λ, based on whether they stem from the IGK or IGL locus, respectively.

[2]It has been shown that the removal process is not truly random, and that preferential combinations exist in reality (Boyd et al. 2010; Kidd et al. 2016).

V_H: QVQLVQSGAEVRKPGASVKVSCKASGYSLKDHYMVWVRQAPGQGLE
WMGWINPQSGGTGYGQKFQGRVTMTRDTSTNTAYMILSSLRSDDTA
VYFCARDGAKTVSNSGLSLLYYHNRLDAWGQGTMVTVSS

V_L: QSVLTQPPSVSAAPGQMVTISCSGSSSNIGKNYVSWYQQLPGAAPK
LLIFDNNKRPSGTPDRFSGSKSGTSATLVITGLQTGDEADYYCGTP
DRSLSVVFGGGTKVTVLG

Figure 1.2: Gene boundaries in an F_V region. Shown are the V_H and V_L sequences of an F_V region. The CDR loops in each case are underlined (CDR1-3 are in order), and approximate boundaries of V- (red), D- (blue), and J-genes (yellow) are shown.

Further sequence diversity in naive antibodies is provided by non-template nucleotide addition, where nucleotides are added at the junction of two genes, increasing junctional diversity (Gilfillan et al. 1995). Further, the imprecise nature of the enzymatic recombination can also lead to a loss of nucleotides (Jackson et al. 2004). Together, this enables a relatively small number of genes to generate a vast space of F_V sequences, from which binding-competent BCRs are sampled by the immune system.

1.2.3.3 Antibody structure

As discussed, the sequence diversity of the F_V domains is much higher compared to that of the constant domains of an antibody. This is especially apparent in the CDR loops, which tend to be the main targets of somatic hypermutation (Jolly et al. 1996), due to their role in binding antigens. Usually, the variable domains of the F_V are separated into the framework (FW) region and CDR loops based on structural and functional properties. An example of an F_V domain is shown in Figure 1.3.

Framework regions The framework region of an F_V is composed mainly of β-sheets in a sandwich configuration. It is marked by its relatively low levels of sequence and structural diversity in crystal structures in comparison to the CDR loops. For the vast majority of antibodies, the FW does not take part in antibody-antigen interactions (Akbar et al. 2021), and is often thought to act as a modality to position and anchor the CDR loops. A set of residues that are part of the FW

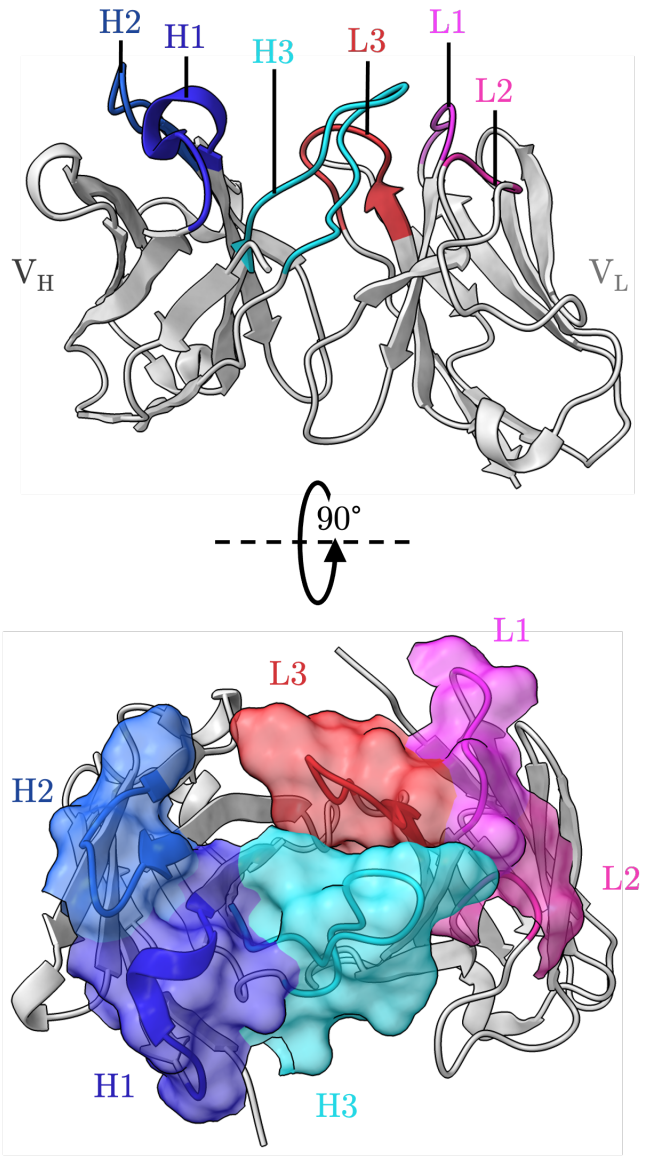


Figure 1.3: Structure of an F_V region. Shown is the crystal structure of an antibody F_V region, with CDR loops labelled according to the IMGT definition shown in cartoon representation, with density of the CDR loops overlain (PDB accession code: 1MLC). The two domains of the F_V exhibit pseudosymmetry, with the CDR loops being oriented and positioned similarly across the V_H and V_L domains. The CDR-H3 loop, which generally is most heavily involved in binding is found in the center of the F_V along with CDR-L3 when regarded from above.

region and are in the vicinity of the CDR loops is collectively known as the Vernier zone (Foote et al. 1992). Further work has shown that the Vernier zone is critical to accommodate and stabilise various conformations of the more structurally variable CDR loops (Makabe et al. 2008). While the majority of the FW region shows little structural diversity, the orientation with which the V_H and V_L domains pack against one another (described using a set of five angles and one distance metric) can be varied across different combinations that make up the F_V region, and is determined primarily by the framework residues (Dunbar et al. 2013).

Minor CDR loops The CDR loops of an antibody are often divided into two groups. The minor CDRs (also referred to as the canonical CDRs) are CDR-H1-2 and CDR-L1-3. These are referred to as minor due to the fact that they usually contribute less to antigen binding (Akbar et al. 2021), making fewer contacts with the epitope. Further, compared to CDR-H3, they contain less sequence diversity (Wong et al. 2019). This is primarily due to the fact that only the V-gene makes sequence contributions to these CDR loops during V(D)J recombination (Fig 1.2). Usually, the minor CDRs are found in predictable, defined conformations based on their sequence, termed canonical forms (Chothia et al. 1987; North et al. 2011; Nowak et al. 2016). The canonical forms have traditionally been defined per loop for each loop length, but recent work has demonstrated that the same canonical form can be found across multiple loop lengths (Nowak et al. 2016).

The CDR-H3 loop The CDR-H3 loop is the most diverse loop across all the CDRs, with markedly higher length, sequence, and structural variance compared to the minor loops (Wong et al. 2019; Weitzner et al. 2015; Regep et al. 2017). It also provides the major modality of antigen binding, contributing on average a larger number of residues to the antibody-antigen interface (Akbar et al. 2021), by virtue of its higher average length (Wong et al. 2019). This provides more surface for it to contribute to the binding interface, and this, coupled with its higher sequence diversity, enables it to target more physicochemically diverse surfaces. Indeed, the conformational diversity of CDR-H3 is higher than the baseline conformational

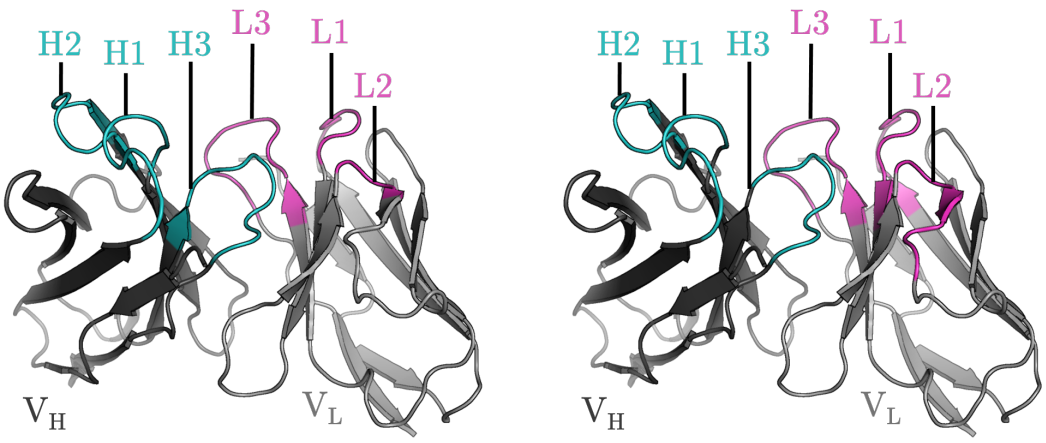
diversity of loops across the proteome, with 30% of CDR-H3 loops in the Protein Data Bank (PDB) representing novel conformations, as opposed to 3% of non-CDR-H3 loops (Regep et al. 2017). CDR-H3 cannot be clustered into canonical forms due to its diversity (Regep et al. 2017). The sequence diversity of naive CDR-H3 loops stems from the genetic contributions to their sequences, with the V-, D-, and J-genes all encoding parts of the loop (Fig 1.2).

1.2.4 Antibody numbering schemes and regions

As antibodies all share this common genetic lineage and use specific areas for binding, numbering schemes have been devised, making use of these features to allow standardised comparison (Wu et al. 1970; Chothia et al. 1987; Abhinandan et al. 2008; Lefranc et al. 2003). The numbering schemes ensure that functionally similar positions in the antibody (for example, the conserved cysteine residues that form the canonical disulphide linkage in the FW, or the CDR loops) are assigned the same numbers, even when one sequence is longer by using insertions or deletions where necessary. In this thesis, we mainly use the IMGT numbering scheme (Lefranc et al. 2003) to define CDR loop boundaries, though our work in Chapter 3 requires use of the Chothia numbering scheme (Chothia et al. 1987), due at times to reliance on previously extracted CDR-H3 loop structures that were annotated using this scheme. A comparison of these two schemes is shown in Figure 1.4.

1.3 Alternate forms of antibodies

Antibodies are a relatively young class of proteins on evolutionary timescales, emerging only with jawed vertebrates. Traditional antibodies, primarily found in mammals, consist of two heavy and two light chains forming a "Y"-shaped structure (Fig 1.1). However, alternate forms like camelid heavy chain antibodies (Mitchell et al. 2018), whose variable domains are referred to as V_HHs (Fig 1.5a) (also referred to as single-domain antibodies [sdAb] or nanobodies), and IgNARs (Immunoglobulin New Antigen Receptors) (Fig 1.5b) (Khalid et al. 2022), diverge from this structure. IgNARs are unique to cartilaginous fish (chondrichthyes) such



V_H: QVQLQESGAEVMKPGASVKISCKAT**GYTFSTYW**IEWVKQRPGHGLE
WIGE**ILPGSGST**YYNEKFKGKATFTADTSSNTAYMQLSTSEDSAVY
YC**ARGDGNYGY**WGQGTTLTVS

V_L: DIELTQSPATLSVTPGDSVSLSCRAS**QSISNN**LHWYQQKSHESPRL
LIK**YVS**QSSSGIPSRFSGSGSGTDFTLSINSVETEDFGMYFC**QQSN**
SWPRTFGGGTKLEIK

Figure 1.4: Comparison of the IMGT and Chothia numbering schemes. Cartoon representation of the F_V region of and antibody is shown (PDB accession code: 1MLC) **(left)** shows CDR loops labelled according to IMGT definitions (Lefranc et al. 2003), whereas **(right)** shows CDR loops labelled according to Chothia definitions (Chothia et al. 1987). Below, the sequences of the V_H and V_L domains are given. CDR loops according to the IMGT scheme are shown in boldface, whereas loops according to the Chothia scheme are underlined.

as sharks and consist solely of heavy chains, with antigen-binding sites markedly different from classical antibodies. They have a single variable domain (VNAR) that binds the antigen, which is notable for its small size. Camelids have both paired antibodies and single-domain antibodies, and it has recently been shown that sdAbs and conventional antibodies are very similar in both the epitopes that they target as well as the antigen-binding surfaces that they possess (Gordon et al. 2023).

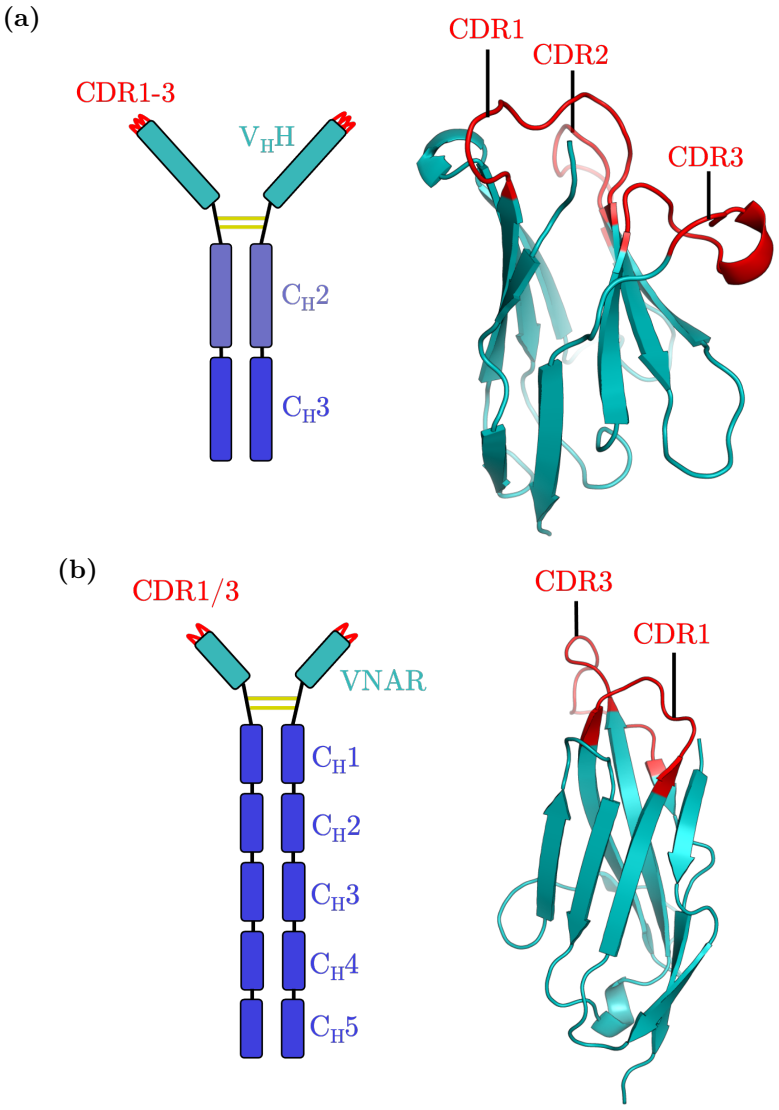


Figure 1.5: Domain architectures and variable domain structures if IgNARs and camelid heavy chain antibodies. On the left, the domain organisation of each type of antibody is shown, with CDR loops coloured red, and disulphide linkages shown in yellow. On the right, the variable domain of each type of antibody is shown, with CDR loops coloured in red. **(a)** shows a camelid heavy chain antibody (PDB accession code: 8BB7). **(b)** shows an IgNAR (PDB accession code: 2YWY). Note the lack of CDR2 in the VNAR.

1.4 Antibody structural modelling

The relative ease and low cost of obtaining the sequence of an antibody has allowed the sequencing of libraries and repertoires on the scale of several million examples. This is largely due to next-generation sequencing technologies that have emerged in recent years (Rouet et al. 2018; Marks et al. 2020; Goldstein et al. 2019), with single-cell sequencing even enabling the sequencing of paired antibodies (Rouet et al. 2018). Experimental structure determination, on the other hand, remains slow and labour-intensive by comparison. *In silico* methods have therefore emerged as a way to enable structure-aware analysis of antibodies.

1.4.1 Homology-based and *ab initio* modelling

Due to the conserved overall structure of the FW region and the minor loop canonical forms, the majority of an antibody structure can reliably be modelled using template-based approaches, also referred to as homology modelling. Indeed, homology modelling was, until recently, the dominant method for antibody structure prediction tasks (Almagro et al. 2014). In the first Antibody Modelling Assessment[3], the FW and minor CDR residues were predicted with an average root mean squared deviation (RMSD) to the target crystal structure of 1.0 Å and 1.2 Å, respectively (Almagro et al. 2011). On the other hand, due to its relative structural diversity, the CDR-H3 loop was predicted with an average RMSD of 3.1 Å (Almagro et al. 2011). In the second Antibody Modelling Assessment[4], however, participants were able to predict the FW, minor CDRs, and CDR-H3 residues with RMSDs of 0.9 Å, 1.1 Å, and 2.8 Å, respectively (Almagro et al. 2014). In these assessments, a state-of-the-art general protein modelling algorithm at the time, I-TASSER (Roy et al. 2011), performed worse than antibody-specific methods.

While homology modelling remained the preferred method for prediction of the FW regions and minor CDR loop structures, it fails to predict the CDR-H3 structure accurately due to the high number of unique conformations that this loop exhibits

[3]The Antibody Modelling Assessment is an antibody-specific blind assessment for structural modelling accuracy, inspired by the general protein Critical Assessment of Structure Prediction.

[4]The second assessment was conducted 2014

(Regep et al. 2017; Weitzner et al. 2015). As such, this loop was often modelled using *ab initio* approaches. A common workflow was to use template-based modelling approaches to model the FW and minor CDR loops, followed by *ab initio* modelling of CDR-H3 (Leem et al. 2016; Weitzner et al. 2017; Zhu et al. 2014). Variants of this approach were used by state-of-the art methods such as RosettaAntibody3 (Weitzner et al. 2017) and the structure prediction method by Zhu *et al.* (Zhu et al. 2014). A further improvement to this approach was proposed by Leem *et al.* with ABodyBuilder (Leem et al. 2016), where a template search is conducted for the FW region based on sequence homology, followed by homology-based modelling of minor loops using FREAD (Choi et al. 2010). Following this, the CDR-H3 loop is modelled in a hybrid approach with Sphinx (Marks et al. 2017b), which uses backbone orientations from known CDR-H3 structures when a sequence match is found, and models the remainder of the loop *ab initio*. Lastly, side chains are modelled into the resultant structure using PEARS (Leem et al. 2018), an antibody- and position-specific side chain modelling algorithm. The resulting decoys are then ranked, and the top decoy is selected resulting in average CDR-H3 RMSD of 2.5 Å across the second Antibody Modelling Assessment structure set.

1.4.2 Deep-learning focused approaches

With the recent dominance of deep learning methods in bioinformatics, newer antibody modelling algorithms have relied extensively on such approaches. DeepAb (Ruffolo et al. 2021), an early one of these types of methods uses a three-step procedure. First, a recurrent neural network architecture is used to learn representations of antibody sequences followed by a convolutional neural network to predict inter-residue distances and orientations. These predictions are then fed into the Rosetta protein modelling suite (Leaver-Fay et al. 2011) in order to realise the structures. This has made significant improvements to the homology-based approach described previously. It should be noted that due to the Rosetta relaxation step, the production of a model using DeepAb requires roughly two minutes, and 10-50 models are required to find an optimal structure. Following the formulation

of graph neural networks (GNN) with properties appropriate for the modelling of biomolecular structures (Satorras et al. 2021)[5], the field shifted to the use of such architectures for structure prediction. ABLooper (Abanades et al. 2021), based on the GNN architecture, models antibody CDR loops at accuracies comparable to DeepAb, but does not rely on Rosetta for structure realisation, with the entire pipeline taking approximately 10 seconds per structure. Due to the increase in available antibody sequence data, such timescales are important for the ability to model entire repertoires prior to structural analysis.

Following the 14$^{\text{th}}$ Critical Assessment of Structure Prediction (Kryshtafovych et al. 2021), AlphaFold2 (AF2) (Jumper et al. 2021) and its derivative, AlphaFold-Multimer (Evans et al. 2021), became the *de facto* method for modelling general protein structures. In part, the success of AF2 was due to its use of multiple sequence alignments (MSA), which allow extraction of residue coupling information (Hopf et al. 2014). Indeed, experiments have demonstrated that a lack of MSAs results in different predictions being made, demonstrating the importance of the residue couplings for the prediction pipeline (del Alamo et al. 2022). In conventional protein structures, residue couplings give information on which residues are likely to be interacting in the three-dimensional structure, as the protein being modelled is likely under evolutionary pressure to maintain its structure. As such, if one residue mutates, an interacting residue is more likely to undergo a compensatory mutation to preserve the geometry and chemistry of the environment. Antibodies; however, often do not follow this paradigm. While the FW region is under selective pressure to maintain its characteristic fold, the requirement to generate diverse antibodies targeting diverse epitopes necessitates selective pressure to maintain FW structures while diversifying CDR loop structures and surface chemistry. As such, antibody MSAs do not contain the same information as general protein MSAs do. To this end, Abanades *et al.* developed ABodyBuilder2 (ABB2) (Abanades et al. 2022), which relies on the same architecture as AF2 to realise structures, but removes the need for MSA generation and processing, while using a weighted

[5]These properties and formulations will be discussed further in Chapter 2.

scoring function that places more importance on the CDR residues, as modelling of the FW residues is a less challenging task. ABB2 has been shown to predict structures with comparable accuracy compared to AF2, but outputs a modelled structure considerably faster than AF2, requiring roughly five seconds in contrast to the 30 minutes required by AF2.

1.5 Antibody dynamics and conformational ensembles

Most cellular processes are orchestrated at the nanoscale through the interaction of proteins and other biomolecular structures with one another. As proteins are dynamic molecules that can undergo large-scale conformational changes, a complete picture of these processes requires an understanding of the underlying motions that drive these interactions and their timescales. Experimentally, conformational rearrangements of proteins (and to a certain extent, even information on whether multiple conformations exist) are difficult to access. Structure determination has been dominated by X-ray crystallography (XRC), wherein the structure of proteins packed into crystals is solved. Data collection during XRC is normally performed at cryogenic conditions, and the crystal environment is highly unlike the physiological environment in which a protein normally exists (Jacobson et al. 2002). The crystal environment is heavily desolvated, has an increased concentration of non-physiological additives, and copies of the protein are tightly packed against each other, resulting in so-called crystal contacts (Janin et al. 1995). XRC usually only provides a single conformation of the protein, although alternate crystallisation conditions have in some cases allowed the resolution of alternate conformations (Spindler et al. 2014; Alanine et al. 2019). Further, it is possible to observe "mixtures" of alternate side chain conformations in crystal structures. These are often small changes, but do indicate side chain flexibility. Approaches such as room temperature crystallography have confirmed that the conformations observed in cryogenic crystal conditions are likely to contain artefacts (Fraser et al. 2011). Nuclear magnetic resonance spectroscopy, on the other hand, allows multiple conformations to be

resolved, as data collection occurs in the solution phase, but it is generally limited to relatively small proteins or protein complexes (Puthenveetil et al. 2019). Lastly, cryogenic electron microscopy, which has seen increased use following a number of technological breakthroughs around 2013 (Callaway 2015), provides a more nuanced picture, as larger systems can be resolved, and local flexibility information is readily extracted from generated density maps. Its average resolution, however, is still lower than for XRC. Other methods to probe flexibility of proteins such as electron paramagnetic resonance and fluorescence resonance energy transfer exist, but are laborious and generate low-resolution data.

Biomolecular simulations such as molecular dynamics (MD) and Monte-Carlo (MC) simulations investigate conformational changes and flexibility with atomic-scale resolution *in silico*. MD allows the monitoring of the time evolution of a biomolecular system, though its main limitation is its computational expense which limits most experiments to timescales that are orders of magnitude shorter than those of functionally relevant conformational transitions (Srivastava et al. 2018). Methods exist to enhance the sampling of such simulations, such as metadynamics or replica-exchange simulations (H'enin et al. 2022). The former, used in this thesis, is detailed in Chapter 2.

1.5.1 Antibodies as conformational ensembles

Proteins exist in solution as an ensemble of multiple conformational states, with different states having differing stabilities (Henzler-Wildman et al. 2007). From a statistical mechanics perspective, the relative likelihood of a copy of a protein existing in one conformation over another is given by the relative difference between the stabilities, or free energies, of these two states (Price et al. 2011). As such, the conformational ensembles of proteins can be regarded as free energy landscapes, where unfavourable conformations appear as hills and favourable ones as wells. The "deeper" a well is (that is, the lower the free energy for a conformation is), the more this conformation will dominate the solution ensemble. Methods such as MD simulations traverse this landscape and allow estimation of both the heights of the

energetic barriers between different stable conformations, as well as the timescales at which transitions from one state to another occur (Henzler-Wildman et al. 2007).

Some antibodies are known to undergo conformational changes that are linked to their function (Jimenez et al. 2002; Alanine et al. 2019). Critically, the CDR loops have been shown to exist in multiple conformations in solution, with even the minor loops being able to occupy multiple canonical forms and cycle between them. These motions have been observed experimentally and computationally (Jimenez et al. 2002; Fernández-Quintero et al. 2020b; Fernández-Quintero et al. 2021; Guloglu et al. 2023). The CDR-H3 loop is often the most flexible loop across CDRs (Fernández-Quintero et al. 2020a; Fernández-Quintero et al. 2019a; Guloglu et al. 2023), in agreement with the structural diversity of the loop that is observed in crystal structures[6](Regep et al. 2017). Further, it has been shown that while the FW region is less flexible when compared to the CDR loops, the inter-domain orientation of the V_H and V_L domains is highly dynamic and occurs on relatively short timescales for the magnitude of the rearrangement (Fernández-Quintero et al. 2021). Since the spatial orientation of the CDR loops is heavily dependent on the orientation of the domains that carry them, this adds another mode by which conformational diversity is introduced to the binding sites of antibodies.

The ability of a single antigen to adopt multiple conformations also explains the paradox outlined in Section 1.2.2, which states that there are fewer possible naive antibodies than possible epitopes. If an antibody is regarded as a dynamic entity rather than a static one, with each conformational state effectively producing a novel binding region, a much smaller number of naive antibodies is able to target a comparatively large space of possible epitopes. The paratope state that is complementary to a given epitope is then stabilised during affinity maturation (Mishra et al. 2018), thereby leading to a highly affine and specific antibody from a lineage that has the potential to bind many possible targets.

[6]Here, we use the term structural variability to refer to the diversity of CDR loop conformations found in solved crystal structures and the term flexibility to refer to the ability and likelihood of an individual CDR loop to adopt multiple different conformations.

1.5.2 The effect of affinity maturation on antibody dynamics

A large number of studies found that as an antibody accumulates mutations during affinity maturation, it becomes less flexible (Fernández-Quintero et al. 2020c; Li et al. 2015; Ovchinnikov et al. 2018; Wong et al. 2011). When regarding an antibody as a conformational ensemble, a naive antibody effectively has the ability to adopt the correct paratope for multiple different epitopes. Once the antigen is then recognised by an antibody, a successful immune response should involve modifying the structure of this antibody such that the "unnecessary" conformations are destabilised while the binding-competent conformation is further stabilised. This makes the antibody more likely to occupy the conformation that will bind the antigen, thereby effectively amplifying binding. As such, the observation that rigidification occurs provides a structural basis for one of the effects affinity maturation.

1.6 The antibody-antigen interaction

The affinity of an antibody for its antigen, or the "tightness" of the interaction is most rigorously measured using the equilibrium dissociation constant, $\mathrm{K_D}$, which for an antibody (Ab) and antigen (Ag) interacting according to the equation

$$\mathrm{Ab} + \mathrm{Ag} \underset{\mathrm{k_{off}}}{\overset{\mathrm{k_{on}}}{\rightleftharpoons}} \mathrm{AbAg} \tag{1.1}$$

is given by the ratio of reactant concentrations to product concentrations, or by the ratio of the dissociation, $\mathrm{k_{off}}$, and binding, $\mathrm{k_{on}}$, rate constants (Price et al. 2011):

$$\mathrm{K_D} = \frac{[\mathrm{Ab}][\mathrm{Ag}]}{[\mathrm{AbAg}]} = \frac{\mathrm{k_{off}}}{\mathrm{k_{on}}} \tag{1.2}$$

More rigorously, $\mathrm{K_D}$ is derived from the binding free energy change, $\Delta\mathrm{G}$ (Price et al. 2011):

$$\mathrm{K_D} = -RT \ln \Delta\mathrm{G}, \tag{1.3}$$

where R is the universal gas constant and T is the absolute temperature. The affinity of the interaction can be measured using a number of commonly used

methods, although only a subset of these, namely surface plasmon resonance and biolayer interferometry (Schasfoort et al. 2008), provide binding and unbinding rate measurements. *In silico* methods have also been used to predict binding free energies using both physics-based (Patel et al. 2017; Wang et al. 2022) and machine learning-based estimators (Liu et al. 2020). Prediction of binding affinities is currently more reliable using physics-based approaches, as recent work has shown that a data gap of several orders of magnitude exists for deep-learning based prediction of antibody affinities (Hummer et al. 2023).

1.6.1 Antibody binding in a dynamic context

There are two commonly cited mechanisms for protein-protein interactions. The induced fit model (Koshland 1958) states that the two binding parts occupy conformational states in solution with near-complementary surfaces that upon collision with one another undergo rearrangement to form the interface. The conformational selection model (Weikl et al. 2014), on the other hand, conceptualises two proteins as conformational ensembles, stating that the proteins independently cycle through conformations and bind to one another when they collide whilst occupying their respective binding-competent conformational states.

The rate at which an antibody binds to an antigen, denoted as the k_{on} rate, is significantly influenced by how readily the antibody's favorable conformation can match the antigen in this view. A higher k_{on} rate indicates that the antibody has a conformation readily available that fits well with the antigen, resulting in a higher frequency of collisions that result in binding. On the other hand, the unbinding rate, k_{off}, is dependent on how easily the antibody-antigen complex can revert to separate entities. A lower k_{off} rate suggests a stronger interaction, where the antibody remains bound to the antigen for a longer duration. Indeed, recent work has provided a strong line of evidence for the conformational selection model in antibody-antigen interactions (Fernández-Quintero et al. 2019b; Fernández-Quintero et al. 2019a).

Furthermore, affinity maturation plays a crucial role in enhancing the efficiency of the conformational selection mechanism. As antibodies undergo affinity maturation,

they accumulate mutations that increase the stability of the conformation that best fits the antigen. This process effectively "narrows down" the conformational ensemble, reducing the prevalence of less relevant structures and amplifying the dominance of the antigen-binding conformation. Consequently, matured antibodies exhibit higher binding rates and lower dissociation rates (or some combination thereof), leading to stronger and more specific antigen-binding. As such, the conformational selection model provides a theoretical framework consistent with experimental observations on antibody specificity, flexibility, and antigen recognition (Fernández-Quintero et al. 2020c; Weikl et al. 2014).

1.7 Therapeutic antibodies

While antibodies have evolved as part of the immune system, recent advances in both technology and our understanding of immunity have enabled therapeutic applications of these molecules. Antibody therapeutics now represent a significant portion of the therapeutics market, with a share of US$115.2 billion in 2018 that is expected to rise to US$300 billion by 2025 (Lu et al. 2020). Further, the number of new antibody therapeutics in yearly approvals by the United States Food and Drug Administration is increasing, with 2022 being the first year where as many biologics were licensed as small-molecule drugs (Senior 2023), with more than half of these being antibody-based formats (Mullard 2022). The Therapeutic Structural Antibody Database (Raybould et al. 2019b), which tracks therapeutic antibodies registered with the World Health Organisation (WHO), indexes 940 clinical stage therapeutics (CSTs), with 164 having gained approval as of November 2023. The dominant therapeutic format involving antibodies is termed monoclonal antibody (mAb) therapy. In this approach, a single effective antibody that is able to neutralise some target antigen is introduced to stimulate the host immune system to attack the target. Monoclonal antibodies were first developed as a result of hybridoma technology (Milstein 1999), which allows singular clones the be produced *in vitro*

and isolated. This resulted in the first therapeutic mAb, Muronomab[7], being licensed in 1990 (Smith 1996).

1.7.1 Discovery of therapeutic antibodies

Approaches to therapeutic mAb development have been framed as occurring across three generations (Sormanni et al. 2018). Within this framework, the first generation refers to *in vivo* discovery, where model organisms (usually mice) are immunised with the target of interest, their B-cells are isolated for hybridoma production, and the antibodies they produce are screened to identify binders. This approach produces non-human antibodies, which then require humanisation to modify the sequence to make it more human-like, enabling the antibodies' therapeutic use in humans (Chin et al. 2021; Tennenhouse et al. 2023). Second-generation approaches are focused on *in vitro* discovery, where large artificial libraries are screened experimentally for target affinity (Smith 1985). The advantage of this approach is that it is faster, owing to the lack of immunisation, and libraries can be constructed using pre-humanised antibodies with known distributions of biophysical properties that allow therapeutic use. The third generation refers to *in silico* discovery, wherein an antibody that binds a given target is either identified from a library through computational predictions, or designed *ab initio* entirely. This usually involves structure-aware tools that model the candidate antibody and antigen and predict whether the antibody will bind to the target antigen through the use of statistical, physical, or machine learning (ML) scoring functions. For instance, the RosettaAntibodyDesign framework (Adolf-Bryfogle et al. 2017) relies on the Rosetta energy function (Alford et al. 2017) to score candidate complexes. This scoring function is a hybrid method combining both physics-based calculations and statistical likelihoods for various structural elements. Deep learning approaches are trained on known antibody-antigen complexes (or even general protein complexes) to "learn" the features of a physically feasible antibody-antigen complex, and can then be used to score candidate complexes. An example for this approach is DLAB (Schneider et al.

[7]This is a murine antibody targeting the CD3 receptor and has therapeutic indications mainly for the treatment of transplant organ rejection.

2021). There are many possible approaches for the *in silico* design of antibodies and the number of ways in which the problem is approached is steadily growing. For instance, a current trend within the field is co-design of antibody structure and sequence (Jin et al. 2021; Peng et al. 2023). *In silico* approaches have been shown to accelerate the development cycle for new antibody therapeutics (Mock et al. 2023), allowing faster transitions from pre-clinical stages to clinical trials.

1.7.2 Developability liabilities

Whether a therapeutic will eventually gain approval for clinical use is not only determined only by its ability to bind a target. In addition to generating the desired biological activity, it must also possess a safe pharmacological profile and must be stable; that is, the drug must be developable. In the case of antibodies, solubility, stability, immunogenicity, and polyspecificty are some of the chief properties that must be optimised alongside binding affinity (Sormanni et al. 2018).

The solubility of an antibody refers to whether it is prone to aggregation, in which case a therapeutic formulation cannot be produced, or whether it produces a viscous therapeutic formulation that affects the pharmacokinetic properties of the therapeutic (Sormanni et al. 2018). Immunogenicity refers to the potential of an administered formulation resulting in a strong immune response by the patient being mounted against the therapeutic. In mild cases this can result in a modulation of the pharmacodynamics of the drug, whereas in extreme cases this can lead to life-threatening responses in the form of cytokine storms (Vaisman-Mentesh et al. 2020; Jarasch et al. 2015). Lastly, polyspecificity[8] (Cunningham et al. 2021) involves an antibody binding multiple targets, resulting in off-target effects. As these properties (including binding affinity) are all functions of the surface topology and chemistry of the antibody, optimising these properties concurrently is a challenging

[8]The definition of polyspecificity in the academic literature is imprecise. While it sometimes refers to off-target binding to unrelated antigens in a negative context, it is sometimes used interchangeably with cross-reactivity, which refers to the ability of an antibody to target multiple variants of the same epitope, such as different serotypes of the same pathogen. Cross-reactive binding is highly desirable in many contexts involving immunity induced by vaccines or natural immunity against pathogens, and antibodies that possess this ability are referred to as cross-neutralising antibodies.

task and the properties must be considered together in order to minimise the trade-off between developability and affinity. A number of computational tools have been developed to address these issues (Lauer et al. 2012; Jain et al. 2017; Raybould et al. 2019a; Norman et al. 2019). For example, the Therapeutic Antibody Profiler (TAP) (Raybould et al. 2019a) allows for the structure-aware analysis of a candidate antibody to determine whether its surface properties deviate significantly from those of known registered CSTs.

1.8 Thesis outline

In Chapter 2, we provide the necessary theoretical background for the main approaches used in this thesis to study antibodies. This includes an account of MD simulations, including enhanced sampling by way of metadynamics, as well as a brief outline of deep learning. We conclude this chapter by describing GNNs, the specific deep learning architecture used later in this thesis.

In Chapter 3, we detail our research on the effect of structural dynamics on the biophysical properties of antibodies. We show that the binding, but not dissociation rates of antibodies are dependent on the flexibility of the CDR-H3 loop. We further show that a static picture of antibodies does not contain information on the binding kinetics, necessitating a dynamic view of antibodies for complete understanding of the binding affinity of these molecules. We assess the flexibility of antibodies using MD simulations and also show that structural modelling algorithms can be effectively exploited to gather this information at much lower computational cost. Using these flexible pictures of antibodies, we examine their surface properties and show how developability metrics can vary across the conformational ensemble, once again highlighting how a dynamic view provides richer understanding of their properties.

Following from our results establishing the importance of recognising the dynamic nature of an antibody, in Chapter 4 we show the effects of the V_L domain on CDR-H3 structure and flexibility. A long-standing hypothesis in the field based on a number of examples is that the exact conformation of the CDR-H3 loop is determined not only by its sequence, but also by the environment that the rest of the antibody

provides. We start our approach by comparing single-domain antibodies and paired antibodies, and show that CDR-H3 loops in these structures pack against the rest of the antibody in distinct ways. We then use metadynamics simulations and static crystal structures to show that the CDR-H3 loop is effectively constrained within a "steric" cage by the minor loops of an antibody, thereby limiting the possible conformations that it can adopt. Lastly, we show that specific canonical forms of V_L minor loops result in antibodies with preferences for high- or low-diversity CDR-H3 loops. The results from this chapter contribute to our understanding of how the dynamic nature of an antibody results in the vast number of possible paratopes, and gives insights into how this structural space is modulated. Further, the results of this chapter provide a structural basis for the observation that antibody structural modelling algorithms that consider the entire antibody at once, rather than its regions in isolation, are more accurate.

In Chapter 5, we focus on refining static predicted structures of antibodies, which, as discussed, are necessary for structure-aware approaches to antibody design and conformational sampling. Based on our conclusion in Chapter 4, we design a model architecture which considers the entire antibody at once, and train GNNs to predict refined inter-Cβ distances based on initial predictions by ABB2. We show that domain-specific modifications to a state-of-the-art structural refinement architecture and training using only antibodies results in improvements to predictions made by ABB2, especially in starting models of low quality. Lastly, we show that the realisation of refined structures using molecular dynamics, modified to include our predicted inter-Cβ distances, fails, despite the predicted distances being physically feasible.

Lastly, in Chapter 6, we discuss the implications of the results presented in this thesis. We argue that consideration of conformational ensembles will be essential for a full understanding of the biophysical properties of antibodies based on our work in Chapter 3. Based on the observed V_L influence on CDR-H3 structure and flexibility in Chapter 4, we hypothesise that the V_L domain may indirectly influence properties that are attributed to the CDR-H3 loop, necessitating a holistic

approach to antibody design. We raise questions about how the structural modelling accuracy of antibodies is assessed for algorithms such as the one detailed in Chapter 5, considering the results of this thesis. Finally, we propose future work to further develop our understanding of antibody structure and function.

2
Theory

Contents

2.1 Chapter abstract 27
2.2 Molecular dynamics simulations 28
2.2.1 Molecular mechanics 28
2.2.2 Molecular dynamics 31
2.2.3 Well-tempered metadynamics simulations 35
2.3 Deep learning and graph neural networks 38
2.3.1 Machine learning 38
2.3.2 Training neural networks 39
2.3.3 Graph neural networks 41

2.1 Chapter abstract

The material detailed in this chapter aims to provide the theoretical background for Chapters 3, 4, and 5. The former two chapters rely heavily on molecular dynamics simulations to probe conformational ensembles of antibodies, with experiments in Chapter 4 using metadynamics simulations to increase coverage. Here we outline the basic methodology of a molecular dynamics simulation and give an account of how metadynamics simulations are performed while allowing recovery of the equilibrium ensembles. In Chapter 5 we describe work involving graph neural networks, a type of machine learning algorithm that allows learning from data in the form

of mathematical graphs. In this chapter we outline the intuition behind machine learning algorithms and discuss the minimal steps required to train a machine learning model. Lastly, we review the basic mathematical formulation of graph neural networks and the key operation needed to train them, the message passing layer.

2.2 Molecular dynamics simulations

Cellular processes depend on and are driven by the interactions between macromolecular structures and their environments, including proteins, carbohydrates, lipids, water, and solutes. However, at the nanoscale, these interactions are orchestrated through interactions between atoms at the nuclear and electronic levels. As the systems of interest in biochemistry are often too large for these interactions to be accessible to experimental methods at a time-resolution that captures important events, they are often studied through the application of molecular dynamics (MD) simulations. These simulations involve the modelling of the system of interest and observing its evolution as a function of the atomic interactions and time.

2.2.1 Molecular mechanics

Rigorous analysis of the interactions that govern biomolecular systems would require an *ab initio* approach at the level of quantum mechanics. The computational cost associated with these calculations makes them intractable for the study of events such as protein-protein interactions which often require simulation of a system at timescales for which quantum mechanical calculations become unfeasible (Leach 2001). To this end, the Born-Oppenheimer approximation, which states that the movement of electrons is near-instantaneous when compared to the movement of atomic nuclei due to their relatively minuscule masses, is applied to effectively ignore the degrees of freedom stemming from the motion of electrons (Leach 2001). This leads to a simplified system for modelling and allows the use of molecular mechanics (MM) approaches.

In a MM approach, we describe particle motions by simple functions based on Newtonian physics. That is, molecular systems and their motions are described

using bond stretching, bending, and dihedral potentials. Such models are referred to as "force fields".

2.2.1.1 Force fields

As described, force fields are functions that can be used to estimate the potential energy of a system based on a set of parameters (Leach 2001). Their accuracy is crucial to the accuracy of any downstream applications and is an active area of research (Leach 2001; Maier et al. 2015; Ponder et al. 2010). Commonly, these parameters are fitted using a combination of experimental data as well as calculations using a quantum mechanics approach (Maier et al. 2015; Huang et al. 2016; Ponder et al. 2010). A typical force field describes the total internal energy, U, of a system as the sum of the bonded and non-bonded interactions for all n particles in the system at their atomic positions $\mathbf{r}_i$:

$$U(\mathbf{r}_1, ..., \mathbf{r}_n) = \sum U_{bonded} + \sum U_{non-bonded} \tag{2.1}$$

which in turn can be further broken down into

$$U_{bonded} = \sum U_{bond} + \sum U_{angle} + \sum U_{dihedral} \tag{2.2}$$

and

$$U_{non-bonded} = \sum U_{electrostatic} + \sum U_{vdW} \tag{2.3}$$

U_{bonded} terms are used to calculate the energetics of bond stretching, bending, and dihedral angles and commonly take on the following forms:

$$U_{bond} = \sum_{i,j} \frac{k_{r,ij}}{2}(r_{ij} - r_0)^2 \tag{2.4}$$

$$U_{angle} = \sum_{i,j,k} \frac{k_{\theta,ijk}}{2}(\theta_{ijk} - \theta_0)^2 \tag{2.5}$$

$$U_{dihedral} = \sum_{i,j,k,l} \sum_{n} k_n(1 + cos(n\phi_{ijkl} - \delta_n)) \tag{2.6}$$

Bond stretching is usually modelled using a harmonic potential (Eqn 2.4), where k_r is the spring constant for the bond between the atom pair ij and r_0 is the

reference distance between the two atoms. Similarly, angles are modelled using harmonic potentials (Eqn 2.5) for triplets of atoms ijk with a similar spring constant k_θ and a reference angle r_θ. Lastly, dihedrals are modelled for four connected atoms $ijkl$ using a periodic function (Eqn 2.6) that is dependent on the number of maxima and minima, n, and the barrier height during rotation, k_n. The term δ_n is a phase factor that places the minima and maxima.

Non-bonded interactions between pairs of atoms ij are modelled using van der Waals (vdW), U_{vdW}, and electrostatic potentials, $U_{electrostatic}$:

$$U_{vdW} = \sum_{i,j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] \tag{2.7}$$

$$U_{electrostatic} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.8}$$

vdW potentials are modelled using a Lennard-Jones potential (Eqn 2.7), where r_{ij} is the speration of the atom pair and A_{ij} and B_{ij} are force field specific parameters, while electrostatic potentials are modelled using Coulomb's law (Eqn 2.8), where q_i and q_j are partial charges of the pair of atoms and ϵ_0 is the electric constant.

It should be noted here that more accurate functions to describe each of the components exist, and that several simplifying assumptions are built into force fields (Leach 2001). For example, bond stretching may be more accurately described using Morse potentials[1] (Leach 2001) and electrostatic interactions do not usually capture the effects of polarisation (Ponder et al. 2010). The aim with many force fields is to balance accuracy with ease of calculation, though modern force fields provide good approximations of first principle quatum mechanical calculations. The choice of force field is therefore a critical consideration for the user, as the force field must "resolve" interactions to an extent that allows the events of interest to be captured accurately[2] while still allowing the computations to be fast enough that such an event can be observed[3]. In this thesis, the Amber ff14SB force field was

[1]This approximation captures bond breaking events at long distances, but is more costly to compute.

[2]For example, modelling of enzymes often makes use of hybrid quantum mechanics/molecular mechanics force fields to allow observations on bond breaking and formation.

[3]For example, protein-lipid interactions in membrane environments are often studied using coarse-grained force fields, as make longer time scales accessible.

used, which models atoms explicitly and has been shown to produce errors <4.18 kJ mol^{-1} when compared to quantum mechanical calculations (Maier et al. 2015)

2.2.1.2 Energy minimisation

When a proteinaceous system is prepared for simulation, the atomic positions are modelled and the system is often soaked into a solution of water molecules and salt. The initial placement of these additional molecules, as well as any errors in the starting structure such as clashes can result in unfavourable geometries, and cause U to become very large, resulting in an unstable simulation. To prevent this, the geometry of the system is usually optimised beforehand with an energy minimisation procedure to find a local optimum. Typically, this involves a numerical procedure, such as the gradient descent algorithm. OpenMM (Eastman et al. 2016), the MD engine used throughout this thesis, relies on the L-BFGS algorithm (Liu et al. 1989).

2.2.2 Molecular dynamics

In order to follow the time evolution of the system given initial atomic coordinates using MD, we essentially make use of Newton's equations of motion,

$$\mathbf{F}_i = m_i \mathbf{a}_i \tag{2.9}$$

where $\mathbf{F}_i$ is the force acting on particle i, m_i its mass, and $\mathbf{a}_i$ its acceleration. Given that the force field allows us to quantify the forces acting on all n atoms in the system, as

$$\mathbf{F}_i(\mathbf{r}_1, ..., \mathbf{r}_n) = -\nabla_i U(\mathbf{r}_1, ..., \mathbf{r}_n) \tag{2.10}$$

we can formulate the following equation that allows us to calculate the positions of the atoms at time t:

$$-\nabla_{r_i} U(\mathbf{r}_i) = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} \tag{2.11}$$

$$\iint -\nabla_{r_i} U(\mathbf{r}_i) dt dt = m_i \mathbf{r}_i(t) \tag{2.12}$$

$$\mathbf{r}_i(t) = \frac{1}{m_i} \iint -\nabla_{r_i} U(\mathbf{r}_i) dt dt \tag{2.13}$$

which can be integrated numerically over short time steps. The need for short time steps arises from the fact that the procedure assumes that the potential function and, therefore, the forces remain stable throughout the time step. Too large a time step would therefore result in inaccuracies of the calculated positions $\mathbf{r}_i(t + \Delta t)$. In extreme cases, this would result in the simulation "exploding" (Braun et al. 2019). Short time steps are therefore chosen to avoid errors during the simulation. The choice of time step depends on the system and the simulation conditions, as well as algorithmic nuances, such as the algorithm used for the integration and whether certain degrees of freedom are removed from the system. Generally, biomolecular simulations are performed using a step size of 1-2 fs (Leach 2001; Braun et al. 2019), though recent work has shown that a step size of 4 fs is achievable by repartitioning mass from heavier atoms onto hydrogens in the system (Hopkins et al. 2015).

Lastly, before a simulation is performed, initial atomic coordinates and velocities are assigned to atoms in the system. The former are usually obtained from experimentally solved structures, whereas the latter are sampled from a Boltzman distribution.

2.2.2.1 Integration algorithms

As outlined, performing an MD simulation requires taking short steps in the time dimension, followed by a recalculation of forces, velocities, and positions before taking another step. The method by which this integration is implemented is crucial to ensuring that errors do not accumulate in the system along the simulation, as these will propagate through the simulation. To perform the integration accurately, a number of integration algorithms have been formulated.

A commonly used integration algorithm is the leapfrog integration algorithm (van Gunsteren et al. 1988). This algorithm calculates positions at $t + \Delta t$ using accelerations, $\mathbf{a}$, and positions, $\mathbf{r}$, from t and velocities, $\mathbf{v}$, from $t - \frac{\Delta t}{2}$:

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \mathbf{a}(t)\Delta t \tag{2.14}$$

$$\mathbf{r}\left(t + \Delta t\right) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{2.15}$$

The algorithm calculates positions at full time steps, but velocities at times staggered by half of the set time step, effectively allowing calculations to "leapfrog" over each other, and has been shown to increase stability.

In this thesis, we performed simulations using a modified Langevin integrator (Izaguirre et al. 2001; Leimkuhler et al. 2016). This applies the Langevin equation when integrating using a leapfrog method:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{r_i} U(\mathbf{r}_i) - \gamma m_i \frac{d\mathbf{r}_i}{dt} + \mathbf{R}_i(t) \tag{2.16}$$

where γ is the friction coefficient and $\mathbf{R}_i(t)$ is a stochastic force representing thermal noise. The first term of this equation is equation 2.9 and two additions are made. The second term is a friction term which is proportional to the velocity of the particle, slowing it down, whereas the last term effectively represents random collisions that particles in the system experience with a hypothetical surrounding water bath at a given temperature. The overall effect of the equation is to ensure that velocities of the particles are consistent with a desired temperature, effectively coupling the simulation to an external heat bath.

2.2.2.2 Temperature and pressure coupling

In the study of biomolecular systems using MD simulations, it is often desirable to study the NPT thermodynamic ensemble, which has a constant number of particles (N), constant pressure (P), and constant tmeperature (T), as this replicates physiological conditions best. The first requirement is trivial, but maintaining temperature and pressure requires modifications to be made to the simulation in the form of a thermostat and a barostat. While multiple thermostats such as the Berendsen (Berendsen et al. 1984), Noose-Hoover (Evans et al. 1985) and Andersen (Andersen 1980) exist, in this thesis Langevin coupling is used. Following from equation 2.16, the velocities can be rescaled at each time step t.

The pressure, P, of a system during an MD simulation is inversely related to its volume, V. It therefore follows that pressure can be increased by decreasing the size of the bounding box and *vice versa*. This is generally done by rescaling

the atomic coordinates of the system. As with thermostats, a number of barostat algorithms exist, such as the Berendsen (Berendsen et al. 1984) and Perinello-Rahman (Parrinello et al. 1981) barostats. In simulations conducted for this thesis, we use the Monte Carlo Barostat, proposed by Åqvist *et al.* (Åqvist et al. 2004). The procedure involves the proposition of a trial volume change. Whether the proposed change in volume is accepted depends on the change in the energy of the system. A favourable volume change will be accepted more frequently while still maintaining stochastic behaviour using the Monte Carlo Method.

2.2.2.3 Periodic boundary conditions

The computational complexity of MD simulations scales heavily with the number of particles simulated, and a goal is to reduce the size of the system as much as possible. Since biological macromolecules usually exist in solution, enough water molecules are still required to (i) fully solvate the solute and (ii) provide enough distance between copies of it to prevent artefacts due to overcrowding. A single copy of the protein is usually simulated in a box of water and salt[4]. However, the small size of the box results in water molecules at the edges of the simulation box experiencing a different environment compared to those further in, resulting in surface effects. To prevent this, periodic boundary conditions are employed. That is, if a particle crosses the boundary of the box, it simply enters the box from the opposite direction. This means that we are effectively simulating an infinite lattice of simulation boxes that are completely identical, thereby preventing surface artefacts. A side-effect of this is that simulation boxes must be of a shape that can form this lattice without any gaps. While cubes are often used, rhombic dodecahedra or truncated octahedra provide an advantage as they require fewer water moelcules to be added. A diagram demonstrating periodic boundary conditions in two-dimensions is shown in Figure 2.1

[4]Additions such as membranes, metabolites, or ligands are also often used.

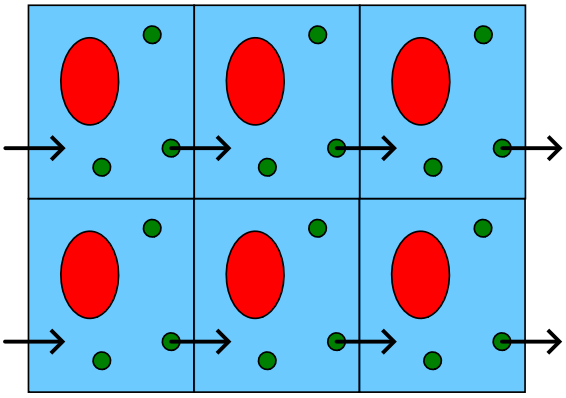


Figure 2.1: Diagram of periodic boundary conditions. By allowing a particle exiting the box to enter it from the other side, the simulated system effectively becomes an infinite lattice, forgoing any issues arising from surface exposure.

2.2.3 Well-tempered metadynamics simulations

Using typical modern hardware, most MD simulations are limited to sampling on the scale of hundreds of ns to tens of µs, though specialised hardware has been used to simulate proteins at the ms scale (Lindorff-Larsen et al. 2011), with further technological advances in hardware like the ANTON3 supercomputer allowing simulation of ribosomes at speeds of 93 $\mu s\,d^{-1}$ (Shaw et al. 2021). However, events that are often studied by MD simulations such as protein folding, loop motions, or allostery can occur on timescales far longer than is accessible through classical MD, with estimates ranging from the scales of ns to seconds. In order to observe such rare events, a number of enhanced-sampling MD protocols have been developed (H'enin et al. 2022). Broadly, these have been categorised into adaptive bias simulations, where some external force is applied to the simulation to manipulate the dynamics of the system; generalised ensemble methods, where exchanges of configurations between parallel simulations are performed; and adaptive seeding methods, where starting configurations of the system are biased (H'enin et al. 2022). In this thesis, we have used well-tempered metadynamics simulations (Barducci et al. 2008), a method that falls within the adaptive bias category.

Metadynamics simulations were originally developed by Laio and Parrinello (Laio et al. 2002) and involve the definition of a collective variable (CV) that describes some slow degree of freedom of the system that is of interest. As the

simulation progresses, time- and configuration-dependent external biases are added to the simulation, thereby "forcing" the simulation to escape local minima.

2.2.3.1 Procedure

Assuming an unbiased MD simulation, we start with a system that evolves according to the total internal energy as a function of its configuration, $U(\mathbf{r})$. In metadynamics simulations, we aim to progressively add biases in the form of repulsive Gaussians to the simulation, such that configurations associated with local minima become less stable the longer the simulation spends time in these.

To do this, we first identify a set of d CVs, $S(\mathbf{r}) = s_1(\mathbf{r}), ..., s_d(\mathbf{r})$ which describe the configuration of the system along the degree of freedom we would like to sample. In practice, choosing these CVs is not a trivial task, and often requires knowledge gained from unbiased simulations, although several techniques exist to accelerate the discovery of effective collective variables based on machine learning methods (Sultan et al. 2017; Bonati et al. 2021). Once CVs have been identified, the width, σ_i, of each bias s_i and the hill height, W, are determined[5]. Once the simulation is started, biases are added to the simulation at regular time intervals, τ, at configuration $\mathbf{r}(k\tau)$. Therefore, the biasing potential experienced by the system at configuration $\mathbf{r}$ and time t can be expressed as

$$V(S(\mathbf{r}), t) = \sum_{k\tau<t} W(k\tau) \exp\left(-\sum_{i}^{d} \frac{(s_i(\mathbf{r}) - s_i(\mathbf{r}(k\tau)))^2}{2\sigma_i^2}\right) \tag{2.17}$$

and the system now evolves according to a modified potential $U'(\mathbf{r}, t) = U(\mathbf{r}) + V(S(\mathbf{r}), t)$. A flaw with these simulations is that biases do not converge to the free energy, but oscillate around it. A variant of the method called well-tempered metadynamics solves this problem by adding an additional term to the calculation of the biases, referred to as the bias factor, ΔT. In this variant, the hill height, W, in Equation 2.17 is scaled according to

[5]Generally, the height of biases is taken to be large enough to effectively accelerate sampling but not so large that diffusion along the CV is lost. The widths of the biases similarly need to be chosen such that they are not so narrow that they effectively modify only a minuscule region of the CV space, but not so large that they result in the simulation being "pushed over" regions that have not yet been explored.

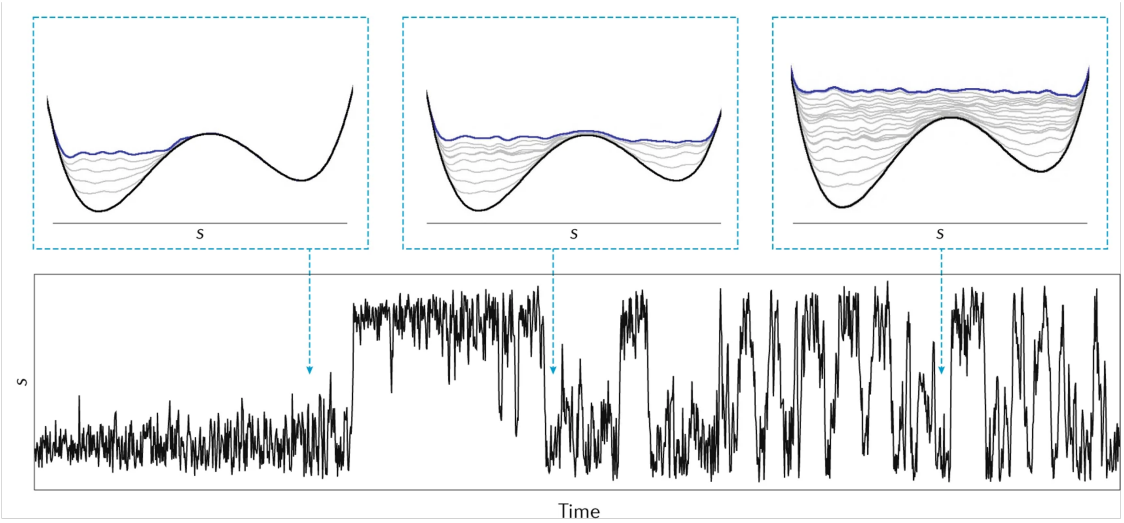


Figure 2.2: Diagram of well-tempered metadynamics simulations. On the bottom, the CV S is plotted against time, showing how more complete sampling is achieved as the simulation progresses. At the top, the modified free energy surface ($U'(\mathbf{r}, t) = U(\mathbf{r}) + V(S(\mathbf{r}), t)$) is shown, with the time-dependent bias potential preferentially filling free energy minima. Note that biases added later during the simulation are smaller, allowing for convergence. Adapted from Bussi and Laio (Bussi et al. 2020).

$$W' = W \exp\left(-\frac{V(s(\mathbf{r}(k\tau)), k\tau)}{\Delta T}\right). \tag{2.18}$$

This effectively means that, as the simulation progresses, newer biases become smaller and smaller, eventually resulting in a completely flat free energy surface along which CVs diffuse (Dama et al. 2014). Indeed, as the simulation time tends to infinity, the change in the bias potential tends to 0.

2.2.3.2 Recovery of equilibrium ensembles

Since the free energy surface of metadynamics simulations is modified, ensemble averages over all observed frames of the simulation trajectory do not represent the unbiased underlying distribution. In an unbiased simulation, the statistical weight for each configuration, $w(\mathbf{r}, t)$, is the same, whereas in an enhanced sampling scheme such as metadynamics this is not the case. In order to obtain unbiased ensembles, it is therefore necessary to estimate the biased weights, $w_b(\mathbf{r}, t)$. If these are accurately identified, the unbiased time average of some observable A, $\langle A \rangle_0$, can be recovered:

$$\langle A \rangle_0 = \frac{\langle Aw \rangle_b}{\langle w \rangle_b} \tag{2.19}$$

Since the biases deposited during the simulation are complementary to the unbiased free energy surface (Fig 2.2), these biases can be used to estimate the required statistical weights. In this thesis, the reweighting scheme proposed by Branduardi *et al.* has been used (Branduardi et al. 2012):

$$w(s) \propto \exp\left(\frac{V(s)}{k_B T}\right) \tag{2.20}$$

2.3 Deep learning and graph neural networks

2.3.1 Machine learning

Machine learning (ML) is a subfield of artificial intelligence that aims to develop algorithms that have the ability to learn patterns from data and can then be used to make predictions and decisions. While pioneering work in the ML field was conducted in the 1950s, the last decade has seen extensive growth of the field and applications to new domains, including drug discovery (Dara et al. 2021). This has primarily been driven by advances in hardware that allow for faster calculations (Paszke et al. 2019), allowing more data to be processed by more complex algorithms; the conception of more expressive algorithms and architectures that are able to capture more nuanced patterns in the input data (Greener et al. 2021; Dara et al. 2021); and a rapid increase in the amount of available data.

At its core, an ML algorithm attempts to map a given input set, $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\}$, to output labels, $Y = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N)\}$, using learned parameters, $\Theta = \{\theta_1, \theta_2, ..., \theta_M)\}$. Thus, any ML algorithm can be expressed as a function, f, such that

$$f : X; \Theta \rightarrow Y$$

Deep learning (LeCun et al. 2015), wherein artificial neural networks are constructed using "layers" of learnable functions[6], has been instrumental in translating the outlined changes in capabilities into practice. The premise of such networks is that early layers extract simple features from input and later layers progressively extract higher-level patterns, thereby allowing complex functions to be learned. This is confirmed by work that has shown that deeper networks do indeed demonstrate higher levels of expressiveness, since they are able to capture more nuanced relationships (Raghu et al. 2016).

Recently, generative artificial intelligence ("AI will transform science — now researchers must tame it" 2023) has allowed for development of revolutionary models across a variety of fields such as image, text, and speech generation, as well as within biomedical sciences (Lin et al. 2022).

2.3.2 Training neural networks

When training a neural network[7], the aim is to learn the parameters, Θ, in the model in order to accurately map the input, X, to the output, Y, making the training process an optimisation problem. Once an architecture has been decided on (that is, once the necessary layers have been constructed in the desired sequence), the parameters for all layers are initialised randomly. The input is then passed through the network using the current parameters, and the accuracy of the resultant mapping (the predictions), $\widehat{Y}$, is computed. This involves the use of a loss function, which measures the distance between the prediction and the true label for each training example, $\mathcal{L}$. (Goodfellow et al. 2016; LeCun et al. 2015) For regression tasks, where the prediction is some continuous value, the mean squared error is often used, whereas classification tasks commonly use the cross-entropy loss function (Goodfellow et al. 2016) (binary cross-entropy loss is used in the case of binary classification):

[6]These are often referred to as neurons. In the most simple form, a neuron is a linear function that takes the form $\mathbf{Y} = \mathbf{\Theta X} + \mathbf{b}$, but different types of layers such as logistic regressions, normalisation layers, and activation layers to add non-linearity are also possible.

[7]Here, we focus only on supervised learning, wherein the aim is to predict some value or label. Another common type of problem is unsupervised learning, where we simply aim to learn patterns within the data, without the use of labelled data.

$$\mathcal{L} = -\sum_{i=1}^{N} \mathbf{y}_i \log(\hat{\mathbf{y}}_i) \tag{2.21}$$

for N training examples. The work in this thesis is concerned with ordinal regression, in which a continuous variable is divided into several bins and is then treated as a classification task (Winship et al. 1984). The aim then is to minimise $\mathcal{L}$ by adjusting Θ.

Once the loss across the entire training set is computed (a single pass over the training examples is referred to as an epoch[8]), the gradient of $\mathcal{L}$ with respect to each parameter $\theta_i, \theta_i \in \Theta$ is calculated, and the parameters are adjusted using some optimisation algorithm, in a process referred to as backpropagation (Goodfellow et al. 2016). While gradient descent optimisation and derivatives thereof can be used, more sophisticated algorithms such as the Adam optimiser (Kingma et al. 2014) are often preferred due to their ability to prevent over-fitting to the training data. Over-fitting refers to the phenomenon wherein the parameters are adjusted to match the input examples perfectly at the cost of generalisation. An over-fitted model has therefore "memorised" the training data, but is unable to make accurate predictions on new inputs. Once optimised after an epoch, the accuracy of the model in making predictions on unseen data[9] (the validation set) is measured to assess learning. Over-fitting can then be assessed by comparing the loss across the training set to the loss across the valdiation set. Another method to prevent over-fitting is the addition of a dropout layer, which sets random inputs/outputs of a specified layer to 0 during each forward pass through the network, thereby preventing over-reliance on a single feature in the data (Srivastava et al. 2014).

While the volume and diversity of the input data is likely the most important factor in the majority of neural network training tasks, several decisions can heavily influence the accuracy of the final model. These are referred to as hyperparameters, and include, but are not limited to, model architecture, the rate at which parameters

[8]In reality, not all training examples are operated on at once due to limitations of computer memory. Instead, the training set is divided into batches during an epoch.

[9]This data is excluded from the training examples, and therefore the model is unable to "cheat" by over-fitting on these examples to artificially inflate its performance.

are adjusted (the learning rate), batch sizes, training epochs, and the choice of loss function. It is common to scan across a variety of combinations of hyperparameters by training multiple models in order to identify the optimal combination of these. Identification of the best set of these hyperparameters is often performed by assessing the accuracy of the model on a second unseen set, the test set (Goodfellow et al. 2016).

2.3.3 Graph neural networks

A common issue arising during the use of artificial neural networks is that their static nature requires inputs to be of a defined shape and size. This has proven to be especially challenging in bioinformatics and cheminformatics applications, as examples from the input data distribution often have varying dimensions (Greener et al. 2021). For instance, molecules may have differing numbers of atoms with different connectivities, and proteins may have sequences containing differing numbers of chains and residues. Further, tasks using molecular structural data are also often expected to be translation- and rotation-equivariant. For instance, rotating or translating a protein structure in three-dimensional space should not affect predictions for whether the protein binds some ligand. While certain approaches, such as convolutional neural networks (LeCun et al. 2015), have been successful in circumventing the input size issue by placing the molecule in question in an n-dimensional grid (Schneider et al. 2021; Misiura et al. 2021), they have largely been unsuccessful in circumventing the challenge of translation and rotation equivariance. Based on their ability to capture molecular structural data effectively, graph neural networks (GNNs) (Scarselli et al. 2009; Greener et al. 2021) and derivative models have in recent years largely replaced the use of other architectures in deep learning for tasks involving molecular structures.

A graph, $G(V, E)$, is a collection of nodes or vertices, $V = \{v_1, v_2, ..., v_n\}$, and edges, $E = \{e_1, e_2, ..., e_m\}$, which connect nodes to one another. Edges may be bidirectional or unidirectional, with the former resulting in an undirected and the latter in a directed graph. For work in this thesis, only the case of undirected

graphs is relevant. Graphs may be fully connected, such that each node shares an edge with every other node; they may be sparsely connected, such that the number of total edges is smaller than that of the fully connected graph constructed using the same nodes; and they may also contain self-edges, where a node is joined to itself via an edge. When considering two nodes, termed v_i and v_j, the edge connecting them is referred to as e_{ij}. For any node v_i, the neighbourhood of that node, $\mathcal{N}(v_i)$, refers to all other nodes that v_i is joined to:

$$\mathcal{N}(v_i) = \{v_j \in V | e_{ij} \in E\}.$$

In GNNs, nodes are associated with node features, with $\mathbf{h}_i$ representing the feature vector of any node $v_i, v_i \in V$. Edges may also be assigned features, denoted here using $\mathbf{e}_{ij}$ for any edge $e_{ij} \in E$. When modelling a protein structure as a graph, the graph can be constructed either at the atom- or residue- level (Jamasb et al. 2020), with both approaches having been successful for a variety of tasks (e.g. Abanades et al. 2022; Chinery et al. 2022). In this thesis, residue-level graphs were constructed, wherein a node is constructed for each residue. Further, prediction tasks for graphs can be at the node level, where nodes are classified or some scalar value is predicted for each node (for example, prediction of whether an antibody residue is in the paratope of the antibody); the edge level, where either some value or class is predicted for an edge (for example, prediction of the type of intermolecular contact between two residues), or the presence of an edge is predicted (for example, whether two atoms are likely to interact); or the graph level, where the entire graph is classified (for example, prediction of enzyme specificity).

The key operation in a GNN is the message-passing layer (MPL) (Fig 2.3) (Scarselli et al. 2009). Intuitively, this type of layer allows information from one node to be passed along the edge to a joining node. At each message passing layer, all nodes receive messages from every node to which they are connected. These messages are then aggregated per node and added to the existing node features. The same can also be done for edge features, where edges receive messages from the nodes that they join. As such, after l message passing layers have been applied, each node will contain

information from nodes that are l degrees away, allowing long-range information to be passed along, even if nodes are not explicitly connected to each other.

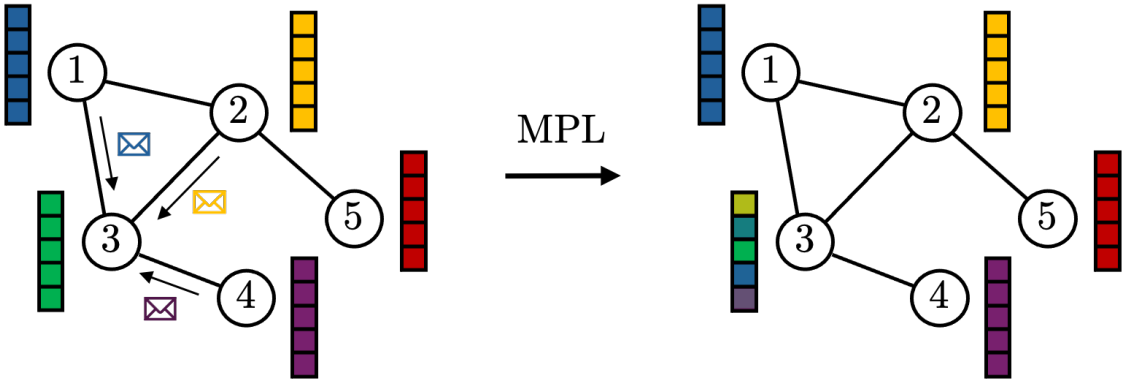


Figure 2.3: Diagram of message passing layers in graph neural networks. At each layer, messages to a target node v_i from all nodes connected to v_i are calculated using some learnable function ϕ. The messages are then aggregated and added to the existing node features of v_i, allowing integration of information from $\mathcal{N}(v_i)$. **NB:** For clarity, the diagram shows only a single node being updated. In practice, all nodes would be updated concurrently at each layer.

A message passing layer consists of three basic steps. Messages are calculated and aggregated and nodes are updated. If edge features are present and also being updated, the new edge features at layer l commonly become the message between two nodes:

$$\mathbf{e}_{ij}^{(l)} = \mathbf{m}_{ij}^{(l)} = \phi_{\mathrm{e}}^{(l)}\left(\mathbf{h}_i^{(l-1)}, \mathbf{h}_j^{(l-1)}, \mathbf{e}_{ij}^{(l-1)}\right) \tag{2.22}$$

where $\mathbf{e}_{ij}^{(l)}$ and $\mathbf{e}_{ij}^{(l-1)}$ represent edge features for the edge connecting nodes v_i and v_j at the current and the previous layers, respectively, and $\phi_e^{(l)}$ is some learnable function. $\mathbf{h}_i^{(l-1)}$ represents the features of node v_i from the previous layer. $\mathbf{m}_{ij}^{(l)}$ represents the message being sent between the two nodes. Following message calculation, all messages coming to a node v_i are aggregated:

$$\mathbf{m}_i^{(l)} = \bigoplus_{j \in \mathcal{N}(v_i)} \left(\mathbf{m}_{ij}^{(l)}\right) \tag{2.23}$$

where $\bigoplus_{j \in \mathcal{N}(v_i)}$ is some aggregation function such as the calculation of the mean, sum, or maximum, and $\mathbf{m}_i^{(l)}$ is the message incoming to node v_i. The aggregation function is usually permutation-invariant, although methods exist to

allow permutation-sensitive functions to be used (Murphy et al. 2018). In this thesis, we use a weighted mean aggregation. The per-node messages are then used to update the features of each node:

$$\mathbf{h}_i^{(l)} = \mathbf{h}_i^{(l-1)} + \phi_{\mathrm{n}}^{(l)}\left(\mathbf{h}_i^{(l-1)}, \mathbf{m}_i^{(l)}\right) \tag{2.24}$$

where $\phi_{\mathrm{n}}^{(l)}$ is some learnable function.

MPLs therefore effectively map the graph onto an increasingly abstract representation of itself. Finally, a neural network such as a multilayer perceptron (Greener et al. 2021) can be used to make some prediction on every node or edge, depending on the task[10].

When a graph onto which a molecular structure is mapped is supplied with *xyz* coordinates of the nodes as node features, the graph loses its translation- and rotation-equivariant character. An equivariant GNN usually uses relative distances between node positions, as these satisfy the equivariance condition. In this work, we have applied this approach when constructing graphs. Alternate formulations of GNNs, such as E(n) Equivariant Graph Neural Networks (EGNN) (Satorras et al. 2021), which have been popularised for molecular graph-based ML approaches, also use this approach to ensure equivariance under rotation and translation operations.

[10]If a graph-level prediction is required, an additional step is usually added to reduce the graph (or some subset of it) into a single feature vector.

3

Contributions of conformational ensembles to biophysical properties relevant to vaccinology and therapeutic antibody design

Contents

3.1 Chapter abstract 46
3.2 Introduction 47
3.3 Methods 51
3.3.1 Prediction of antibody binding rates 51
3.3.2 Analysis of therapeutic surface profiles 58
3.4 Results 60
3.4.1 A diverse set of antibody structures shows that antibody binding and dissociation rates are independent of each other 60
3.4.2 Physicochemical descriptors of static structures are poor predictors of binding rates in antibodies 61
3.4.3 CDR-H3 length predicts binding rate constants of antibodies 66
3.4.4 CDR-H3 flexibility correlates with binding rates and can be probed using computationally inexpensive methods . 68
3.4.5 Sphinx efficiently identifies differences in CDR-H3 flexibility 72
3.4.6 Secondary structure and surface flexibility of antigens influence antibody binding rates 76
3.4.7 Limited exploration of antibody conformational ensembles enriches profiling of therapeutic properties 77
3.5 Discussion and conclusions 85

3.1 Chapter abstract

Antibodies have proven potential as highly selective therapeutics, yet the underlying biophysics of the antibody-antigen interaction is not well understood. In particular, many questions remain about how antibodies can be designed to show selectively improved biophysical properties. An example which motivated this thesis is the finding that antibodies with high binding rates were more effective at preventing erythrocyte invasion by the malaria parasite *Plasmodium falciparum*. This may also hold true for other intracellular pathogens, or diseases in which the antigen is transiently exposed. In this chapter, we aim to determine which antibody properties cause them to show a higher binding rate. We first collated a data set of antibody-antigen complexes with binding rate information and compared their properties. We found that information from static antibody-antigen complexes did not allow prediction of their binding rates. We next explored whether the flexibility of the antibody impacted binding rates. We used both molecular dynamics simulations and antibody modelling algorithms to examine the flexibility of the third hypervariable loop of the antibody heavy chain and found that this flexibility is a determinant of binding rates, but not of dissociation rates. We also show that flexibility of the targeted epitope influences binding rates. Given the importance of flexibility for binding rates, we also investigated how flexibility influences the surface properties of antibodies. We found that analysis of these can be enriched by considerations of conformational ensembles, enabling a deeper understanding of possible liabilities that may arise during therapeutic antibody discovery and design. Our results raise considerations for immunogen and antibody design and demonstrate that an understanding and consideration of conformational ensembles enriches our understanding of biophysical properties.

This chapter contains reproduced material from the following publication:

Raybould, M.I.J., Turnbull, O.M., Suter A., **Guloglu B.**, Deane, C.M. (2023) Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling. BioRXIV. doi.org/10.1101/2023.06.28.546839 (Raybould et al. 2023)

Declaration: Results presented in Section 3.4.7 are part of a project jointly conducted with Dr. Matthew I.J. Raybould. All simulations were performed and analysed for convergence and trajectories were converted into PDB files by the author of this thesis. Dr. Raybould conducted TAP calculations for resultant frames. All data presented in this thesis were visualised and interpreted by the author of this thesis.

3.2 Introduction

Therapeutic antibodies have demonstrated great potential in treatments of a number of diseases, with rapidly increasing market shares, as discussed in Chapter 1. However, prediction and optimisation of biophysical properties such as binding affinity, kinetics, and development liabilities are still challenging for antibodies. Technical advancements in both algorithm design and experimental data generation are currently fuelling improvements (Hummer et al. 2022). Tools such as the Therapeutic Antibody Profiler (TAP) (Raybould et al. 2019a) are regularly updated with all new clinical-stage therapeutic (CST) antibodies, while other work has shown that there still exists a data-gap of several orders of magnitude to confidently predict the binding affinity of antibodies to their cognate antigens (Hummer et al. 2023) (see Chapter 1).

Proteins do not exist as single conformers in solution (Henzler-Wildman et al. 2007), as described in Chapter 1, where we also discussed studies demonstrating that antibody CDR loops can undergo large-scale conformational changes in solution, with CDR-H3 being able to occupy many different states (Fernández-Quintero et al. 2020b; Fernández-Quintero et al. 2019b; Fernández-Quintero et al. 2020a; Guloglu et al. 2023). In line with this, recent work has shown that the conformational

selection model (Weikl et al. 2014), in which the binding rate is a function of how likely the binding partners are to be in the correct conformation at the time of any intermolecular collisions, provides an explanation of antibody binding (Blackler et al. 2022; Fernández-Quintero et al. 2019a). Further, it has been shown that affinity maturation results in a decrease in flexibility and rigidification of the binding pose (Fernández-Quintero et al. 2020c; Wong et al. 2011; Palma et al. 2017; Schmidt et al. 2012), although it is contested whether this happens in all cases of affinity maturation (Jeliazkov et al. 2017). In keeping with the conformational selection model, we hypothesise that the flexibility of the CDRs may be a unique determinant of binding rates of antibodies, as a more rigid antibody that is locked into the binding-competent pose would undergo a higher proportion of productive collisions with the antigen.

Traditionally, the property most frequently used to demonstrate strong binding of an antibody to its target antigen is the equilibrium dissociation constant (K_D), which in turn can be expressed as the ratio of the dissociation rate constant (k_{off}) and the association or binding rate constant (k_{on}). However, tight binding in the form of a low K_D value is not always associated with strong neutralising activity for a given pathogen, as the antibody might not be targeting a key epitope. Further, recent work has shown that, for some pathogens, k_{on} is a more powerful predictor of neutralising activity than K_D, such as in the case of the malarial antigen PfRH5 (Alanine et al. 2019) or in Simian Immunodeficiency Virus variants (Steckbeck et al. 2005). In the former case, it has been shown that a high binding rate is preferable to overall strong binding, since any circulating antibodies only have a short period before the pathogen invades host cells and becomes inaccessible to them. Few efforts have been made to investigate this phenomenon in different pathogens, yet it is plausible that a similar effect might be observed in other cases where the pathogen is only transiently available to antibodies such as Human Immunodeficiency Virus or herpesviruses. We therefore asked what causes an antibody to have a rapid binding rate.

Studies of binding rates for different protein-protein and protein-ligand interaction systems have indicated that different biophysical properties are associated with rapid binding. Electrostatic attractive forces may play a role (Chiu et al. 2015; Pan

et al. 2013; Bai et al. 2011), in a manner similar to electrostatic steering observed in enzyme-substrate interactions (Vascon et al. 2020). In the case of small-molecule inhibitors binding the HIV protease, it has been shown that electrostatic interactions increase the binding rate (Chiu et al. 2015), although charge complementarity does not appear to predictably increase binding rates in every case (Pan et al. 2013), as has been shown for phosphate binding to periplasmic phosphate-binding protein (Ledvina et al. 1998) or a number of carbonic anhydrase inhibitors (Krishnamurthy et al. 2008). A study involving a data set of 62 protein-protein complexes spanning a variety of folds including enzymes and antibody-antigen complexes has also found a correlation between electrostatic complementarity and binding rates among other determinants, such as hydrophobicity and the buried surface area (Bai et al. 2011). Hydrophobicity and the ease of desolvation of the binding site have also been indicated as predictors of binding rates of small molecules binding heat shock protein 90 using molecular dynamics simulations (Schmidtke et al. 2011). Beyond its chemistry, the topology and curvature of the binding surface has also been found to modulate binding rates in general protein-protein interactions, through modulation of the entropy change of association (Bai et al. 2011). Yet more studies have formulated the problem in terms of transition state theory (Amaral et al. 2017; Schoop et al. 2015; Moal et al. 2012), arguing that the free energy barrier of necessary conformational changes to adopt the binding-competent pose determines binding rates in the case of small molecules binding pockets on protein surfaces, such as heat shock protein 90 inhibitors (Amaral et al. 2017) and other pharmaceutical targets (Schoop et al. 2015). This has also been shown in a set of 44 protein-protein complexes, 10 of which were antibody-antigen complexes (Moal et al. 2012). To our knowledge, no systematic analysis of these possible determinants has been conducted specifically in the case of antibodies, and the question of which factors influence binding rates remains an open one.

During therapeutic antibody discovery, binding is not the only consideration; developability must also be optimised (see Chapter 1). Developability encompasses properties such as self-association of an antibody, which can result in high viscosity,

impacting the pharmacokinetics of the therapeutic, and aggregation, which can result in instability of the formulation and in the worst case can lead to anti-drug immune responses (Jarasch et al. 2015). These properties are usually a result of the surface properties of the antibody. For example, aggregation often occurs due to the presence of hydrophobic patches on the surface of the antibody and antibodies with such patches are often removed from pools of candidate antibodies (Dobson et al. 2016). Typically, this relies on structural modelling of the sequences, followed by analysis of the static models. This approach does not take into account alternate conformations that may be observed in solution or how they might contribute to the surface hydrophobicity of the ensemble.

In this chapter we first examine what causes an antibody to bind more rapidly to its cognate antigen. We begin by curating a data set of antibody-antigen complexes where data on the kinetics of the binding event are available. We then analyse the static complexes, examining the interface and interactions between binding partners. Then, we turn our attention to the flexibility of the antibody and use both molecular dynamics and recent deep learning methods to estimate CDR flexibility. Our results indicate that CDR-H3 flexibility is a stronger predictor of binding rates. Further, we show that the secondary structure of the epitope also influences the binding rate of antibodies, likely through differences in flexibility.

In the latter sections of the results, we show that assessment of developability can be enriched by considering the conformational ensemble of an antibody, rather than a single static structure. Our work demonstrates that calculations using tools for structural analysis of antibodies, such as TAP (Raybould et al. 2019a), can simply be performed over MD simulation outputs to get a better understanding of how different conformational states of an antibody contribute to observed biophysical properties.

3.3 Methods

3.3.1 Prediction of antibody binding rates

3.3.1.1 Kinetic data set curation

We downloaded the 2019 release of the PDBbind-CN database (Liu et al. 2015) and all antibody structures in SAbDab (Dunbar et al. 2014) as of 26$^{\text{th}}$April, 2020. We filtered the SAbDab data set to select only structures bound to proteins (amino acid chains longer than 50 residues) and cross-referenced this with the PDBBind-CN data set to obtain all protein-bound antibody structures with known affinity. In order to maximise the size of our data set, we did not include a resolution cutoff. However, we removed all structures with missing residues in any of the CDR loops.

We manually inspected the primary literature for each entry cited in the PDB and selected structures where kinetic data meeting specific criteria were available. The criteria for this selection were that the kinetic data had been generated using surface plasmon resonance or bio-layer interferometry; that the antibody had been used as the ligand and antigen as the analyte; and that there was good agreement between sensorgrams and fitted models, resulting in a final data set of 191 structures (167 paired antibodies [Ab], 24 nanobodies/single-domain antibodies [sdAbs]). We also allowed experiments where either the antigen or a Fab fragment was used as the ligand, but omitted structures where affinity experiments used the whole antibody as the analyte, as this would result in multivalent binding events, therefore modulating the binding kinetics.

The structures in the data set were further processed through removal of crystallographic factors (defined as non-polypeptide small molecules other than water) and waters. In structures where more than one copy of the antibody-antigen complex was present, we kept only the first copy. The binding rate constants associated with each structure are given in Appendix A.1.

We used ANARCI (Dunbar et al. 2015) to number sequences using the IMGT numbering scheme (Lefranc et al. 2003) (CDR1: IMGT residues 27-38, CDR2: IMGT residues 56-65, CDR3: IMGT residues 105-117, Framework: IMGT residues

1-128 and not within CDR definitions). Where comparisons to Sphinx predictions were necessary, we instead used the Chothia numbering scheme definitions for CDR-H3 only. This corresponds to IMGT residues 108-116, resulting in a slightly shorter definition of the loop.

3.3.1.2 Static analyses of epitope and paratope structures

To examine the effects of interface composition and size on binding rates we examined the composition of the interface by analysing both residue burial and interactions between the antibody and antigen.

Buried surface area calculations We examined residue burial using an in-house implementation of the Shake-Rupley solvent-accessible surface area (SASA) algorithm (Shrake et al. 1973) using a probe size of 1.4 Å. When stratifying by residue types, we used the groupings listed in Table 3.1 (Lide 1991). We also defined two additional groups. All hydrophobic residues are those that are either in the aliphatic or aromatic hydrophobic groupings, whereas all charged residues are residues that are either in the acidic or basic groupings.

Side Chain Group	Residues
Aliphatic Hydrophobic	Ala, Ile, Leu, Met, Val
Aromatic Hydrophobic	Phe, Trp, Tyr
Polar Neutral	Asn, Gln, Cys, Ser, Thr
Acidic	Asp, Glu
Basic	Arg, His, Lys
Unique	Pro, Gly

Table 3.1: Residue groupings used for SASA calculations

Intermolecular contact analysis We used Arpeggio (Jubb et al. 2017) to analyse the intermolecular interactions in antibody-antigen complexes. Arpeggio uses a set of rules defined for CREDO (Schreyer et al. 2009) to systematically count the occurrence of different types of interactions. CREDO is a database detailing

protein-ligand interactions and includes a definition by which different pairs of atoms and local environments are classed into different types of interactions, thereby allowing a systematically consistent approach to enumerating such contacts in a protein structure. A list of interactions analysed by Arpeggio is found at `https://github.com/harryjubb/arpeggio`, and flags used are found in Appendix A.2.

Epitope secondary structure analysis We began analysis of epitope secondary structures by generating a non-redundant data set based on antigen sequences. To do this, we extracted antigen sequences from the structures using Biopython (Cock et al. 2009; Hamelryck et al. 2003), and generated a non-redundant set by using cd-hit (Li et al. 2006) at a clustering threshold of 0.9 (90% sequence identity across antigen sequence), retaining each cluster representative identified by cd-hit. This is defined as the longest sequence in that cluster. If there were multiple sequences longer than all others, the first sequence in the input FASTA file was chosen. This resulted in 59 of 167 antibody structures being removed from the data set. To calculate the dominant secondary structure element in each epitope, we used the DSSP (Kabsch et al. 1983) implementation in Biopython (Cock et al. 2009). We defined three classes by grouping secondary structure elements: helical regions (α-helices, π-helices, or 3_{10}-helices), strand regions (β-sheets and β-bulges), and coiled regions (high curvature, turns, and residues which are otherwise unclassified). We then calculated the fraction of each secondary structure element in each epitope (defined as all antigen residues that have a <4.5 Å heavy atom distance to any heavy atom in the antibody):

$$\text{fnat}_x = \frac{N_x}{N} \tag{3.1}$$

where N_x refers to the number of epitope residues classed into secondary structure class x, and N refers to the total number of residues in the epitope. We then classified each epitope by whether it is dominated by defined secondary structure ($\text{fnat}_{\text{helix}} + \text{fnat}_{\text{strand}} > 0.5$) or by coils ($\text{fnat}_{\text{coil}} > 0.5$). For further analysis, we also stratified epitopes by separating helices ($\text{fnat}_{\text{helix}} > \text{fnat}_{\text{strand}} + \text{fnat}_{\text{coil}}$), strands

($\text{fnat}_{\text{strand}} > \text{fnat}_{\text{helix}} + \text{fnat}_{\text{coil}}$), and coils ($\text{fnat}_{\text{coil}} > \text{fnat}_{\text{strand}} + \text{fnat}_{\text{helix}}$) into distinct classes rather than structured *versus* unstructured epitopes.

Analysis of epitopes using AlphaFold2 predictions We assessed the predicted error of AlphaFold2 (AF2) (Jumper et al. 2021) on the antigens in our data set by extracting the Uniprot ID (Bateman et al. 2022) of each antigen from the PDB (Berman et al. 2000), excluding fusion proteins and epitope grafts, and downloaded corresponding entries from the AlphaFold database (Váradi et al. 2021). We then extracted and calculated the average predicted error of the epitope (defined as antigen residues that have a <4.5 Å heavy atom distance to any heavy atom in the antibody).

3.3.1.3 Prediction of flexibility using molecular dynamics simulations

We performed unbiased molecular dynamics simulations of structures in the data set for which no missing residues were found in the entire F_V region, resulting in 144 systems being simulated. All systems were prepared and simulations performed using OpenMM v7.7 (Eastman et al. 2016). Antigens and constant domains in the PDB files were removed and N-methyl groups were used to cap C-termini using an in-house script. Next, using pdbfixer (Eastman et al. 2016), we protonated the models at a pH of 7.5 (protonation states were assigned using the Modeller module of OpenMM (Eastman et al. 2016)), soaked them in truncated octahedral water boxes with a padding distance of 1 nm, and added sodium or chloride counterions to neutralise charges and then NaCl to an ionic strength of 150 nM. We parameterised the systems using the Amber14-SB forcefield (Maier et al. 2015) and modelled water molecules using the TIP3P-FB model (Jorgensen et al. 1983). Non-bonded interactions were calculated using the particle mesh Ewald method (Darden et al. 1993) using a cut-off of distance of 0.9 nm, with an error tolerance of 5×10^{-4}. Water molecules and heavy atom-hydrogen bonds were rigidified using the SETTLE (Miyamoto et al. 1992) and SHAKE (Ryckaert et al. 1977) algorithms, respectively. We used hydrogen mass repartitioning (Hopkins et al. 2015) to allow for 4 fs time steps. Simulations were run using the mixed-precision

CUDA platform in OpenMM using the Middle Langevin Integrator with a friction coefficient of 1 ps^{-1} and the Monte-Carlo Barostat set to 1 atm. We equilibrated systems using a multi-step protocol detailed in Table 3.2. Following equilibration, we performed 100 ns of unrestrained simulation of the NPT ensemble at 300 K, estimating flexibility over the final 90 ns of each simulation.

Stage	Ensemble	Restrained Atoms	Restraint Strength (kJ mol^{-1} nm^{-2})	Duration	Start T (K)	End T (K)
1	Minimisation	Protein Heavy	4184.00	5000 steps	-	-
2	NVT	Protein Heavy	4184.00	0.2 ns	100	300
3	NPT	Protein Heavy	4184.00	0.2 ns	300	300
4	NPT	Protein Heavy	2092.00	0.5 ns	300	300
5	Minimisation	Backbone Heavy	2092.00	5000 steps	-	-
6	NPT	Backbone Heavy	2092.00	0.2 ns	300	300
7	NPT	Backbone Heavy	418.40	0.2 ns	300	300
8	NPT	Backbone Heavy	41.84	0.2 ns	300	300
9	NPT	-	-	1 ns	300	300

Table 3.2: Parameters and steps used during the equilibration of simulated systems

To calculate Cα root mean squared fluctuation (RMSF) values of residues in

regions of interest, we aligned the region of interest on its starting structure using mdtraj v1.9.6 (McGibbon et al. 2014) and calculated the time average of the root mean squared deviation (RMSD) with respect to the average coordinates of each atom in this region over the course of the simulation, similarly to how this calculation is implemented in mdanalysis (Gowers et al. 2016):

$$\mathrm{RMSF}(\mathbf{x}) = \left\langle \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \langle \mathbf{x}_i \rangle)^2} \right\rangle \tag{3.2}$$

where $\mathbf{x}_i$ refers to the coordinates of atom i, N refers to the total number of atoms in the region of interest, and $\langle \cdots \rangle$ refers to the time average of some quantity.

3.3.1.4 Conformational ensemble analysis using Sphinx and MODELLER

Modelling using Sphinx In order to examine the flexibility of CDR-H3 using Sphinx (Marks et al. 2017b), we remodelled the CDRs onto the crystal structure framework using Sphinx after removing the antigen from each structure. Sphinx is a loop modelling method that combines both *ab initio* and knowledge-based approaches. Given a target CDR-H3, the algorithm searches for shorter loops within a CDR-H3 loop library (CDR boundaries are defined using the Chothia numbering scheme (Chothia et al. 1987)) based on sequence similarity with the target loop. Next, each candidate loop is assessed based on whether its termini fit the anchors of CDR-H3 in the given framwork. The top hits based on these criteria are then ranked according to sequence identity, and the algorithm builds the loop backbone using each selected fragment. The backbone is built by using the dihedral angles of the template in positions where a sequence match is found, and the remaining dihedral angles are selected from position-specific distributions of dihedral angles. Loops are then closed using the cyclic coordinate descent algorithm (Canutescu et al. 2003). Lastly, the built loops are scored to choose the best 500 decoys. These are then ranked after side chains are modelled using PEARS (Leem et al. 2018).

It has previously been demonstrated that Sphinx is able to differentiate between rigid and flexible loops (Marks et al. 2018). Each decoy that is generated by

Sphinx represents a possible conformation of the loop based on both templates extracted from crystal structures as well as the position-specific dihedral angle distributions that are used to complete the models. We used Sphinx to remodel and generate decoys of the CDRs (flags used available in Appendix A.3) for all of the structures in our set.

Modelling using MODELLER To model the CDR-H3 loops using MODELLER (Webb et al. 2016), we used its *ab initio* loop modelling procedure. We modelled CDR-H3 without the antigen present using the same Chothia definitions used for Sphinx and generated 500 decoys for each structure using the SOAP-loop scoring function (Dong et al. 2013). When modelling loops, we used the `very slow` refinement setting.

Analysis of ensembles from Sphinx and MODELLER In order to enable thresholding for analysis based on how much of an antibody-antigen interaction is mediated through the V_H domain CDR loops, we calculated the fraction of antibody heavy atoms within 7.5 Å of any antigen heavy atom:

$$\text{fnat}_{\text{VH}} = \frac{N_{\text{VH}}}{N} \tag{3.3}$$

where N_{VH} and N stand for the number of heavy atoms from the V_H and the total number of antibody heavy atoms within 7.5 Å of antigen heavy atoms, respectively. 7.5 Å was empirically chosen as the distance cut-off after scanning across a range of 2.5-15 Å and identifying the distance providing the strongest signal.

For MODELLER and Sphinx decoys, we separately calculated the pairwise RMSD of Cα positions of all decoys belonging to each CDR loop using Biopython, resulting in distributions with 124,750 RMSD values. For decoys a and b with N Cα positions $\mathbf{x}^a$ and $\mathbf{x}^b$, respectively, the RMSD is calculated as follows:

$$\text{RMSD}(\mathbf{x}^a, \mathbf{x}^b) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\mathbf{x}_i^a - \mathbf{x}_i^b|^2} \tag{3.4}$$

The spread of these gives an estimate of how varied the possible conformations of the CDR loops are, as predicted by each method.

Joint clustering of Sphinx decoys and molecular dynamics frames To probe the relative diversity of CDR-H3 structures generated by Sphinx and MD simulations, we pooled all 500 Sphinx decoys and 500 snapshots of the MD simulation taken at regular intervals (for our trajectories of 90 ns, this results in a frame every 0.18 ns) for each antibody structure after aligning them on the V_H domain of the starting structure. We then extracted CDR-H3 Cα coordinates and clustered them using average-linkage hierarchical clustering using SciPy (Virtanen et al. 2019) with a cut-off distance of 1.5 Å. After assigning a cluster to each conformation, we calculated the fraction of all clusters identified by both Sphinx and MD simulations.

ABodyBuilder2 and AbLooper To build loops using AbLooper (Abanades et al. 2021), we provided the crystal structure of the antibody without the antigen and used default settings. We then saved the resultant models and extracted the error estimates for the CDR-H3 residues from the B-factor column of the generated PDB files. We did not rebuild any missing density in the antibody before doing this.

To model whole antibodies using ABodyBuilder2 (ABB2) (Abanades et al. 2022), we extracted the sequences of each antibody in our data set and modelled these using default settings. We then saved the resultant models and extracted the ABB2 error estimates for the CDR-H3 residues.

3.3.2 Analysis of therapeutic surface profiles

3.3.2.1 Therapeutic antibody data set curation

In order to examine whether conformational ensembles provided by relatively short molecular dynamics simulations give an enriched picture of the surface properties of therapeutic antibodies, we modelled the F_V regions of 14 clinical-stage therapeutics (CSTs) using ABB2. We chose half of these CSTs whose structures would have been deposited in the PDB before the training of ABB2 and seven for which structural information would not have been available. This acts as an internal control and

ensures that results from the analysis are not skewed by artificially inflated model quality. We then aligned the modelled F_V structures with their crystal structures, and grafted the C_H-1 and C_L domains onto the F_V models, thereby generating Fab fragments for simulation. All missing residues in the constant domains were modelled using MODELLER v10.2, using the `very slow` refinement setting and selecting the lowest energy model from 10 decoys. A full list of CSTs and the crystal structures from which Fab fragment models were derived is provided in Table A.2.

3.3.2.2 Generation of conformational ensembles using molecular dynamics simulations

All systems were prepared and simulations performed using OpenMM v7.7 (Eastman et al. 2016). N-methyl groups were used to cap C-termini using an in-house script. Next, using pdbfixer (Eastman et al. 2016), we protonated the models at a pH of 7.5 (protonation states were assigned using the Modeller module of OpenMM (Eastman et al. 2016)), soaked them in truncated octahedral water boxes with a padding distance of 1 nm, and added sodium or chloride counter-ions to neutralise charges and then NaCl to an ionic strength of 150 nM. We parameterised the systems using the Amber14-SB forcefield (Maier et al. 2015) and modelled water molecules using the TIP3P-FB model (Jorgensen et al. 1983). Non-bonded interactions were calculated using the particle mesh Ewald method (Darden et al. 1993) using a cut-off of distance of 0.9 nm, with an error tolerance of 5×10^{-4}. Water molecules and heavy atom-hydrogen bonds were rigidified using the SETTLE (Miyamoto et al. 1992) and SHAKE (Ryckaert et al. 1977) algorithms, respectively. We used hydrogen mass repartitioning (Hopkins et al. 2015) to allow for 4 fs time steps. Simulations were run using the mixed-precision CUDA platform in OpenMM using the Middle Langevin Integrator with a friction coefficient of 1 ps^{-1} and the Monte-Carlo Barostat set to 1 atm. We equilibrated systems using the multi-step protocol detailed in Table 3.2 and performed unrestrained simulations of the NPT ensemble at 300 K for a duration of 200 ns.

3.3.2.3 Profiling of surface properties

To calculate surface properties, we assessed convergence of the simulations by monitoring F_V Cα RMSD values relative to the input structure across the simulations using mdtraj (McGibbon et al. 2014). Once converged, we extracted frames from the simulations at equal intervals and calculated therapeutic surface properties using the Therapeutic Antibody Profiler (TAP) (Raybould et al. 2019a). While TAP calculates a variety of of metrics describing the surface of the structure, we focus here only on the ones that are conformation-dependent:

- Patches of Surface Hydrophobicity (PSH) score: scores an F_V structure based on the extent to which hydrophobic residues are clustered together in patches.
- Patches of Positive Charge (PPC) score: scores an F_V structure based on the extent to which positively charged residues are clustered together in patches.
- Patches of Negative Charge (PNC) score: scores an F_V structure based on the extent to which negatively charged residues are clustered together in patches.
- Structural F_V Charge Symmetry Parameter (SFvCSP) score: scores an F_V structure based on the extent to which the overall surface charges of the heavy- and light-chains differ from one another.

3.4 Results

3.4.1 A diverse set of antibody structures shows that antibody binding and dissociation rates are independent of each other

To study the structural determinants of antibody binding rates, we first collected a data set of antibody structures bound to their antigens. Briefly, we cross-referenced the PDBBind-CN 2019 (Liu et al. 2015) set with SAbDab (Dunbar et al. 2014), and extracted entries whose PDB codes appear in both databases. This yielded a total of 191 structures of both antibodies and nanobodies/single-domain antibodies (sdAb). The binding rate constants in this set cover a wide range of values (4.6×10^3

- 2.80×10^8 $M^{-1} s^{-1}$) (Fig 3.1). We began our analyses by probing the correlation between binding rates and dissociation rates and found these to be uncorrelated ($\rho = -0.12$, $p = 0.080$), suggesting that these are influenced by different factors (Fig 3.1a). We also examined the difference between the antibody and single-domain antibody structures in our set and found no difference in binding rates (Fig 3.1b).

To assess coverage, we also examined a range of genetic attributes (Fig 3.2). Our data set contains both antibodies and single domain antibodies (Fig 3.2a) and the antibodies in this set include both κ and λ light chains (Fig 3.2b). The antibodies come from a range of organisms, with the majority arising from human and mouse lineages (Fig A.1). Within these, both the human (Fig 3.2c) and the mouse (Fig 3.2d) structures show a distribution of germline lineages, with the former including both κ and λ chains. In the mouse set, all light chains come from κ lineages (background distributions for these genes across SAbDab are shown in Fig A.2). We also examined the distribution of CDR lengths in the data set and found that the length distributions match those found in surveys across repertoire sequences and solved antibody structures (Briney et al. 2018; Raybould et al. 2019a; Wong et al. 2019; Gordon et al. 2023) in all loops (Fig A.3). Especially in the case of CDR-H3, we found that the data set mirrored the natural distribution, with a range from 5 to 26 residues ($\mu = 13.95$ residues, $\sigma = 4.31$). We refer to this data set as the kinetic data set. We conclude that, although binding and dissociation rates both influence the overall affinity of an antibody for its cognate antigen, they are likely to be dependent on distinct factors.

3.4.2 Physicochemical descriptors of static structures are poor predictors of binding rates in antibodies

Based on the lack of correlation between binding and dissociation rate constants, we hypothesised that there might be properties that selectively modulate binding rates. We started by analysing the physicochemical descriptors of the paratope to understand whether the types of intermolecular contacts in the antibody-antigen interface are predictive of binding rates. We calculated the buried surface area of

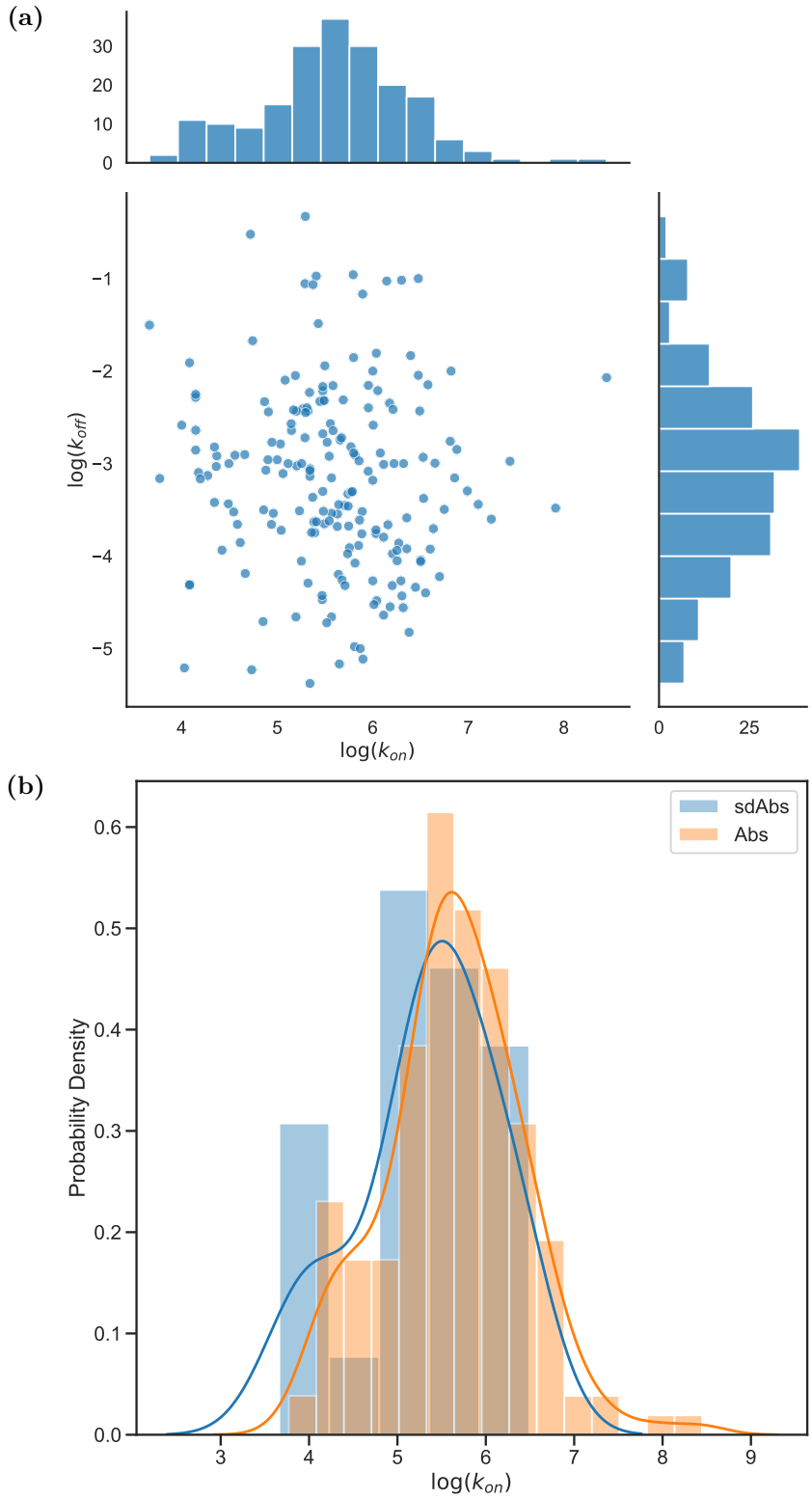


Figure 3.1: Coverage of kinetic data in the kinetic data set. **(a)** shows the marginal and bivariate distribution of binding and dissociation rates of antibodies across the data set while **(b)** shows the distribution of binding rates comparing antibodies (Abs) with nanobodies/single-domain antibodies (sdAbs)

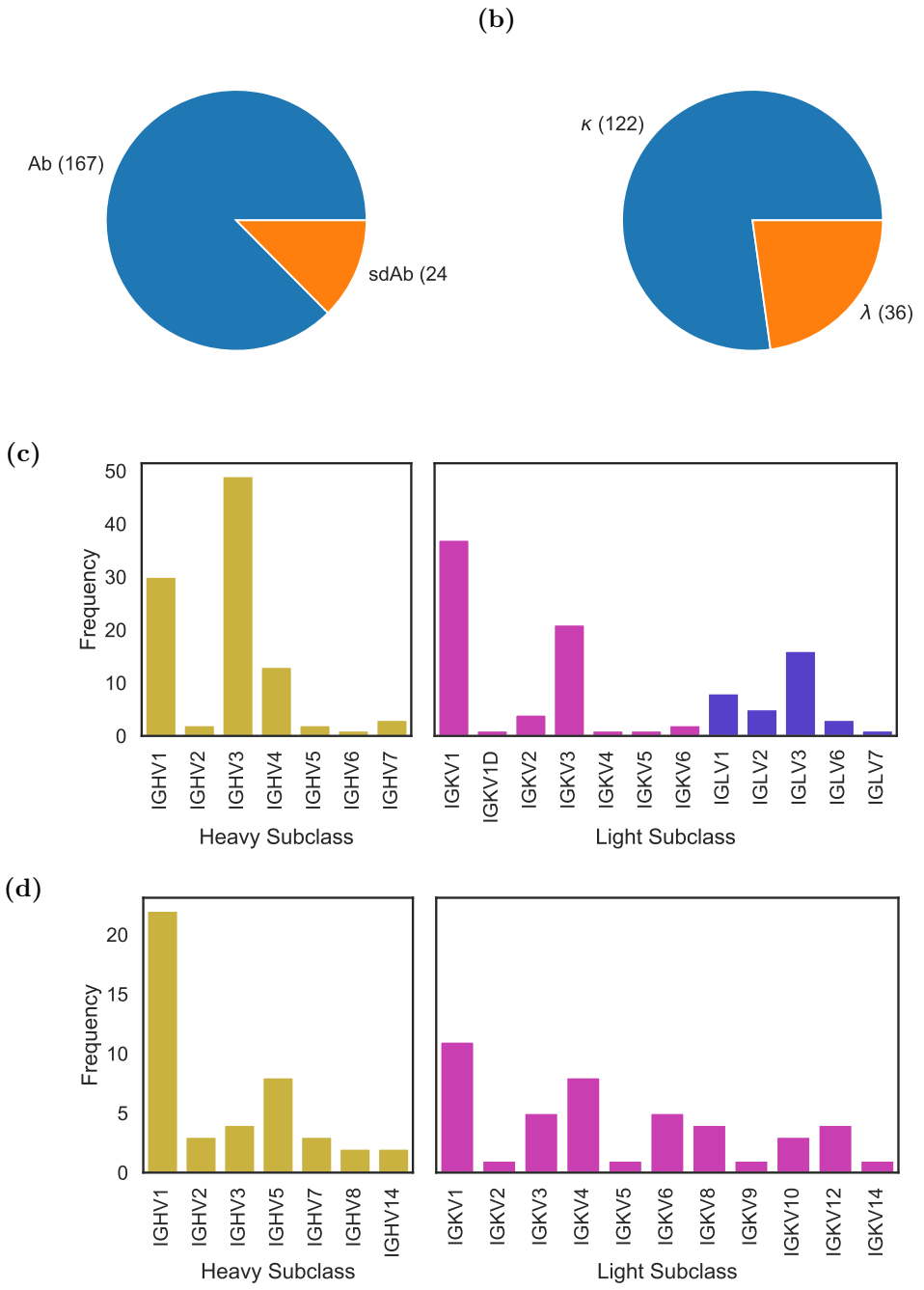


Figure 3.2: Coverage of genetic attributes in the kinetic data set. **(a)** shows numbers of antibodies (Ab) and nanobodies/single-domain antibodies (sdAb), **(b)** the distribution of κ and λ antibodies, **(c)** the germline genes for human antibodies in the data set, as listed in SAbDab, and **(d)** the germline genes for mouse antibodies in the data set, as listed in SAbDab. *NB:* Panels **(b-d)** do not include structures for which genetic attributes could not be identified by SAbDab and represent 158 examples from the total set of 167 paired antibody structures. All attributes were successfully identified for these 158 structures.

the paratope of each antibody-antigen complex and stratified this by the type of residue. We found no statistically significant correlations between the size and/or composition of the paratope and binding rates (Table 3.3).

Side Chain Group	Spearman's ρ	p-value
Aliphatic Hydrophobic	-0.14	0.08
Aromatic Hydrophobic	0.03	0.73
All Hydrophobic	-0.06	0.46
Polar Neutral	-0.12	0.14
Acidic	-0.04	0.64
Basic	-0.02	0.52
All Charged	-0.05	0.52
Unique	-0.03	0.73
All Residues	-0.15	0.05

Table 3.3: Correlations between binding rates and buried surface area (BSA). BSAs for different residue groupings (as described in the Section 3.3) on the paratope in each complex were calculated using an in-house implementation of the Shrake-Rupley algorithm, and correlations were assessed using Spearman rank correlation coefficients.

We next tested whether intermolecular contacts in antibody-antigen complexes influence binding rates. We counted and classified the contacts in the kinetic data set using Arpeggio (Jubb et al. 2017). Arpeggio is a computational tool that relies on rules defined in the CREDO database (Schreyer et al. 2009) to consistently classify intermolecular contacts. We calculated the interface fingerprint for each antibody-antigen complex by counting the number of times each type of interaction is observed between antibody and antigen within a complex, and calculated Pearson correlation coefficients (r) for each type of interaction and the experimental binding rates. We chose Pearson's r in this case as we were not able to identify any non-linear dependence structures in scatter plots of interaction frequencies and $\log(k_{on})$(Fig 3.3). We were only able to identify a weak correlation between ionic interactions and binding rates (Pearsons's $r = 0.17$) (Fig 3.3). We argue that the effect that ionic interactions have on binding rates likely occurs because electrostatic interactions

operate over a much greater distance compared to other types of interactions. This suggests that electrostatic steering may influence binding rates (Vascon et al. 2020) but is unlikely to be the major determinant. We note that a correlation was also identified for carbonyl interactions, although we argue that this is unlikely to be an important factor, since these polar contacts are relatively weak and are rare, with no more than three identified in any interface (Fig 3.3).

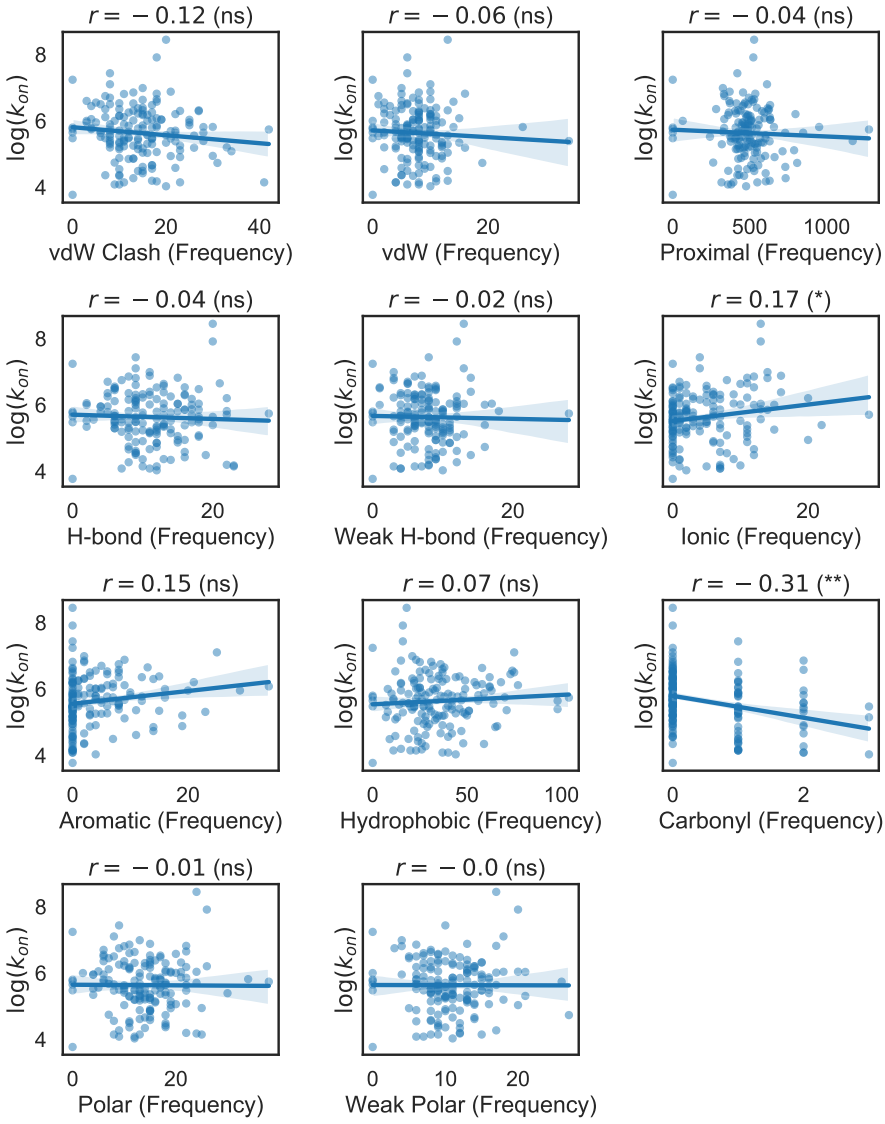


Figure 3.3: Correlations between binding rates and contacts made between antigens and antibodies. We enumerated different types of antibody-antigen interactions in the kinetic data set using Arpeggio and calculated correlations between these and binding rates. ns: not significant, *: $0.05 > p \geq 0.005$, **: $p < 0.005$

Based on our results, we conclude that static structures alone contain very little

signal that can be used to predict the binding rate of an antibody.

3.4.3 CDR-H3 length predicts binding rate constants of antibodies

Given that neither the size nor the composition of the interface correlates with the binding rate, we examined whether the flexibility of CDR loops relates to binding rate following previous studies investigating the effects of affinity maturation on antibody flexibility. We began by examining the effect of CDR length on binding rates. We hypothesized that a longer loop, when regarded as a simple polymer, is likely to be more flexible due to the increased number of degrees of conformational freedom and that this might result in a decrease in binding rates, given previous studies on the conformational selection model in antigen recognition (Blackler et al. 2022; Fernández-Quintero et al. 2019a). We examined the correlation of three metrics (CDR-H3 length alone, the total length of all V_H CDRs and the total length of all CDRs) with $\log(k_{on})$ at various $fnat_{VH}$ cutoffs as described in section 3.3 (Fig 3.4). $fnat_{VH}$ can vary from 0 to 1 and indicates how much of the paratope is contributed by the V_H domain. For example, at an $fnat_{VH}$ cutoff of 0.4, we removed all structures where $fnat_{VH} < 0.4$ from the calculation. This means that at a cutoff of 0.4 only structures where at least 40% of interactions are mediated by CDR-H1-3 remain.

We found that CDR-H3 length alone correlates with $\log(k_{on})$ (Fig 3.4a). Further, we observed that the correlation becomes stronger when the involvement of CDR-H1-3 in antigen binding is larger. The highest correlation observed is at $fnat_{VH} = 0.5$, where $\rho \approx 0.33$, after which the reliability of the correlations drops off sharply.

We observed the same trends in both correlations and associated p-values for the total length of CDR-H1-3 as we had for CDR-H3 length alone, although the magnitude of the correlations is slightly lower while the associated p-values are slightly higher (Fig 3.4b). We hypothesize that this is due to the low variance in CDR-H1 and CDR-H2 lengths. As such, the variance in the total V_H metric originates almost entirely from the variance of CDR-H3 (Fig A.3). Next, we

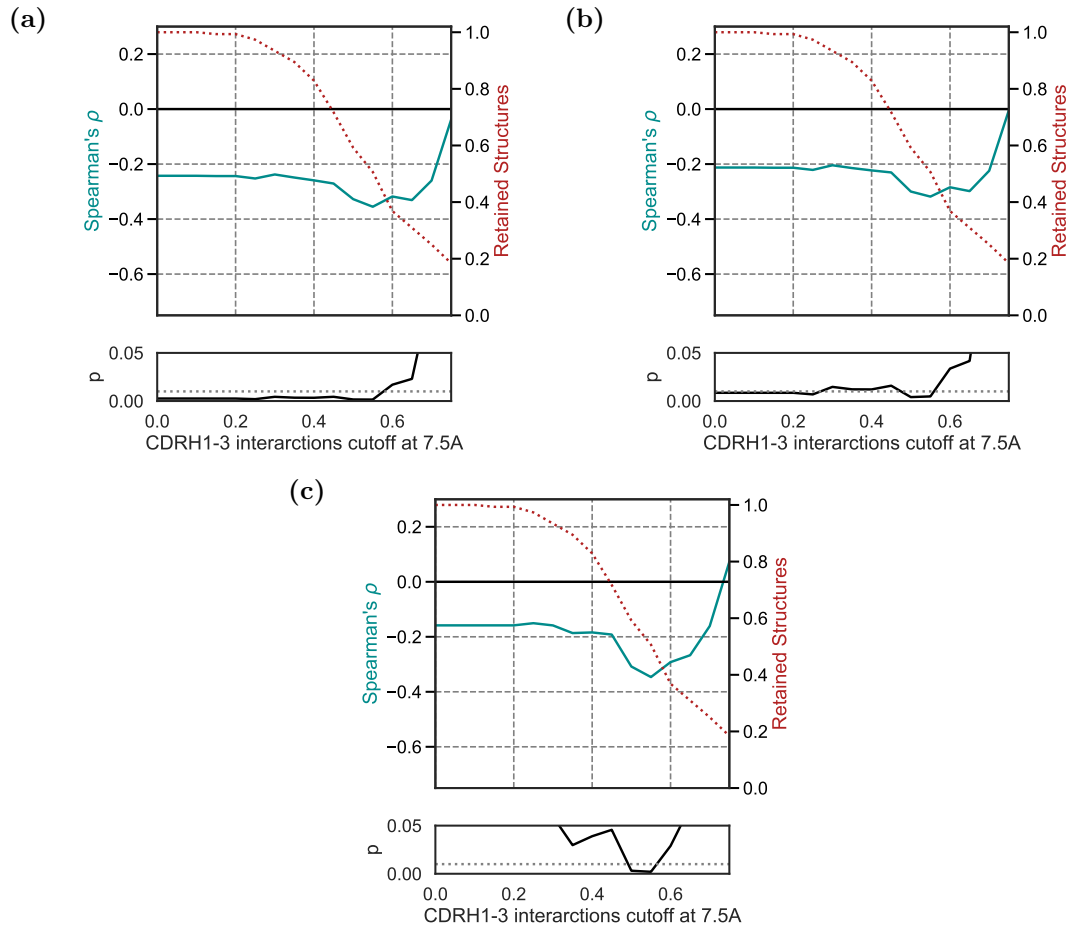


Figure 3.4: Correlations between CDR loop lengths and binding rates. At each fnat_{VH} cutoff (CDR-H1-3 interactions cutoff at 7.5 Å), we filtered the kinetic data set and correlated loop lengths and $\log(k_{on})$. The fraction of structures from the original data set remaining at each cutoff is shown in dark red while the Spearman rank correlation coefficient (ρ) is shown in cyan for each cutoff. Below each graph, p-values for the correlation at that cutoff are reported (dotted grey line: 0.01). **(a)** shows the correlations calculated using CDR-H3 length, **(b)** for CDR-H1-3 total length, and **(c)** for the total length of all CDR loops.

examined the relationship between total CDR length and $\log(k_{on})$ (Fig 3.4c). We generally find weaker correlations when considering all six CDR loops, and associated p-values indicate a loss of significance. This may be due to the low variance found in V_L CDRs (Fig A.3). We again hypothesize that the reason that correlations between total CDR length and $\log(k_{on})$ mirror the trend seen when using CDR-H3 length alone is that appreciable variance in loop length is only observed in the case of CDR-H3. Taken together, these data suggest that CDR-H3 flexibility may be a predictor of binding rates.

3.4.4 CDR-H3 flexibility correlates with binding rates and can be probed using computationally inexpensive methods

Based on our findings that CDR-H3 loop length correlates with binding rates, we next decided to examine whether flexibility of the antibody affects $\log(k_{on})$. We performed 100 ns molecular dynamics simulations of antibodies in the kinetic data set that did not have missing residues in the F_V region (144) and calculated RMSF (time average of RMSD aligned to the average structure in the simulation) for whole F_V , the FW region, all CDR loops, CDR-H1-3, paratope residues, and CDR-H3 alone, after aligning the simulations on each region independently. We unsurprisingly found that the flexibility of the F_V and the FW region correlate only weakly with the binding rate of the antibody, if at all, since the FW region is not usually involved in antigenic recognition and makes up the majority of the F_V (Fig 3.5a). We then turned our attention to the CDR loops, and found once again that the average RMSF values of the full set of CDR loops in the simulations were poor predictors of binding rates. Based on our previous observation that it is the length of the CDR-H3 loop specifically which is correlated with binding rates, we decided to examine V_H CDRs as well as CDR-H3 in isolation. We found that flexibility of V_H CDRs provides a predictor of binding rates but found that the effect was greater using CDR-H3 alone (Fig 3.5a).

Both the length and flexibility of CDR-H3, which are linked parameters, correlate with binding rate. Flexibility showed the stronger effect, with a maximum Spearman's ρ of -0.44 at an $fnat_{VH}$ cutoff of 0.65, compared with $\rho = -0.27$ for both CDR-H1-3 and all CDR loops, though the correlations were not statistically significant in the latter two cases. We again found that the predictive power of flexibility was dependent on most of the interaction being mediated through the heavy chain.

Lastly, we examined the flexibility of the paratope itself as a predictor of binding rates, including only residues that interact with the antigen, and found that CDR-H3 flexibility was still a better predictor of binding rates (Fig 3.5a). This suggests

that even regions of the CDR-H3 loop that do not make direct contact with the antigen affect binding rates, since not all of the CDR-H3 forms interactions with the antigen and because CDR-H3 flexibility provides a better predictor than the paratope alone. We posit that this might be due to the fact that these residues still provide a scaffold around other residues in the paratope, and therefore are still able to influence the binding rate.

Since our simulations were relatively short at 100 ns, we hypothesized that greater coverage of the conformational space of the antibodies would further increase the quality of the correlations, since it has been shown that conformational transitions of the CDR-H3 loop can occur over timescales as long as hundreds of microseconds (Fernández-Quintero et al. 2020b). As performing MD simulations of this magnitude would be prohibitive when considering the entire kinetic data set, we aimed to probe the conformational ensemble of these antibodies using less resource-intensive methods. To this end, we investigated the use of modelling tools MODELLER, Sphinx, ABLooper, and ABB2 as methods to approximate flexibility. In the case of MODELLER and Sphinx we examined the conformational diversity of the potential CDR-H3 loop conformations they generated. These conformational ensembles have previously been used to analyse structural flexibility (Marks et al. 2018). In the case of AbLooper and ABB2, we used their internal error estimates as proxies for flexibility, since uncertainty estimates of deep learning models like AF2 have similarly been shown to correlate with flexibility (Ma et al. 2023).

We found that the Sphinx, AbLooper, and ABB2 determinants of flexibility all provided stronger correlations with binding rates than short MD simulations (Fig 3.5b). We therefore argue that these methods do indeed probe the conformational ensemble more effectively. MODELLER, on the other hand, performed less effectively than the simulations (Fig 3.5b). We argue that this might be due to the fact that MODELLER, being a *de novo* loop modelling method, does not use heuristics based on known antibody structures to build models. Indeed, we found that mean pairwise RMSD values of decoys generated with MODELLER correlated strongly with loop length (Fig A.4), suggesting that sequence-based information

on loop conformations is poorly captured by this method. On the other hand, the three methods that performed better than short MD simulations had much lower correlations with loop length (Fig A.4). ABB2, the method that predicts crystal loop structures with the lowest RMSD to the target loop performed the least effectively in generating flexibility information, likely due to the fact that accurate prediction of protein structures using deep learning methods does not necessarily indicate that the model is aware of the underlying biophysics (Outeiral et al. 2022).

Next, we tested whether methods like Sphinx do in fact provide greater coverage of conformational space than short MD simulations. To do this, we pooled the 500 decoys generated by Sphinx for each antibody with 500 snapshots from the MD simulation of each antibody, taken at regular intervals. We then clustered the resultant pooled structures and assigned a cluster to each conformation. We examined the number of clusters identified by each method and found that Sphinx decoys sampled a larger number of possible conformations than MD simulations (Fig 3.6c). As a case study, we visualised the decoys generated by Sphinx (Fig 3.6a) and MD frames (Fig 3.6b) for the antibody in PDB structure 1BJ1. We found that Sphinx provided a much broader set of conformations and that these conformations were more likely not to be sampled in the simulation. Indeed, comparisons across the entire kinetic data set showed that most of the conformations found in the MD simulation were also identified by Sphinx (3.6d).

Lastly, we calculated correlation coefficients between the dissociation rates and predicted flexibility values and found no signal (Fig A.5). This shows that antibody flexibility is a predictor of binding rates alone. This explains the apparent lack of correlation between binding and dissociation rates of antibodies (Fig 3.4a) and suggests that engineering flexibility would result in selective modulation of binding rates without an impact on dissociation rates.

We conclude that CDR-H3 flexibility is a specific predictor of antibody binding rates and that this can be probed either by MD simulations or by using less computationally intensive methods that make use of heuristics based on known antibody structures.

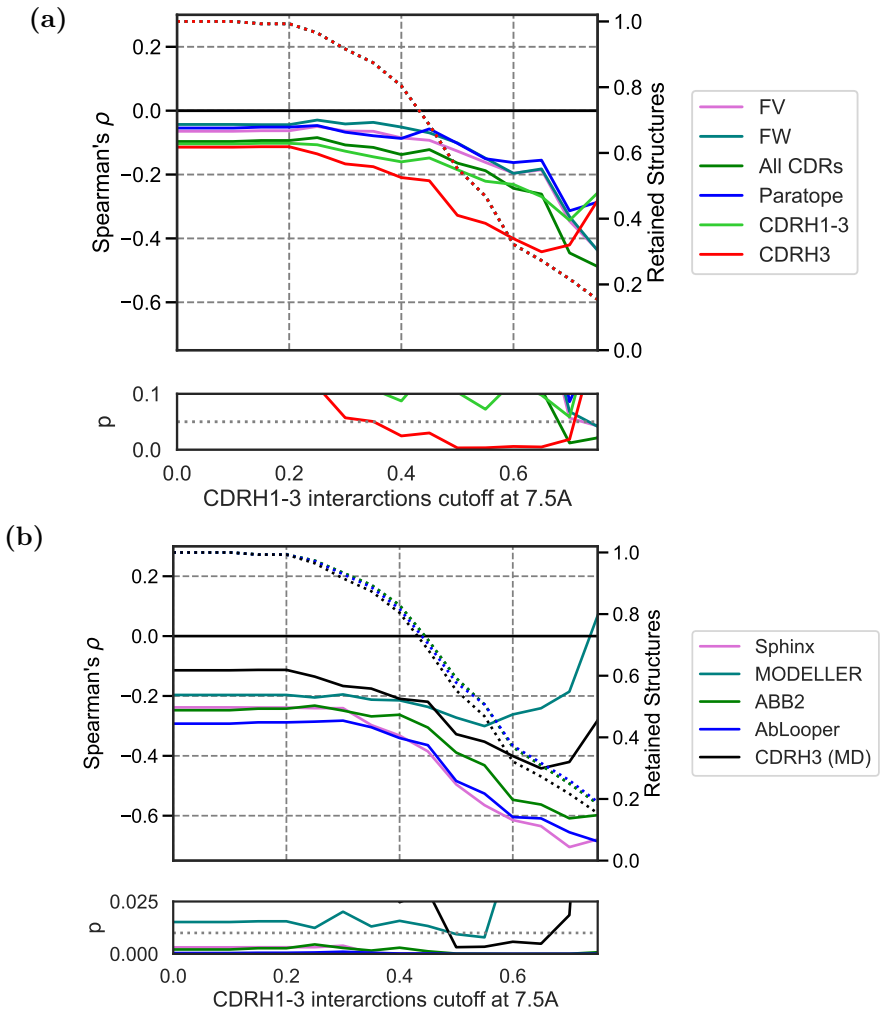


Figure 3.5: Correlations between antibody flexibility as predicted through MD simulations and antibody modelling methods and binding rates. At each fnat_{VH} cutoff (CDR-H1-3 interactions cutoff at 7.5 Å), we filtered the kinetic data set and correlated loop lengths and $\log(k_{on})$. The fraction of structures from the original data set remaining at each cutoff is shown using dotted lines (right y-axis) while the Spearman rank correlation coefficient (ρ) is shown in cyan for each cutoff (left y-axis). **(a)** shows the correlations between binding rates and RMSF values of labelled regions of the antibody, with p-values for each correlation shown below (gray dashed line: 0.05). **(b)** shows the correlations between predicted flexibility of the CDR-H3 loop and binding rates using different methods. For Sphinx and MODELLER, the median pairwise RMSD is used as the predictive variable. For ABB2 and AbLooper, the average predicted error across the loop is used for correlations. p-values for each correlation shown below (gray dashed line: 0.01).

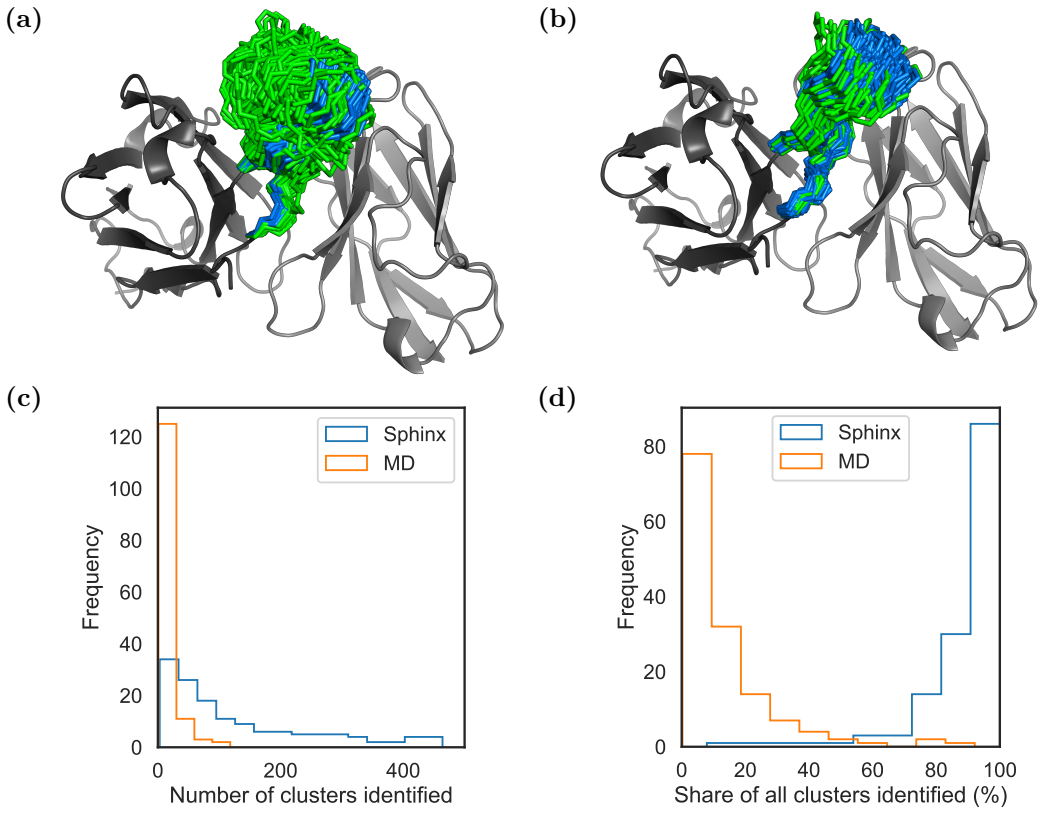


Figure 3.6: Comparisons of Sphinx and MD conformational ensembles. **(a)** shows the CDR-H3 loop structures of 1BJ1 predicted by Sphinx, with unique conformations not identified by MD shown in green. **(b)** shows the CDR-H3 loop structures of 1BJ1 identified in the MD simulation, with unique conformations not identified by Sphinx shown in green. **(c)** shows the numbers of structural clusters identified by Sphinx and MD simulations across the entire kinetic data set, while **(d)** shows the fraction of all clusters identified by either method.

3.4.5 Sphinx efficiently identifies differences in CDR-H3 flexibility

In order to validate that the method developed here for flexibility analysis and binding rate prediction does in fact extract information about the flexibility of antibody CDR-H3 loops on a scale beyond the short 100 ns simulations, we used two approaches. Firstly, we identified a non-redundant set of unbound antibodies at 99% sequence identity for which X-ray crystallographic structures at a resolution cutoff of 2.8 Å were available. This yielded a set of 84 structures. We then searched through this data set and identified 15 structures wherein at least one of the Fab fragments in the asymmetric unit had a disordered CDR-H3 loop. We then applied

our flexibility prediction method using Sphinx to these 84 structures and showed that 15 unresolved and 69 resolved structures have different predicted flexibility scores. We also assessed whether filtering the generated decoys and using only the top n decoys (as ranked by the internal scoring function of Sphinx) would have an effect on the predictor. Assessing differences with a Mann-Whitney U test, we found that regardless of whether decoys are filtered, the unresolved structures had a higher predicted flexibility score, although the significance of this dropped slightly when decoy filtering was introduced (Fig 3.7).

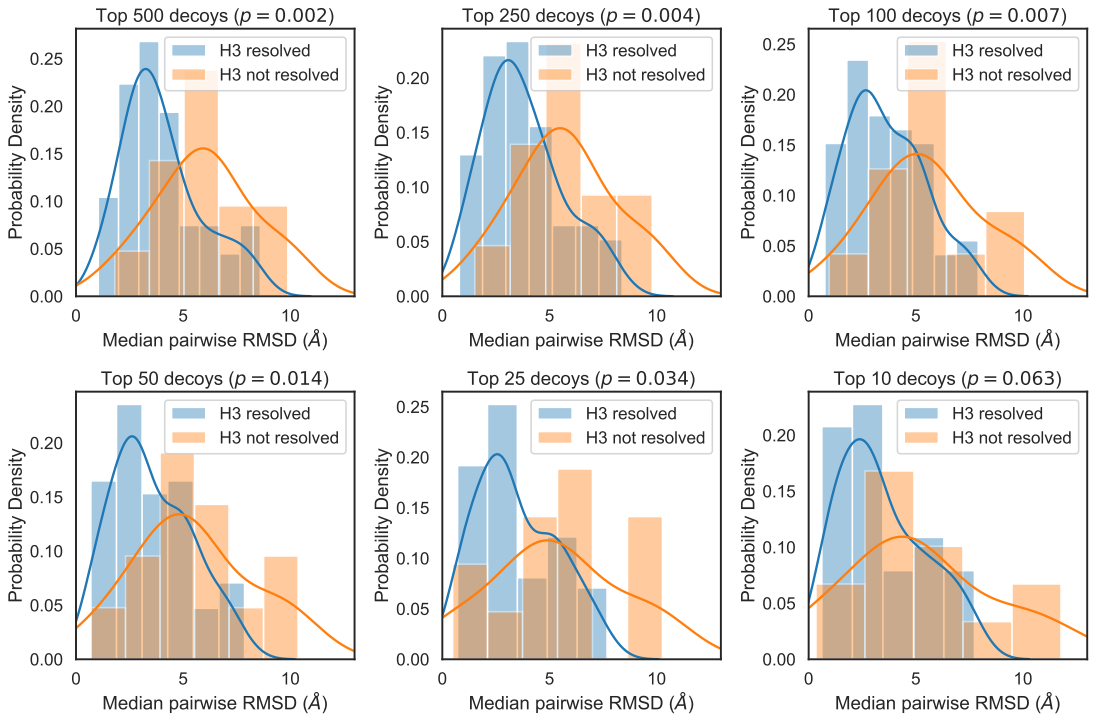


Figure 3.7: Differences in CDR-H3 flexibility as predicted by Sphinx for unresolved loops. We calculated the predicted CDR-H3 flexibility of a non-redundant set of unbound antibodies using Sphinx and differences observed between resolved and unresolved CDR-H3 loops when the top n decoys were selected for pairwise comparison. Reported p-values are for Mann-Whitney U tests, as the dsitribution of median pairwise RMSD values did not exhibit normality.

Since our non-redundant data set is likely to include antibodies of various different clonotypes, we also tested whether our method could be used to pick out more subtle differences in flexibility that may arise during affinity maturation as a result of small changes to antibody sequences. To answer this question, we validated our method

against data generated by Fernández-Quintero *et al.* (Fernández-Quintero et al. 2020c). The authors of this study used metadynamics simulations using CDR-H3 dihedral angles as collective variables to exhaustively sample CDR-H3 conformations of antibody pairs, where one antibody is naive and the other has undergone affinity maturation, and showed that the antibodies used all had more rigid CDR-H3 loops in the affinity-matured state. Of note is the fact that while many of these antibodies do have mutations in the CDR-H3 loop, several only have mutations surrounding CDRs or the framework regions and lack any changes in the CDR-H3 loop itself. The CDR-H3 sequences of these antibodies are given in Table 3.4

naive	mature	naive CDR-H3	mature CDR-H3	CDR-H3 difference (%)
1mlc	1p2c	GDGNYGY	GDGFYVY	28.6
1fl6	1kel	DGSYAMDY	WGSYAMDY	12.5
1aj7	1gaf	YYGIY	YYGIY	0
1d6v	1axs	GHSYYFYDGDY	GHSYYFYDGDY	0
1dqj	1ndg	WGGDV	WDGTY	60
4yhi	4yho	GTSGYDYFDY	GTSGWDYFDY	10
4nja	4nj9	RGGYGIRGYFDV	RGGYGSRGYFDV	8.3
1t66	4fab	YGYHGAY	SYYGMDY	71.4
1n7m	1ngw	RDSDY	RDMDY	20

Table 3.4: Summary of structures used for molecular dynamics simulations by Fernández-Quintero *et al* (Fernández-Quintero et al. 2020c). The structures span a range of CDR-H3 loop lengths (according to the Chothia definition) and have varying degrees of sequence diversity between the naive and mature antibodies. PDB accession codes are also provided.

Using Sphinx to assess the flexibility of naive and mature pairs (Figure 3.8), we found that Sphinx was unable to identify rigidifcation in the majority of cases when all 500 decoys generated by Sphinx were used in pairwise comparisons. The only cases in which rigidification was predicted were 1N7M and 1T66. We then tested whether using a loop scoring function to remove poorly modelled decoys from our analysis would give a more accurate description of the number of viable conformations for each loop. When the top 10% of decoys (50 decoys out of 500)

were retained for analysis, the method was able to predict rigidification for six out of nine pairs. Notably, 1D6V (0% change), 4YHI (10% change) and 4NJA (8.3% change) were correctly predicted as being rigidified during affinity maturation, suggesting that including the scoring function filtering step in our analysis allows Sphinx to identify stabilising interactions that the CDR-H3 makes with the rest of the antibody. The predictive power of our method decays when more stringent thresholding is used. We believe that this may be due to the fact that there are too few conformations left for comparison when this is the case.

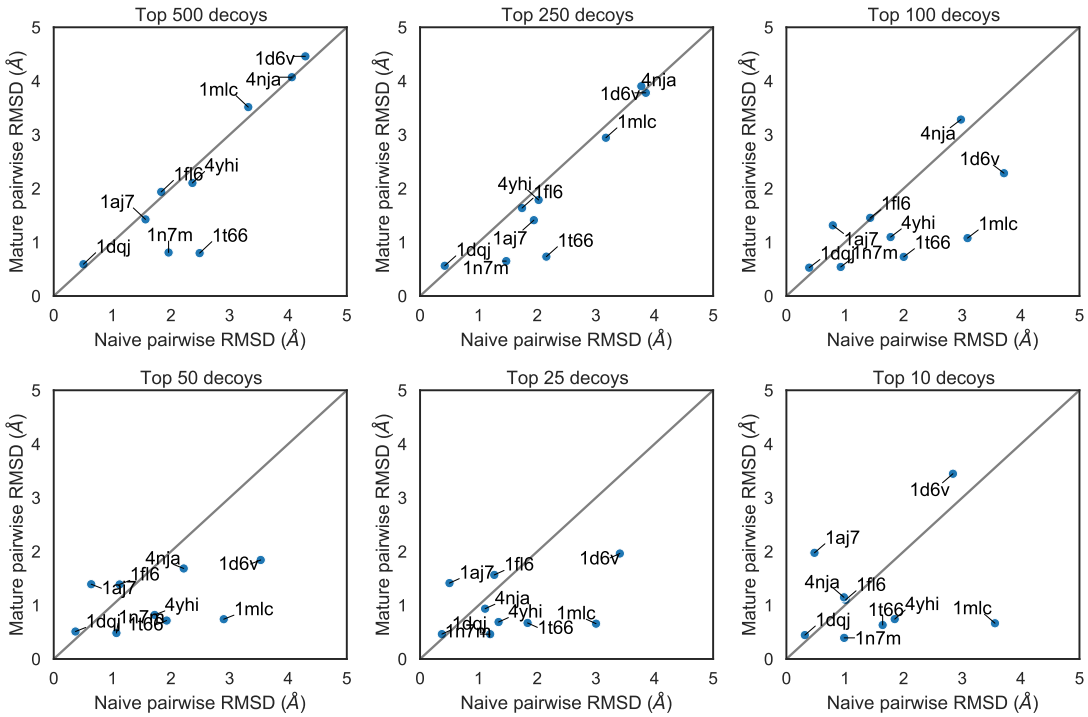


Figure 3.8: Prediction of CDR-H3 rigidification during affinity maturation. We compared the predicted flexibility of naive and affinity matured antibodies given in Table 3.4 using Sphinx (each pair is labelled with the accession code of the naive antibody) and used the built-in scoring function to pick out only the top n decoys. Points below the $x = y$ line indicate antibodies that are predicted to be rigidified during affinity maturation.

These results suggest that, using Sphinx, we are able to predict flexibility. Further, we conclude that there are two ways in which this occurs. When there are large enough differences in the CDR-H3 sequence, Sphinx is able to predict flexibility based on conformational diversity of fragments that are available in the CDR-H3 library. When differences are more subtle, the loop scoring function is

able to pick out stabilising interactions caused by changes in sequence, even though templates used may be similar. Taken together with the fact that Sphinx was able to predict binding rates, we conclude that its ability to predict binding rates is a result of Sphinx being able to generate conformational ensembles of CDR-H3 that are in line with the flexibility of the loop.

3.4.6 Secondary structure and surface flexibility of antigens influence antibody binding rates

We next examined whether antigen flexibility also influenced binding rates. In line with our results on CDR-H3 flexibility and loop modelling algorithms, recent work has shown that AF2 (Jumper et al. 2021) confidence scores are predictive of local flexibility (Ma et al. 2023; Guo et al. 2022). Thus, we began by using AF2 models of antigen structures in our kinetic data set and examined the correlation between the binding rate and confidence score, but we did not identify a significant correlation ($\rho = 0.16$, $p = 0.14$). We posit that this is because AF2 was trained on the structures in this data set, and therefore predicts them with near perfect accuracy. Indeed, upon further examination, we found that all bar one of the epitopes in our data sat had an average confidence score of 70 or greater across the epitope, which AF2 categorises as a confident prediction. The small range of confidence scores likely prevents us from extracting meaningful information on the flexibility of the epitopes. Extracting flexibility information from AF2 predictions using subsampled multiple sequence alignments as demonstrated by del Alamo *et al.* (del Alamo et al. 2022) may be one way to generate more insightful results.

As a proxy for flexibility we next examined the effect of secondary structure on binding rates. It has previously been shown that alpha-helical and beta-strand regions of proteins show lower flexibility than regions made up mainly of coils (Yaseen et al. 2016). We generated a non-redundant data set from our kinetic data set based on antigen sequences (see Section 3.3) and assigned secondary structure to the antigen using DSSP (Kabsch et al. 1983). We then calculated the fraction of secondary structure elements in each epitope and categorised them by the dominant

secondary structure. We found that antibodies targeting epitopes predominantly made up of beta-strands or alpha helices had significantly higher binding rates, 2-fold on average (Fig 3.9a), but found no statistically significant difference in the case of dissociation rates (Fig 3.9c). We then further separated the data set into helix-, sheet-, and coil-dominated epitopes and found that helical epitopes had higher binding rates than coil-based epitopes (Fig 3.9b). We also found that helical epitopes had a greater average binding rate than beta-strand-dominated epitopes, with an average binding rate difference of 1.4-fold (this difference was not statistically significant), indicating that the effect observed in the two-class analysis is driven by helical epitopes. Once again, we found that the type of epitope did not influence the dissociation rate (Fig 3.9d). Lastly, we found that the differences in binding rates have an observable effect on the overall affinity of antibodies targeting helix- *versus* coil-dominated epitopes (Fig 3.9f), but not for antibodies simply targeting epitopes with defined secondary structure that is either helix- or strand-based (Fig 3.9e).

We conclude that the secondary structure of the epitope has a significant effect on binding rates of antibodies. Based on previous literature and our findings when examining the antibody side of the complexes in our kinetic data set, we interpret this as being due to the fact that different secondary structure elements show different levels of flexibility and that less flexible epitopes correlate with higher on-rates.

3.4.7 Limited exploration of antibody conformational ensembles enriches profiling of therapeutic properties

Therapeutic antibody development not only hinges on optimisation of binding affinity and kinetics, but also places great importance on the engineering of biophysical properties that are likely to result in liabilities during production and administration. Raybould *et al.* (Raybould et al. 2019a; Raybould et al. 2023), demonstrated that clinical-stage therapeutic (CST) antibodies occupy a subset of surface property space compared to antibodies in general. Namely, they identify as properties that may require optimisation the total length of the CDR loops, the extent and magnitude of patches of surface hydrophobicity (PSH), positive charge (PPC) and negative

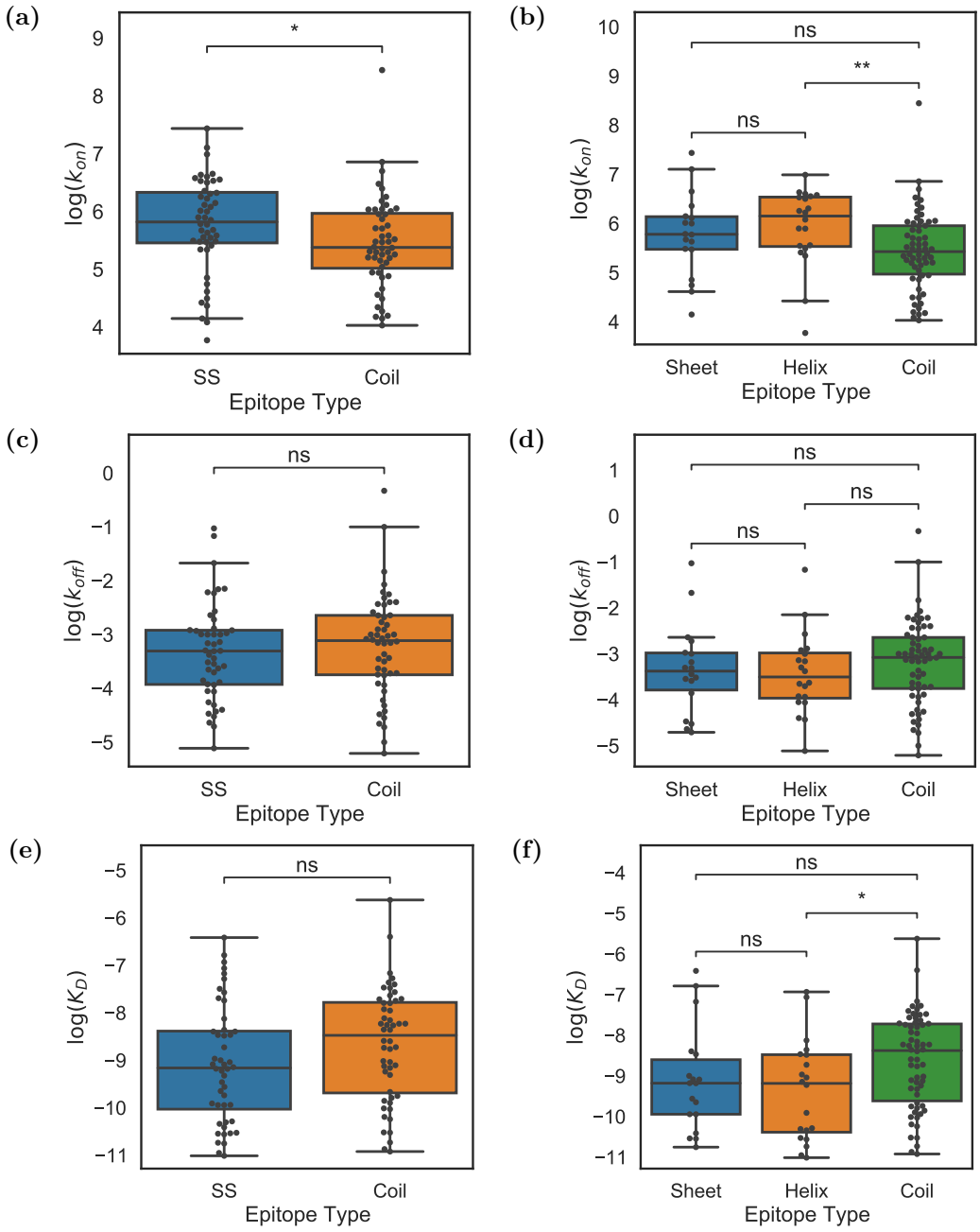


Figure 3.9: Differences in binding kinetics based on epitope secondary structure. Epitopes are characterised using DSSP at a heavy atom distance cutoff to the antigen of 4.5 Å. **(a)**, **(c)**, **(e)** show the $\log(k_{on})$, $\log(k_{off})$, and $\log(K_D)$, respectively, using the binary classification system, with "SS" denoting helix/strand-dominant epitopes and "Coil" denoting coil-dominated epitopes. **(b)**, **(d)**, **(f)** show results when "SS" is further decomposed into the type of secondary structure that is dominant in each epitope. Significance was tested using a Mann-Whitney U test as stratified samples could not be assessed for normality based on D'Agostino and Pearson's omnibus test due to low sample sizes.

charge (PNC) in the CDRs, and asymmetry in the net heavy- and light-chain surface charges (SFvCSP). They also recommend guidelines on the range within which these should be kept during therapeutic antibody development in the form of numerical thresholds based on the values seen for therapeutic antibodies. TAP assigns flags to a structure, based on the calculated scores for the current structure compared to all known CST structures. For the PSH score, an amber flag is assigned to a structure if it is within either the upper or lower 5^{th}percentile of the CST distribution, and a red flag is assigned if it falls outside of the CST distribution. For PPC and PNC scores, amber flags are assigned if the structure is in the upper 5^{th}percentile of the CST distribution and red flags are assigned to structures with scores above the CST distribution. Finally, SFvCSP amber flags are assigned if the structure is in the lower 5^{th}percentile of the CST distribution and red flags are assigned to structures with scores below the CST distribution. Scores that do not fall within these extrema are assigned green flags. Values used to define flags are shown in Table A.3.

Antibodies do not exist as single states in solution, but are able to undergo dramatic conformational shifts, especially in the CDR loops (Fernández-Quintero et al. 2020b). We hypothesised that the use of conformational ensembles during analysis of these properties might enable a richer understanding of a therapeutic candidate. To this end, we used ABB2 to model a total of 14 clinical stage therapeutic antibodies (CSTs) found in TheraSAbDab (Raybould et al. 2019b) whose structures are known (Table A.2) and performed 200 ns MD simulations on Fab fragments that were generated by grafting C_H1 and C_L domains from crystal structures onto the F_V models generated. We assessed equilibration of the systems by following the RMSD of the F_V region to the starting structure and calculated conformer-dependent properties (PSH, PPC, PNC, and SFvCSP) for each frame in the trajectories. We also calculated the simulation average of each property over the last 120 ns of simulation based on visual assessment of the RMSD curves (Fig A.6).

We saw that whether changes in properties were observed under simulation conditions was heavily dependent on the system and the property. Generally, we found that when TAP flags a modelled structure as having possible liabilities, it

also does so for the simulation. However, we found that for many of the antibodies, multiple states for a property were observed, with the antibody switching between these states, indicating that behaviour of the conformational ensemble is indeed more complex. We discuss specific examples here, but provide traces for all CSTs tested in Appendix A.4

We observed the largest fluctuations in the case of surface hydrophobicity (PSH) (Fig 3.10, A.7). In almost all cases, the simulated antibody visited regions beyond the amber and red flags, even if the modelled antibody and simulation average had similar values. For example, the ABB2 model of Erenumab shows the same PSH score as the simulation average (ABB2: 183.9 (amber flag), MD average: 182.18 (amber flag)), but we note that there are two predominant states that result in this. One of these states has a PSH score of approximately 160, whereas the second, dominant state has a higher PSH score of approximately 190 (Fig 3.10a). In other cases, the simulation and modelled structure have different scores, but are still given the same flag by TAP, with Dostarlimab (ABB2: 114.47 (green flag), MD average: 132.12 (green flag)) being an example of this (Fig 3.10b). Lastly, we identified cases where TAP assigns different flags to the average from the simulation and the input model. Fepixnebart is an example of this. While the input model is assigned a green flag by TAP (PSH score: 148.83), the simulation average is assigned an amber flag (PSH score: 180.99) (Fig 3.10c).

Surface positive charge density (PPC) (Fig 3.11, A.8) mirrored the behaviour observed in PSH scores. Boserolimab showed the same PPC score over the course of the simulation (0.06, green flag) as it does for the model average (0.13, green flag), with only rare fluctuations (Fig 3.11a). We observed that other antibodies showed greater changes in PPC scores, with some retaining the TAP flag assigned to modelled structures in simulation such as Fepixnebart (ABB2: 0.04 (green flag), MD average: 0.43 (green flag)) (Fig 3.11b), and others being assigned different flags by TAP, such as Simaravibart (ABB2: 1.50 (amber flag), MD average: 0.88 (green flag)) (Fig 3.11c). In these cases, fluctuations between different PPC scores are rapid, suggesting that PPC score calculation is sensitive to small changes in

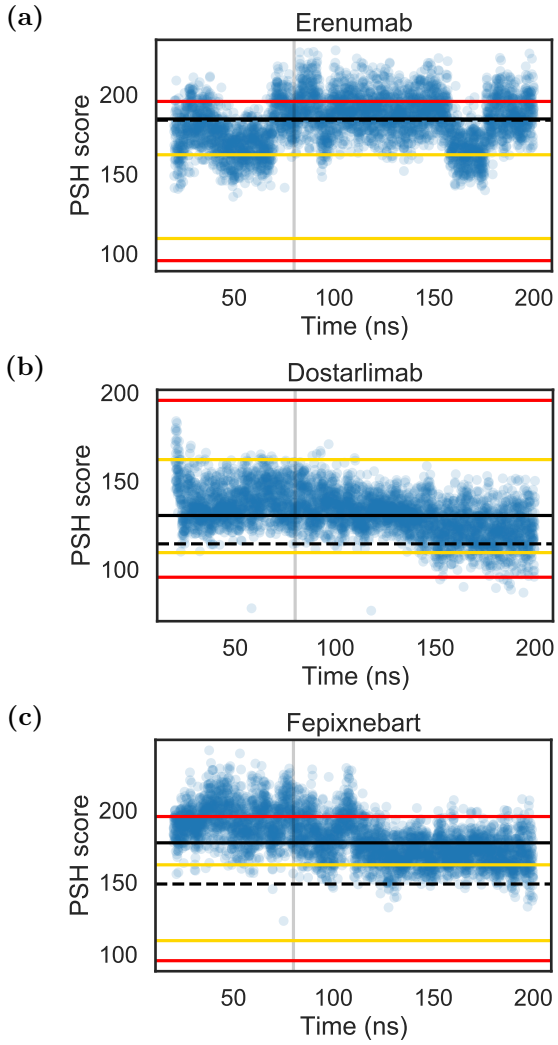


Figure 3.10: Changes in surface hydrophobicity (PSH) scores of antibodies during simulation. The region between the amber lines corresponds to PSH scores that would be assigned a green flag by TAP and the regions outside the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average PSH score from three ABB2 models, whereas the solid black line shows the average PSH score over the final 120 ns of the trajectory (marked by the grey line). **(a)** shows a case where the simulation average corresponds well to the model average, **(b)** shows a case where TAP would still assign the same flag to the simulation and model averages, but the PSH score has shifted substantially, and **(c)** shows an example where the simulation average would be assigned an amber flag, but the model average would be assigned a green flag.

conformation that can happen over small timescales. We posit that this is likely due to movement of side chains of charged residues.

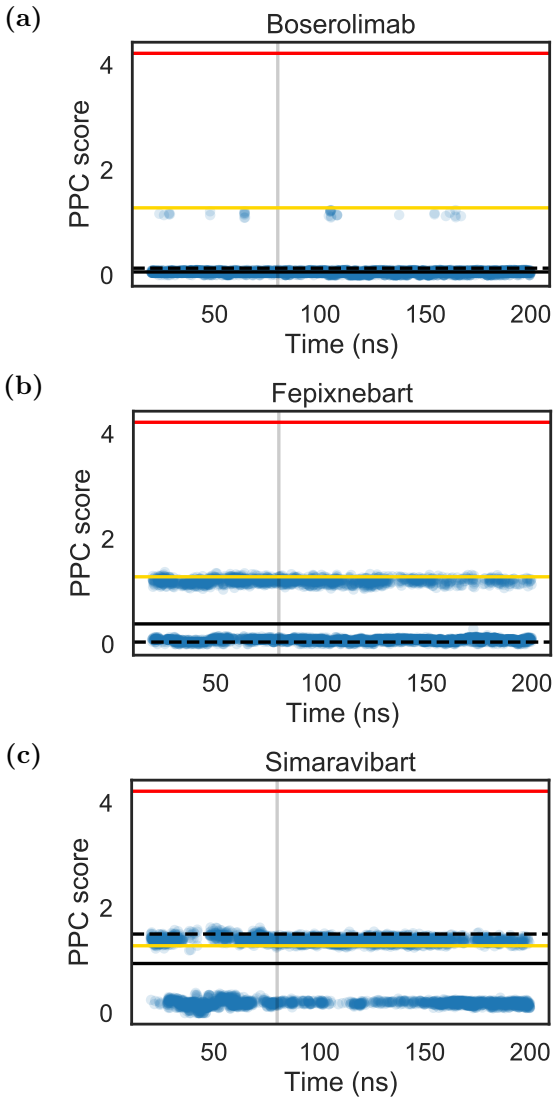


Figure 3.11: Changes in surface positive patch (PPC) scores of antibodies during simulation. The region under the amber lines corresponds to PPC scores that would be assigned a green flag by TAP and the regions above the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average PPC score from three ABB2 models and the solid black line shows the average PPC score over the final 120 ns of the trajectory (marked by the grey line). **(a)** shows a case where the simulation average corresponds well to the model average, **(b)** shows a case where TAP would still assign the same flag to the simulation and model averages, but the PPC score has shifted substantially, and **(c)** shows an example where the simulation average would be assigned an amber flag but the model average would be assigned a green flag.

For surface negative patches (PNC) (Fig 3.12a, 3.12b, A.9) and asymmetry in

the net heavy- and light-chain surface charges (SFvCSP) (Fig 3.12c, 3.12d, A.10), we observed similar changes, but did not identify cases where the flag assigned by TAP changed as a result of the simulation. The behaviour of PNC scores is similar to that of PPC scores, with changes being a result of fast transitions (Fig 3.12b). For SFvCSP scores, we observed greater variation, likely because it takes into account both positive and negative patches on the surface of the antibody. Here, we found that antibodies adopt conformations that are beyond the assigned flag (Fig 3.12d), but similar to PSH scores. Since these are minor states, the assigned flag does not change.

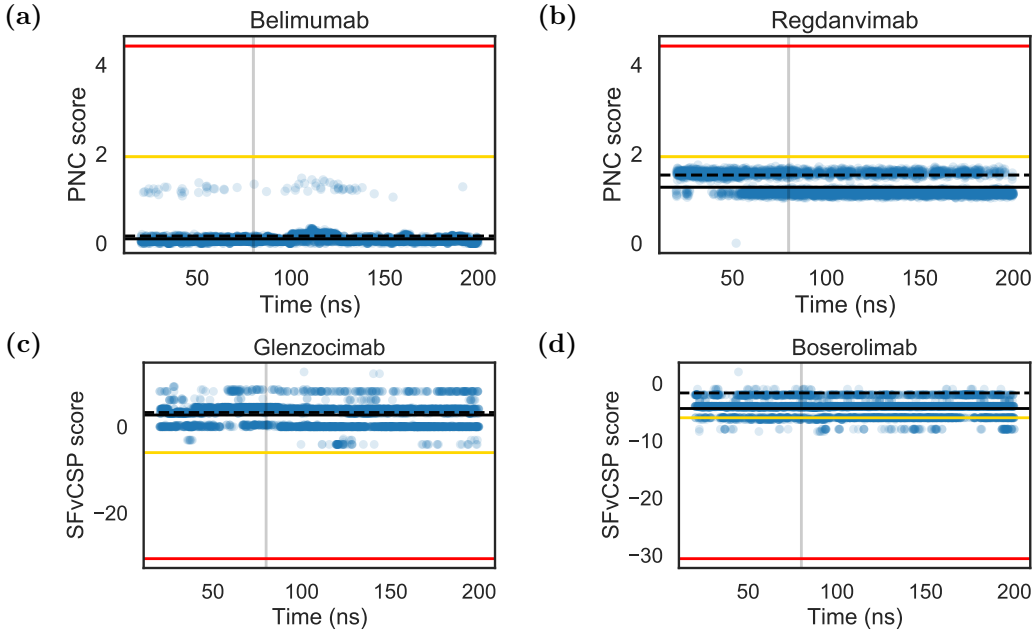


Figure 3.12: Changes in surface negative patch (PNC) and asymmetry in the net heavy- and light-chain surface charges (SFvCSP) scores in of antibodies during simulation. The region below the amber line (for PNC) corresponds to scores that would be assigned a green flag by TAP (for SFvCSP, the region above the amber line is assigned a green flag) and the regions outside the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average score from three ABB2 models and the solid black line shows the average score over the final 120 ns of the trajectory (marked by the grey line). **(a)** and **(c)** show cases where the simulation average corresponds well to the model average, and **(b)** and **(d)** show cases where TAP would still assign the same flag to the simulation and model averages.

The simulations employed here are relatively short, yet reveal the extent to which changes that can occur over relatively short timescales can influence surface

properties of antibodies. They also suggest that more extensive sampling with enhanced sampling methods and longer simulations may reveal conformations that are inaccessible to short, unbiased simulations that result in drastic changes to these properties. Based on our results, we conclude that examination of conformational ensembles can indeed give a richer understanding of surface properties relevant to therapeutic antibody design.

3.5 Discussion and conclusions

Understanding the biophysical properties of antibodies is crucial to the development of both therapeutic antibodies and novel vaccines, as different pathogens and antigens will require optimisation of different aspects of an antibody therapeutic. Here, we show that the binding rate of an antibody, which is a key influence on the overall affinity, is determined at least in part by the flexibility of the CDR-H3 loop. Similarly, a less dynamic epitope also correlates with increased binding rate.

Using a hand-curated data set, we first show that classical analyses of the antibody-antigen complex, such as the buried surface area or interactions within the interface, yield little information about the binding rate that characterises the complex. We identify only electrostatic interactions as a weak predictor of binding rates, likely due to their comparatively long-range effects which allow binding partners to be "pulled" into the correct binding geometry through electrostatic steering (Vascon et al. 2020). However, based on the relatively weak correlation that we find, we argue that it is unlikely that this plays a major role in determining antibody binding rates.

We then show that flexibility, on the sides of both the antigen and the antibody, is a more powerful predictor of the binding rate. We show that the flexibility of the CDR-H3 loop specifically predicts binding rates and that this is the crucial loop for paratope dynamics. This follows from previous studies that have shown that CDR-H3 is more flexible than the other CDR loops (Fernández-Quintero et al. 2020b; Fernández-Quintero et al. 2020a; Guloglu et al. 2023). When examining the antigen, we show that certain epitopes result in a faster association of the binding partners. This is a significant point for immunogen design, especially for pathogens such as malaria, where circulating antibodies have a limited time frame to bind the target and should be considered in cases where multiple candidate epitopes are able to generate neutralising antibody responses.

Marked by the increases in both repertoire sequencing efficiency and antibody modelling accuracy, computer-aided approaches to high-throughput screening of antibodies are becoming more commonplace (Hummer et al. 2022). A key challenge

lies in predicting the biophysical properties of antibodies without the need to purify and test each candidate sequence. Extracting information on protein flexibility is challenging both *in vitro* and *in silico*, due to the traditionally low-throughput methods that are commonly used for this, as discussed in Chapters 1 and 2. We show that recent advances in antibody structure prediction, as well as proxies for epitope flexibility, can be used to gain at least partial information on the flexibility of the binding partners in the complex. Importantly, we note that many deep learning-based approaches fail to generalise on non-redundant antibodies when predicting binding affinities without sufficient training examples (Hummer et al. 2023). We argue that this is in part due to the lack of flexibility information to which the models are exposed during training and hypothesize that inclusion of conformational ensembles during training would greatly improve the accuracy of such models and aid in their ability to generalise across different clonotypes.

We also tested how the assessment of conformational ensembles enables richer insights into surface properties of candidate therapeutic antibodies using MD simulations. We found that even relatively short simulations give information on the multitude of states of an antibody, and that these can result in ensemble averages that differ from values derived from calculations on static models. We found that in many cases, the simulation average corresponds well to results from static analyses, but we propose that assessment of ensembles may be a useful tool during later stages of therapeutic development. For instance, we propose that if a possible therapeutic shows very favourable properties for antigenic recognition but is shown to be prone to aggregation or has other liabilities, MD simulations could be employed to identify states of the antibody that result in unfavourable surface properties. These can then be selected against through computational engineering using approaches such as multi-state design (Sauer et al. 2020), thereby allowing a more nuanced approach to the engineering pipeline. Further, this type of analysis will likely become much less resource-intensive in the future with computational methods that accurately predict alternate conformations of proteins becoming increasingly accurate (del Alamo et al. 2022; Hou et al. 2023; Jing et al. 2023).

The results of this chapter will be valuable to the rational vaccinology and antibody engineering communities, where detailed understanding of the conformational ensemble of an antibody will yield greater insights into its properties, thereby enabling more accurate design of both effective therapeutics and vaccines. Taken together with the recent advancements in multi-state prediction of protein structures, we believe that, as methods to generate conformational ensembles become faster and less resource-intensive, approaches like ours will become more commonplace in the future.

4

The structure and structural variability of CDR-H3 loops are influenced by specific attributes of the V_L domain

Contents

4.1 Chapter abstract 89
4.2 Introduction 90
4.3 Methods 92
4.3.1 Comparison of paired and single-domain antibody CDR-H3 loops 92
4.3.2 Analysis of V_L domain influences on CDR-H3 conformations 96
4.4 Results 101
4.4.1 Single-domain antibody CDR-H3 loops show different conformations from paired antibody CDR-H3 loops . . . 101
4.4.2 Antibody V_H and V_L CDR loops are characterised by non-equivalent inter-loop contacts in crystal structures . 105
4.4.3 Conformational rearrangements in CDR-H3 do not require major conformational changes in minor loops . . . 107
4.4.4 Minor CDR loops create steric boundaries within which CDR-H3 samples conformational space 111
4.4.5 Antibody subtypes, genes, and CDR canonical forms show preference for CDR-H3 conformations with different levels of diversity 114
4.4.6 Identical V_H sequences show differences in CDR-H3 conformation when paired with different V_L sequences . . . 119
4.5 Discussion and conclusions 121

4.1 Chapter abstract

The binding domains of standard antibodies are their V_H and V_L domains, both of which present analogous CDR loops. It has long been known that the V_H CDRs contribute more heavily to the antigen-binding surface (paratope), with the CDR-H3 loop providing a major modality for the generation of diverse paratopes. As discussed in Chapter 1, this is thought to be due to both the combinatorial nature of V(D)J recombination and the conformational flexibility of the paratope. Here, we provide evidence for an additional role of the V_L domain as a modulator of CDR-H3 structure, thereby further expanding the possible V_H paratope states even further. To demonstrate this, we use a diverse set of antibody crystal structures and a large set of molecular dynamics simulations. We first show that CDR-H3 loops in single-domain and paired antibodies adopt distinct conformational states using a coarse-grained approach to structural analysis of these loops, and that this is likely due to the V_L domain of paired antibodies creating steric bulk and thereby forcing the CDR-H3 loop into different conformations. We show that specific attributes of the V_L domain such as subtypes, CDR canonical forms, and genes can influence the structural diversity of the CDR-H3 loop in paired antibodies, and we provide a physical model for how this effect occurs through inter-loop contacts and packing of CDRs against each other. Our results indicate that the rigid minor loops fine-tune the structure of CDR-H3, thereby contributing to the generation of surfaces complementary to the vast number of possible epitope topologies. They also provide insights into the interdependent nature of CDR conformations, an understanding of which is important for the rational antibody design process.

This chapter contains reproduced material from the following publications:

Guloglu, B., Deane, C.M. (2023) Specific attributes of the V_L domain influence both the structure and structural variability of CDR-H3 through steric effects. Frontiers in Immunology. doi.org/10.3389/fimmu.2023.1223802 (Guloglu et al. 2023)

Gordon, G.L., Capel, H.L., **Guloglu, B.**, Richardson, E., Stafford, R.L., Deane, C.M. (2023) A comparison of the binding sites of antibodies and single-domain antibodies. Frontiers in Immunology. doi.org/10.3389/fimmu.2023.1231623 (Gordon et al. 2023)

Declaration: Results presented in Section 4.4.1 are part of a project jointly conducted with Gemma L. Gordon. All of the work presented in this thesis was conducted, analysed, interpreted, and visualised by the author of this thesis.

4.2 Introduction

While the V_H and V_L domains are largely structurally symmetrical in their framework regions, the degree of symmetry is much lower in the CDR loops. Analogous CDRs of the V_L domain present distinct conformations when compared to those of the V_H (Chothia et al. 1987; North et al. 2011; Nowak et al. 2016). Most significant is the conformational space occupied by CDR-H3. While all other CDR loops can be classed into distinct canonical forms (Chothia et al. 1987; North et al. 2011; Nowak et al. 2016), the CDR-H3 loop is highly variable and does not lend itself to structural clustering (Regep et al. 2017). This is due to the much larger sequence space explored by this loop as a result of V(D)J recombination, as the CDR-H3 loop has sequence contributions from the V, D, and J genes, with additional diversity stemming from nontemplated nucleotide addition at each junction during the V(D)J recombination event (Alt et al. 1982), as discussed in Chapter 1. The analogous CDR-L3 loop shows much lower sequence diversity due to the lack of the D segment in the light chain. Here, we use the term structural variability to refer to the diversity of CDR loop conformations found in solved crystal structures and the term flexibility to refer to the ability and likelihood of an individual CDR loop to adopt multiple different conformations.

The dissimilarities between V_H and V_L CDRs are also reflected in their contributions to the paratope, with the CDR-H3 loop generally dominating the paratope and V_L CDRs making smaller contributions than their V_H counterparts, thereby

leading to imbalanced roles of the two variable domains in antigen binding (Akbar et al. 2021), even though both CDR-H3 and CDR-L3 are poised for interaction with the antigen by virtue of occupying the centre of the CDR loops. Indeed, there are many cases where the V_L domain makes no contacts with the antigen at all (Akbar et al. 2021; Phillips et al. 2021). Further, the presence of single-domain antibodies in chondrichthyes (IgNARs) (Khalid et al. 2022) and camelids (V_HHs) (Mitchell et al. 2018) (an account of their structures is provided in Chapter 1) shows that the presence of the V_L domain is not necessary for antigen recognition. These observations have raised questions about the precise role of the V_L domain and how the presence and sequence of this region influences the V_H domain.

Due to the spatial proximity of the CDR-H3 loop and the V_L CDRs, it has long been hypothesised that one possible role of the V_L domain, in addition to antigen binding, might be modulating the CDR-H3 structure. This has been confirmed in specific cases, showing that the crystal structure (Teplyakov et al. 2016) and conformational ensemble (Fernández-Quintero et al. 2021) of the CDR-H3 loop are indeed influenced by different pairings of V_H and V_L domains. Further, as it is placed at the interface between the V_H and V_L domains, the structure of the CDR-H3 loop may be influenced simply by the presence of the V_L domain; that is, by virtue of either having or missing a V_L domain, CDR-H3 loops of paired and single-domain antibodies may adopt different structures.

In this study, we show that specific attributes of the V_L domain influence both the structure and structural variability of CDR-H3. We begin by showing that CDR-H3 loops of paired antibodies (henceforth referred to as antibodies or Abs in this chapter) and single-domain camelid antibodies (henceforth referred to as single-domain antibodies or sdAbs in this chapter) show differences in how they fold against the rest of the structure due to the presence or absence of the V_L domain. We then show that CDR loops are tightly packed against each other in a distance-dependent manner in antibody crystal structures and that V_L loops are more heavily compacted than their V_H counterparts. Further, we show that

the CDR-H3 loop is unique in its packing, making interactions with more loops than the analogous CDR-L3 loop.

We next examine these observations in a more detailed way using a set of molecular dynamics simulations of eight antibody structures. We find that the observed packing patterns are recapitulated in simulations and that the V_L loops are all more rigid than their V_H counterparts, and we note that this effect would be expected given the greater compaction of the V_L loops. We also show that the CDR-H3 loop explores a significantly greater conformational space than other CDRs and makes contacts with all of these, suggesting that the remaining CDR loops (referred to as the minor loops) act as a "cage" around the CDR-H3, constraining the number of conformations that it is able to adopt.

Using a length-independent structural similarity calculation, we show that different V_L genes, subtypes, and CDR canonical forms have a significant effect on the structural variability of CDR-H3, with shorter CDR-L1 and longer CDR-L3 loops increasing conformational diversity of the CDR-H3 loop. We rationalise this observation in the context of our constrained CDR-H3 model. We provide a case study that shows the expected effects of V_L domain attributes on CDR-H3 conformation, and confirm using a structural data set that pairing the same V_H domain with different V_L sequences leads to a change in CDR-H3 conformation, with this change being correlated with the diversity of the V_L sequences.

4.3 Methods

4.3.1 Comparison of paired and single-domain antibody CDR-H3 loops

4.3.1.1 Data set curation

To compare the orientation of the variable domains of single-domain antibodies/nanobodies (sdAbs) and antibodies (Abs) to one another, we compared two structural data sets. We downloaded sdAb and Ab structures from SAbDab on 8.8.2022, removed any structures with missing backbone atoms in the CDR-H1-3 loops or their anchors (defined as three residues on either side of a loop), and

clustered these using cd-hit (Li et al. 2006) at a threshold of 0.95. We then extracted cluster representatives for the generation of non-redundant data sets. We note that the sdAb and Ab sets were generated individually, meaning that identical sequences could in theory be found across sets, but not in the same set. This yielded 116 unbound sdAbs, 388 sdAbs bound to their cognate antigens, 862 unbound Abs and 1977 bound Ab structures.

4.3.1.2 Analysis of CDR-H3 loop orientations

We analysed the general orientation of CDR-H3 loops by examining their centers of geometry in reference to an $\mathbb{R}^3$ coordinate system. We then describe the center of geometry of the CDR-H3 loop using spherical coordinates. To orient the coordinate system identically in all examined structures, we used the following fitting procedure.

Let $\mathbf{c}$ be the centre of geometry of the backbone atoms of IMGT positions 102, 103, 118, and 119. Let $\mathbf{p}$ be the first principal component vector of the anchor positions that minimises the distance to the Cα atom of IMGT position 103 when extended from $\mathbf{c}$. Let $\mathbf{a}$ be the centre of geometry of all V_H anchor positions (we take the anchors to be the three residues on either side of the loop as defined by the IMGT numbering scheme.). Let $\mathbf{\Pi}$ be the plane that crosses $\mathbf{c}$ whose normal vector is $\mathbf{p}$. Therefore for all $\mathbf{r} = \langle x, y, z \rangle$,

$$\mathbf{p} \cdot (\mathbf{r} - \mathbf{c}) = 0 \tag{4.1}$$

$$\mathbf{p} \cdot \mathbf{r} = \mathbf{p} \cdot \mathbf{c} \, . \tag{4.2}$$

We calculate the shortest distance of any point $\mathbf{r}$ from $\mathbf{\Pi}$ using

$$d(\mathbf{r}) = \frac{\mathbf{p} \cdot \mathbf{r} - \mathbf{p} \cdot \mathbf{c}}{||\mathbf{p}||} \, . \tag{4.3}$$

Let $\mathbf{u} = \mathbf{a_\Pi} - \mathbf{c}$, where $\mathbf{a_\Pi}$ is the orthogonal projection of $\mathbf{a}$ onto $\mathbf{\Pi}$, given by

$$\mathbf{a_\Pi} = \mathbf{a} - d(\mathbf{a}) \frac{\mathbf{p}}{||\mathbf{p}||} \, . \tag{4.4}$$

Let $\mathbf{v}$ be a vector orthogonal to both $\mathbf{u}$ and $\mathbf{p}$, such that the resultant coordinate system is left-handed,

$$\mathbf{v} = \underset{\mathbf{x}\in\{\mathbf{u}\times\mathbf{p},-\mathbf{u}\times\mathbf{p}\}}{\operatorname{argmax}} \begin{vmatrix} x^{(1)} & u^{(1)} & p^{(1)} \\ x^{(2)} & u^{(2)} & p^{(2)} \\ x^{(3)} & u^{(3)} & p^{(3)} \end{vmatrix} . \tag{4.5}$$

Given the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{p}$, such that $\mathbf{u}\perp\mathbf{v}\perp\mathbf{p}$, we aim to describe the CDR-H3 backbone atomic positions using spherical coordinates.

Let $\mathbf{h}$ be the centre of geometry of the CDR-H3 backbone positions, as defined by the IMGT numbering scheme. To describe $\mathbf{h}$ relative to $\mathbf{c}$ using spherical coordinates, $\mathbf{h_c}(\rho, \theta, \phi)$, we define $\rho = ||\mathbf{h} - \mathbf{c}||$. θ describes the elevation or polar angle and ϕ describes the azimuth. θ decreases as the CDR-H3 loop projects directly up and away from the $\mathrm{V_H}$ domain. ϕ increases as the CDR-H3 loop points out from the $\mathrm{V_H}$ domain. In the case of antibodies, this refers to a CDR-H3 loop that is oriented more towards the $\mathrm{V_L}$ domain.

We use $\mathbf{p}$ as the polar reference vector. Thus,

$$\theta = \cos^{-1} \frac{(\mathbf{h} - \mathbf{c}) \cdot \mathbf{p}}{||\mathbf{h} - \mathbf{c}||\ ||\mathbf{p}||} . \tag{4.6}$$

Lastly, we use $\mathbf{u}$ as the azimuthal reference vector. Therefore, ϕ will be the angle between $\mathbf{u}$ and $\mathbf{h_\Pi} - \mathbf{c}$, where $\mathbf{h_\Pi}$ is the orthogonal projection of $\mathbf{h}$ onto $\mathbf{\Pi}$,

$$\mathbf{h_\Pi} = \mathbf{h} - d(\mathbf{h}) \frac{\mathbf{p}}{||\mathbf{p}||} \tag{4.7}$$

and

$$\phi = \cos^{-1} \frac{(\mathbf{h_\Pi} - \mathbf{c}) \cdot \mathbf{u}}{||\mathbf{h_\Pi} - \mathbf{c}||\ ||\mathbf{u}||} . \tag{4.8}$$

In order to consistently measure the azimuth angle, we calculate

$$\begin{vmatrix} h_\Pi^{(1)} & u^{(1)} & p^{(1)} \\ h_\Pi^{(2)} & u^{(2)} & p^{(2)} \\ h_\Pi^{(3)} & u^{(3)} & p^{(3)} \end{vmatrix} . \tag{4.9}$$

As we aim to measure the clockwise displacement angle using **p** as a reference, we use the calculated ϕ as the azimuth angle when the determinant is positive. Otherwise we define the azimuth reference angle as $2\pi - \phi$.

Using this spherical coordinate system, the parameter ρ represents the reach of the CDR-H3 loop away from the V_H domain, ϕ describes the horizontal orientation of the CDR-H3 towards the rest of the V_HH (for sdAbs), or against the V_L domain (for Abs), and θ describes loop extension where a low value corresponds to a CDR-H3 extending up and away from the rest of the V_H domain. We also define an additional parameter referred to as compaction, which is given by the number of residues of the loop divided by the average reach, ρ. Figure 4.1 depicts these parameters on a V_H domain.

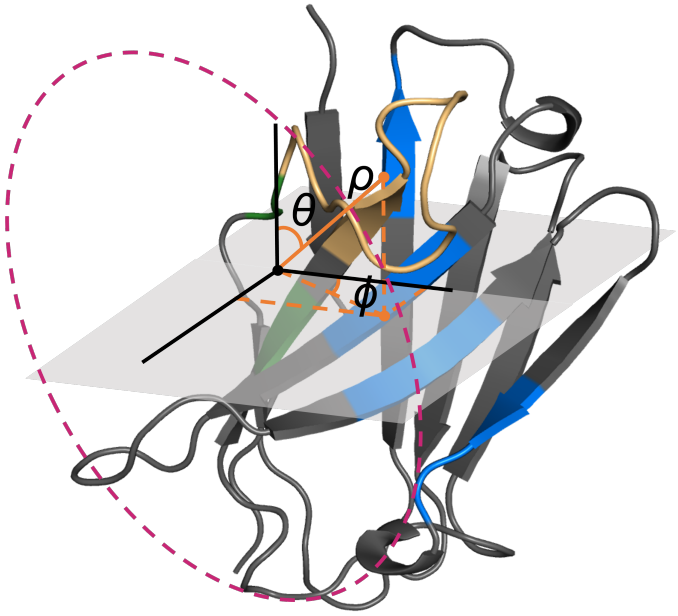


Figure 4.1: Depiction of how CDR-H3 loop orientation and reach is described using spherical coordinates. The V_H domain is shown, with the approximate location and orientation of the V_L domain marked by the dashed pink oval. The CDR-H3 loop is shown in wheat, with IMGT positions 102, 103, 118, and 119 (whose centre of geometry defines **c**) shown in green. CDR-H1/2 anchor residues are shown in blue.

4.3.2 Analysis of V_L domain influences on CDR-H3 conformations

4.3.2.1 Crystal data set curation

We mined SAbDab (Dunbar et al. 2014) on 09.08.2022 and filtered based on a 95% sequence identity threshold across the F_V region in Abs. We further filtered this data set to exclude structures that did not have paired heavy and light chains, single-chain F_V formats, structures not solved using X-ray crystallography, and structures with a resolution greater than 3Å. Lastly, we filtered the data set to remove structures with unresolved loops. Only the first F_V in the asymmetric unit was included in the final set. This procedure yielded 2357 non-redundant F_V structures.

4.3.2.2 Region definitions

ANARCI (Dunbar et al. 2015) was used to number sequences using the IMGT numbering scheme (Lefranc et al. 2003). We chose this numbering scheme to enable comparison across the V_H and V_L domains (CDR1: IMGT residues 27-38, CDR2: IMGT residues 56-65, CDR3: IMGT residues 105-117, Framework: IMGT residues 1-128 and not within CDR definitions).

4.3.2.3 Crystal radial distribution functions

Radial distribution functions (RDFs) for each loop pair were calculated by using a modified version of the mdtraj (McGibbon et al. 2014) implementation that does not normalise by unit cell volume. Namely, we generated histograms of pairwise heavy atom distances (up to 10 Å) using a bin width of 0.1 Å. We then normalised the densities by dividing the counts at each bin by the volume of the spherical shell located at that bin. When compared to RDFs generated from simulation trajectories, this results in comparable trends but not absolute frequencies, since the calculation also involves normalisation based on the box volume for simulation data, though this is a roughly constant value.

4.3.2.4 Molecular dynamics simulations

The simulated systems were the first Fabs in the asymmetric units of PDB entries 3GJF, 4DKF, 1E6J, 3L5X, 2CMR, 4ZS7, 3L5W, and 5BK1. We chose this set of antibodies as they provided a representative sample in terms of inter-CDR loop distances, a range of CDR loop lengths (CDR-H1: 8-10 residues, CDR-H2: 7-8 residues, CDR-H3: 11-19 residues, CDR-L1: 5-9 residues, CDR-L2: 3 residues, CDR-L3: 8-11 residues), and sequence diversity. They also span a diverse set of IGHV and IG(K/L)V genes.

Simulation protocol All systems were prepared and simulations performed using OpenMM v7.7 (Eastman et al. 2016). We capped C-termini using N-methyl groups. Next, we protonated the models at a pH of 7.5 (protonation states were assigned using the Modeller module of OpenMM (Eastman et al. 2016)), soaked them in truncated octahedral water boxes with a padding distance of 1 nm, and added sodium or chloride counterions to neutralise charges and then NaCl to an ionic strength of 150 mM. We parameterised the systems using the Amber14-SB force field (Maier et al. 2015) and modelled water molecules using the TIP3P-FB model (Jorgensen et al. 1983). Non-bonded interactions were calculated using the particle mesh Ewald method (Darden et al. 1993) using a cut-off of distance of 0.9 nm, with an error tolerance of 0.0005. Water molecules and heavy atom-hydrogen bonds were rigidified using the SETTLE (Miyamoto et al. 1992) and SHAKE (Ryckaert et al. 1977) algorithms, respectively. We used hydrogen mass repartitioning (Hopkins et al. 2015) to allow for 4 fs time steps. Simulations were run using the mixed-precision CUDA platform in OpenMM using the Middle Langevin Integrator with a friction coefficient of 1 ps^{-1} and the Monte-Carlo Barostat set to 1 atm. We equilibrated systems using a multi-step protocol: (i) energy minimisation over max. 10,000 steps, (ii) heating of the NVT ensembles from 100 K to 300 K over 200 ps, (iii) 200 ps simulation of the NPT ensembles at 300 K, (iv) cooling of the NVT ensembles from 300 K to 100 K over 200 ps, (v) energy minimisation over 10,000 steps, (vi) heating of the NVT ensembles from 100 K to 300 K over 200 ps, and (vii) 5 ns

simulation of the NPT ensembles at 300 K. We then initialised well-tempered metadynamics (Barducci et al. 2008) using a bias height of 10 kJ mol^{-1} with width 0.3 rad and a bias factor of 10, depositing biases every 500 steps. We biased a linear combination of sine values of CDR-H3 and CDR-L3 ψ angles in a manner similar to Fernández-Quintero *et al.* (Fernández-Quintero et al. 2019b), resulting in two collective variables. We performed triplicate simulations of 1 µs duration and assessed convergence using block analysis of free energy estimates. To recover unbiased ensembles, we calculated weights (w) for each configuration (s) using the time-independent reweighting scheme (Branduardi et al. 2012):

$$w(s) \propto \exp\left(\frac{V(s)}{k_B T}\right) \tag{4.10}$$

where $V(s)$ is the bias at configuration s and $k_B T$ is the product of the Boltzmann constant and temperature.

Analysis of simulation data Analysis of MD trajectories was performed using mdtraj v1.9.6 (McGibbon et al. 2014) and scikit-learn v1.0.2 (Pedregosa et al. 2011). Minimum heavy atom distances of crystal data and MD trajectories were calculated using mdtraj. MD frames were clustered using average linkage hierarchical clustering (cutoff=1.25 Å) based on CDR RMSD after alignment on framework regions excluding 2 residues on either N-terminus. We chose to exclude these residues as they showed considerably higher flexibility than the rest of the framework regions. To calculate RMSD and RMSF values (defined as the time average of RMSD), we used the positions at the beginning of the simulation as the reference for RMSDs ($\mathbf{x}_{i,0}$) and the average positions for RMSF calculations ($\langle \mathbf{x}_i \rangle$):

$$\mathrm{RMSD}(\mathbf{x}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\mathbf{x}_i - \mathbf{x}_{i,0}|^2} \tag{4.11}$$

$$\mathrm{RMSF}(\mathbf{x}) = \left\langle \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i - \langle \mathbf{x}_i \rangle)^2} \right\rangle \tag{4.12}$$

4.3.2.5 Dynamic time warping analysis

To compare CDR-H3 loops of differing lengths, we used dynamic time warping (DTW) (Nowak et al. 2016; Muller 2007) after superposing CDR-H3 loops on anchoring residues, defined as IMGT positions 100-104 and 118-122. DTW is analogous to the Needleman-Wunsch algorithm (Needleman et al. 1970), using dynamic programming to find the optimal path in a backbone atom RMSD matrix. When two loops of the same length are compared, the algorithm reduces down to a conventional RMSD calculation. When the two loops are of differing length, the optimally aligning residues are included in the comparison. To obtain the optimal alignment for two loop backbones, $\mathbf{A} = \{(x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_N, y_N, z_N)\}$ and $\mathbf{B} = \{(x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_M, y_M, z_M)\}$, we generate a cost matrix $\mathbf{C} \in \mathbb{R}^{N \times M}$ where $\mathbf{C}_{i,j}$ is the Euclidean distance between $a_i \in \mathbf{A}$ and $b_j \in \mathbf{B}$. We aim to find the optimum path $\mathbf{P}$ through an accumulated cost matrix $\mathbf{C}^A$, starting at $\mathbf{C}^A_{1,1}$ and ending at $\mathbf{C}^A_{N,M}$. $\mathbf{P}$ and $\mathbf{C}^A$ are calculated using Algorithm 1.

The distance of the two loops based on the optimal path is then given by the average cost of traversing $\mathbf{C}$ via $\mathbf{P}$:

$$d = \frac{\mathbf{C}^A_{N,M}}{L} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{C}_{q_i} \tag{4.13}$$

where $q_i \in \mathbf{P}$ and L is the length of $\mathbf{P}$. An example of DTW being used to compare two CDR loops of differing lengths is given in Figure B.1.

We used DTW on our crystal data set to generate a pairwise comparison matrix of all CDR-H3 loops. We then stratified the data set based on antibody subtype, F_V genes, and minor loop canonical forms (calculated using SCALOP (Wong et al. 2018)) and compared the average distance in the stratified data to the overall average distance. To test for significance, we used a bootstrapping procedure to randomly sample as many points from the whole data set as there were in each stratum, repeated the comparison 10,000 times, and calculated a p-value based on how many times an effect of greater size was observed.

Algorithm 1 Dynamic Time Warping

Require: $\mathbf{C} \in \mathbb{R}^{N \times M}$
  $\mathbf{C}^A \leftarrow \mathbb{R}^{N \times M}$ ▷ Initialise empty accumulated cost matrix
  while $n \leq N$ **do** ▷ Fill first column of $\mathbf{C}^A$
    $\mathbf{C}^A_{n,1} \leftarrow \sum_{k=1}^{n} \mathbf{C}_{k,1}$
  end while
  while $m \leq M$ **do** ▷ Fill first row of $\mathbf{C}^A$
    $\mathbf{C}^A_{1,m} \leftarrow \sum_{k=1}^{m} \mathbf{C}_{1,k}$
  end while
  for $n = 2, ..., N$ **do** ▷ Fill the rest of $\mathbf{C}^A$
    for $m = 2, .., $ **do**
      $\mathbf{C}^A_{n,m} \leftarrow \mathbf{C}_{n,m} + \min\{\mathbf{C}^A_{n-1,m-1}, \mathbf{C}^A_{n-1,m}, \mathbf{C}^A_{n,m-1}\}$
    end for
  end for
  $l \leftarrow 1, q_l \leftarrow (N, M)$
  repeat ▷ Find the optimal path through $\mathbf{C}^A$
    $l \leftarrow l + 1, q_{l-1} \leftarrow (n, m)$
    if $n = 1$ **then**
      $q_l \leftarrow (1, m - 1)$
    else if $m = 1$ **then**
      $q_l \leftarrow (n - 1, 1)$
    else
      $q_l \leftarrow \text{argmin}\{\mathbf{C}^A_{n-1,m-1}, \mathbf{C}^A_{n-1,m}, \mathbf{C}^A_{n,m-1}\}$
    end if
  until $q_l = (1, 1)$
  $L \leftarrow l, \mathbf{P} \leftarrow \{q_L, q_{L-1}, ..., q_1\}$
  return $L, \mathbf{P}, \mathbf{C}^A$

4.3.2.6 V_H-V_L pairing analysis

To search for antibodies with identical V_H but dissimilar V_L sequences, we extracted all V_H sequences from SAbDab (Dunbar et al. 2014) on 10.12.2022 where the structure was solved using X-ray crystallography and retained only entries where F_V backbones were resolved completely. We clustered these antibodies on V_H sequences using global alignment at a clustering threshold of 1.0 using cd-hit (Li et al. 2006). We then grouped resultant non-singleton antibodies together and clustered them again, this time using a threshold of 0.4 and clustering on the V_L sequences. We extracted pairs from within each subcluster if the V_L sequence identity was below 99%. This yielded a set of 86 antibody pairs. Every pair obtained contained at least one different residue in the V_L CDR loops without our method specifically

selecting for pairs with CDR loop sequence identity. For each pair, we aligned the V_H structures on the framework regions using Biopython v1.78 (Hamelryck et al. 2003; Cock et al. 2009) and calculated backbone RMSD values. Lastly, we calculated Pearson correlation coefficients using scipy v1.8.1 (Virtanen et al. 2019).

4.4 Results

4.4.1 Single-domain antibody CDR-H3 loops show different conformations from paired antibody CDR-H3 loops

We started by asking whether the presence of the V_L domain in antibodies (Abs) results in CDR-H3 loops adopting different conformations when compared to single-domain antibodies/nanobodies (sdAbs). Our analyses show that Ab and sdAb CDR-H3 loops show distinct differences in their orientation relative to the rest of the V_H domain. Using the described spherical coordinates, ρ describes the reach of the CDR-H3 loop away from the rest of the V_H domain. A CDR-H3 loop in an extended conformation will have a high ρ value, whereas a loop of identical length that is folded against the V_H domain will have a lower ρ value. ϕ gives an indication of whether the CDR-H3 loop is horizontally oriented towards the rest of the V_H domain or away from it. In the case of Ab structures, a high ϕ value indicates packing against the V_L domain. θ gives a measure of the elevation of the loop. A low value corresponds to a CDR-H3 that extends directly up and away from the rest of the V_H domain, whereas a high value indicates that the loop is "folding" down. In the case of Ab structures, a high θ value corresponds to a loop that is packed into the groove created by the V_H -V_L interface. Lastly, we divide the loop length by ρ to give a measure of compaction. A loop with low compactness uses its entire length to reach away from the V_H domain, whereas high compactness corresponds to a loop that is packed against the V_H .

A comparison of Ab and sdAb structures using these measures (Fig 4.2) shows that both types of antibodies have similar reach, as shown by the near-identical distributions of ρ values (Fig 4.2a). This indicates that sdAbs are not necessarily able to provide extended paratopes via their CDR-H3 loops compared to Abs. In

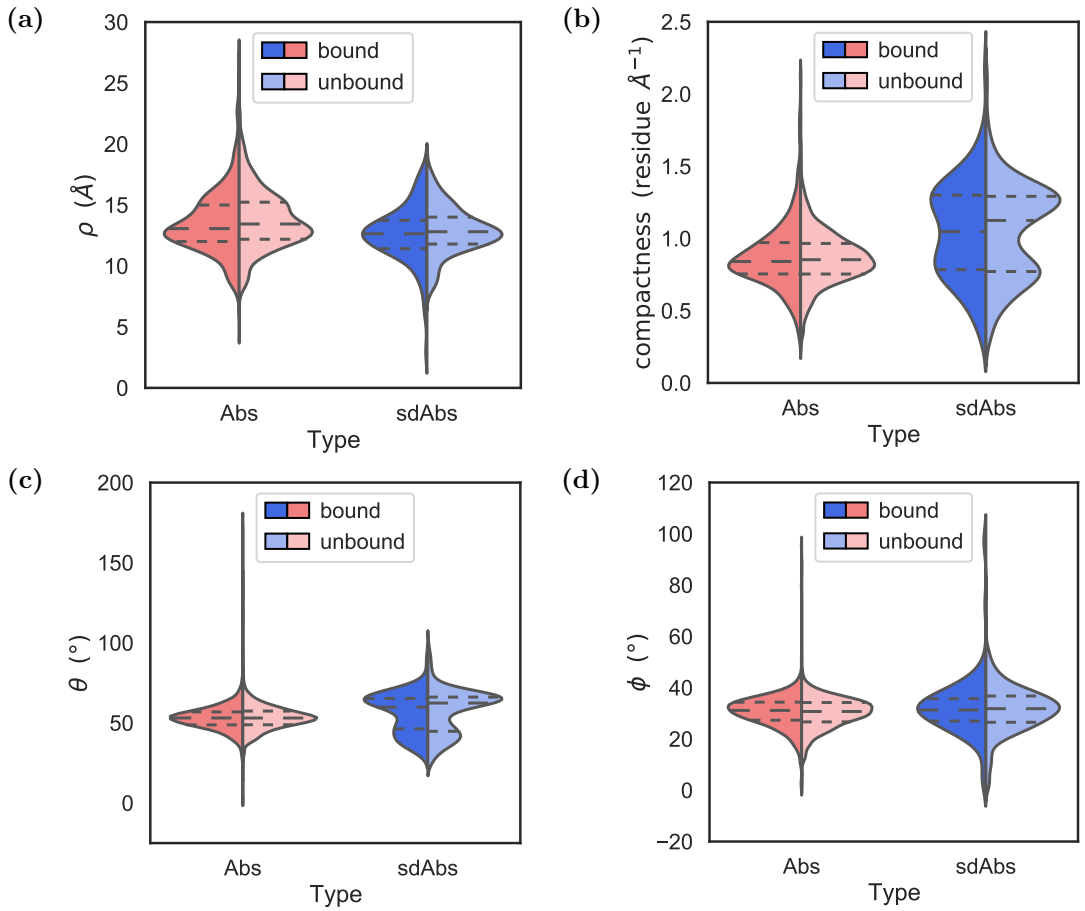


Figure 4.2: Distributions of spherical coordinate descriptors of CDR-H3 conformations, comparing sdAbs and Abs. **(a)** shows the average reach of the loop and describes how far the loop extends out from its anchors, ρ, **(b)** shows the compactness of the loop, **(c)** shows *theta* and describes the downwards orientation of the loop, while **(d)** shows ϕ and describes the "horizontal" orientation of the loop away from the V_H domain. Dashed lines on each violin plot refer to the 75^{th}percentile, the median, and the 25^{th}percentile, from top to bottom.

fact, we find a shoulder in the distribution of ρ values for Abs above the median value, suggesting that Abs are more likely to be able to target deep epitopes that require long reach. This also shows that the two types of antibodies have similar reach on average, even though sdAb CDR-H3 loops tend to be longer (Gordon et al. 2023) than their Ab counterparts, suggesting differences in how the length of these loops is organised around the V_H domain.

The observation that sdAb CDR-H3 loops tend to be longer than their Ab counterparts while still having the same reach is explained through loop compaction. Indeed, we found that sdAb CDR-H3 loops are much more compacted than Ab loops

on average (Fig 4.2b). The distribution of compactness scores for sdAbs is bimodal, with the first peak corresponding to the distribution found in Abs. This suggests the existence of two types of sdAb CDR-H3 loops: one population that behaves similarly to Ab CDR-H3 loops and one population that is more significantly folded up against the V_H domain[1]. A scatter plot of CDR-H3 length against ρ confirms this to be the case (Fig 4.3a). sdAbs can either increase their reach with CDR-H3 length with a rate similar to Abs or keep their loops in a more heavily compacted state.

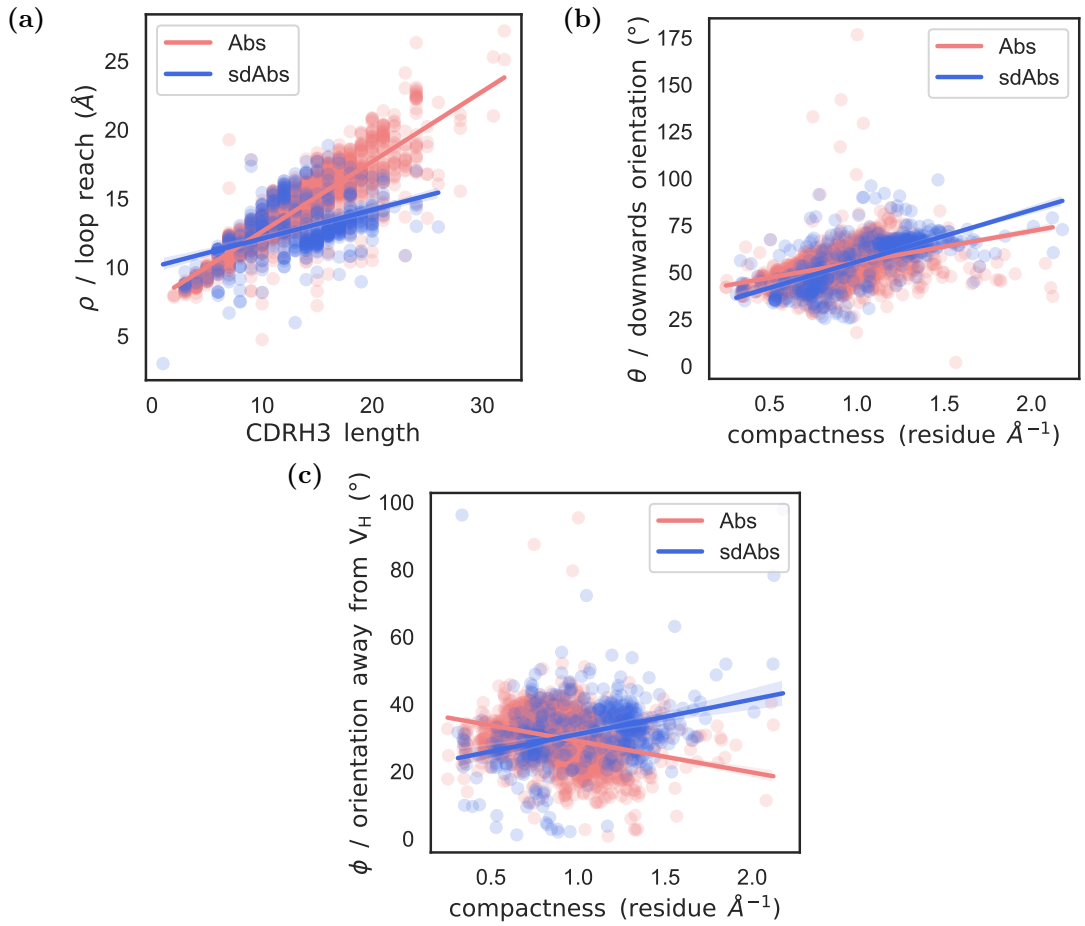


Figure 4.3: Relationships between spherical coordinate descriptors of CDR-H3 conformations, comparing sdAbs and Abs. **(a)** shows how the average reach of a loop, ρ, increases as loop length increases, **(b)** shows the relationship between compactness and θ, the downwards orientation of the loop, while **(c)** shows the relationship between compactness and ϕ, the "horizontal" orientation of the loop away from the V_H domain. Shaded regions around regression lines show the 95% confidence interval for these regressions.

Next, we compared the orientation of the CDR-H3 loops of Abs and sdAbs

[1]We presume that this may be representative of kinked and extended sdAb CDR-H3 loops (Dizicheh et al. 2023).

and examined the differences in both the azimuth and elevation angles. Compared to Ab CDR-H3 structures, sbAbs show a much wider bimodal distribution of θ values (Fig 4.2c), with the major peak corresponding to θ values in excess of those observed for Ab structures and another minor peak below the Ab distribution. This indicates that the majority of sdAb CDR-H3 loops do not extend upwards away from the V_H domain, but rather lie flat against it, therefore folding down. We observe a slight shift in θ in the distribution for bound sdAbs, but note that the position of the peaks still remains stable. We conclude that sdAbs generally do not extend their CDR-H3 loops upon binding, as has previously been hypothesised. Lastly, we show near-identical values of ϕ for sdAbs and Abs with sdAb ϕ values having a slightly wider distribution (Fig 4.2d).

In order to examine how CDR-H3 loops pack against the V_H and/or V_L domains, we analysed the relationship between the spherical angles and compactness (Fig 4.3). We show that both sdAb and Ab CDR-H3 loops become more compacted through an increase in θ (Fig 4.3b). That is, packing of the loop down towards the rest of the V_H domain decreases its reach. We hypothesise that this is a mechanism to stabilise the paratope structure by allowing the loop to pack against the rest of the V_H domain. On the other hand, we find inverse relationships between compactness and ϕ for sdAbs and Abs. As ϕ increases (that is, as the CDR-H3 loop is horizontally oriented away from the V_H domain), sdAbs show an increase in compactness whereas the opposite is true for Abs (Fig 4.3c). In the former case, an increase in ϕ results in the loop extending away into empty space, whereas in Abs this corresponds to the loop being positioned towards the V_L domain. As the presence of the V_L domain provides steric hindrance, the CDR-H3 loop is forced into a conformation that orients it away from the Ab, therefore reducing compactness and increasing reach.

Taken together, these results suggest that the lack of the V_L domain in sdAbs may result in changes of the CDR-H3 loop conformation. We argue that the presence of the V_L domain in Ab structures provides a surface for the CDR-H3 loop to pack against, thereby providing the necessary stability for an extended conformation. In sdAbs, these interactions are missing and therefore CDR-H3 loops

exhibit different packing patterns. Further, this may explain the observation that sdAbs and Abs generally target paratopes of similar shapes and accessibilities even though sdAbs tend to have longer CDR-H3 loops (Gordon et al. 2023), as the added length of these loops in sdAbs may simply be a compensatory mechanism that makes up for the lack of the V_L domain.

4.4.2 Antibody V_H and V_L CDR loops are characterised by non-equivalent inter-loop contacts in crystal structures

Our comparison of CDR-H3 loop structures in Abs and sdAbs demonstrates clear differences between the two types of antibodies, although this might also be an effect of CDR-H3 sequence differences, as different genetic lineages are used to produce each type. To investigate this, we aimed to identify whether differences in CDR-H3 conformations within paired antibodies alone can be due to differences in the V_L domain. We began by analysing the structural relationships between antibody CDR loops by examining a data set of antibody crystal structures. As described in Section 4.3.2, we mined SAbDab for a set of antibodies. This yielded a set of 2357 non-redundant structures for which all CDR loops are resolved. We then examined the relative packing of loops against each other by calculating radial distribution functions (RDF). For each loop, we selected heavy atoms and tallied the number of heavy atoms belonging to other loops at a given distance interval and normalised this value by dividing by the number of possible pairs. Our results are generally in line with what would be expected: loops that are close to each other in the three-dimensional antibody structure tend to make more contacts with each other. These trends are largely mirrored for the V_H and V_L loops, owing to the symmetry that marks antibody structures (Fig 1.3).

We found that loops sandwiched between other loops make contacts on both sides. The CDR-1 loops make contacts with the CDR-2 and CDR-3 loops of the same chain, and the CDR-2 loops, owing to their positions on the edges of the V_H -V_L complex, only make contacts with their neighbouring CDR-1 loops (Fig 4.4). We did, however, find two major differences between the V_H and V_L loops.

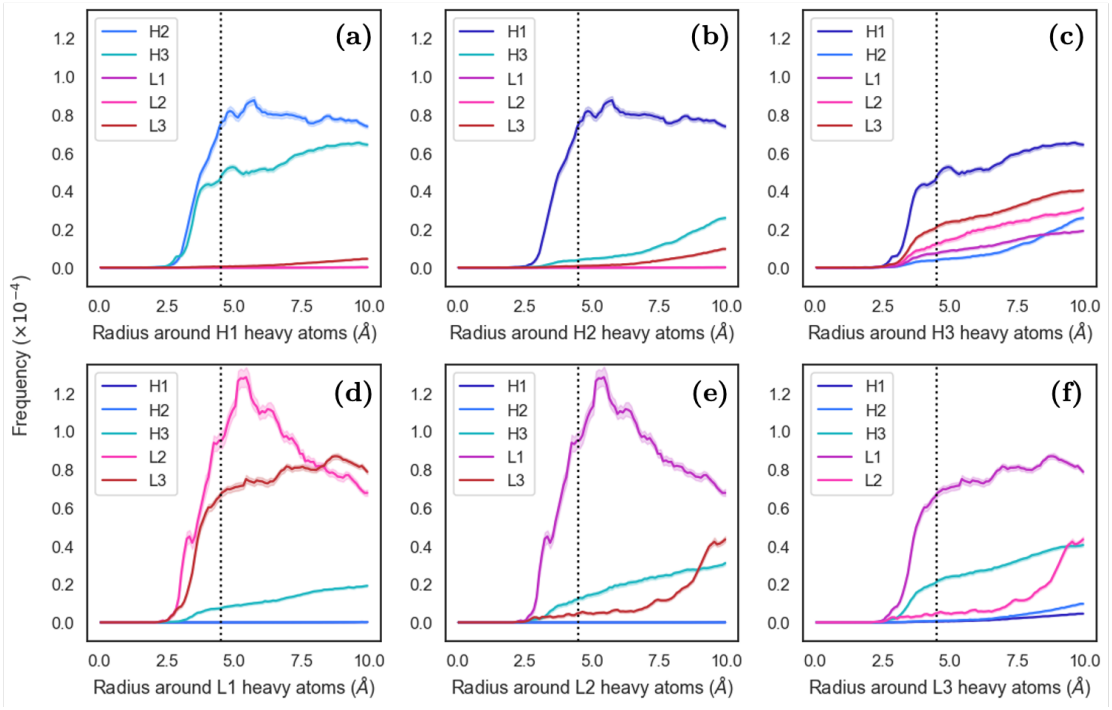


Figure 4.4: Radial distribution functions of CDR loops in the crystal data set. **(a–f)** We generated histograms for all pairwise CDR loop heavy atom distances using a bin width of 0.1Å, where **(a–f)** correspond to CDR-H1, -H2, -H3, -L1, -L2, and -L3, respectively. Each bin was then normalized by dividing by the product of the volume of a spherical shell centered at 0 and located on the edges of the bin and the number of possible pairs of heavy atoms. Shaded areas are 95% confidence intervals.

Firstly, the CDR-L1/2 loops show tighter packing than the corresponding CDR-H1/2 loops. The interactions between CDR-L1 and CDR-L2 show a larger peak compared to those between CDR-H1 and CDR-H2 (Fig 4.4a, 4.4d). Based on the number of residues, the CDR-L2 loop is considerably shorter ($\mu = 3.05$ residues, $\sigma = 0.45$) than the CDR-H2 ($\mu = 7.90$ residues, $\sigma = 0.74$) loop, while the CDR-L1 ($\mu = 7.60$ residues, $\sigma = 2.17$) and CDR-H1 ($\mu = 8.19$ residues, $\sigma = 0.69$) loops are more comparable in their lengths. This suggests tighter packing of the CDR-L2 loop against the CDR-L1 loop than is observed with CDR-H2 and CDR-H1 based on the RDF trace, despite the shorter average length of CDR-L2.

Secondly, the pattern of contacts made by the CDR-H3 loops differs considerably from that of the CDR-L3 loop. The CDR-H3 loop makes significant contacts with CDR-H1, CDR-L3, and CDR-L2 loops, with small contributions from the CDR-H1 loop (Fig 4.4c), whereas the CDR-L3 loop appears to make contacts mainly with the

CDR-L1 and CDR-H3 loops, with a small number of contacts made with CDR-L2 (Fig 4.4f). While CDR-H3 packs less densely against CDR-H1 than CDR-L3 with CDR-L1, it packs more densely against all other loops than CDR-L3 does against its analogous loops. This is in line with the larger structural variability of CDR-H3. CDR-H3 loops are known to adopt multiple different conformations and can pack against minor loops in different ways, which is thought largely to be due to their length and sequence diversity as a result of V(D)J-recombination (Regep et al. 2017). Conversely, owing to the limited conformational space afforded by the canonical forms and limited sequence space, CDR-L3 loops are more likely to pack against the same loops, in the same way (Wong et al. 2019).

We also stratified our data set on light chain subtype. We did not observe any significant differences in RDF densities in κ/λ antibodies (Fig B.2), showing that the density of loop packing is the same in both V_H and V_L loops, irrespective of the light chain subtype.

Taken together, our results suggest that the V_L loops appear to pack more tightly against each other than the V_H loops and that the CDR-H3 loop is differentiated from its light chain counterpart by a much broader set of configurations with which it packs against the minor loops. We also show that the CDR loops of the V_L domain are more densely packed than those in the V_H domain, thereby likely generating a more rigid structure on one side of the F_V .

4.4.3 Conformational rearrangements in CDR-H3 do not require major conformational changes in minor loops

In order to test the observations made on crystal structures in a dynamic context, we created a set of MD simulation data. Briefly, we performed metadynamics simulations using eight antibody structures (simulation protocol details are provided in Section 4.3.2.4). Each simulation was allowed to proceed for 1 µs and simulations were performed in triplicate. To recover equilibrium ensembles, we re-weighted

our trajectories. The simulations were biased on both the CDR-H3 and CDR-L3 loops and therefore extensive sampling of conformational space is expected for these loops only.

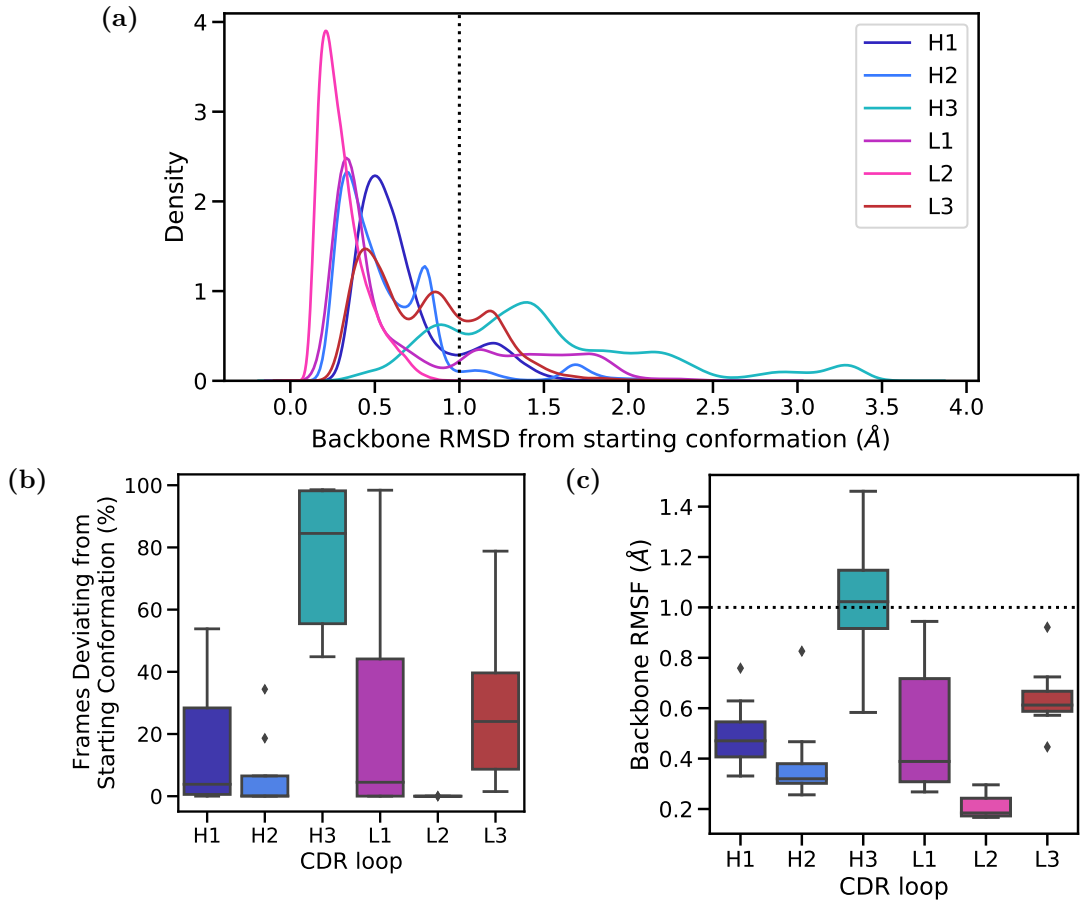


Figure 4.5: Flexibility of CDR loops in metadynamics simulations. All loops were individually aligned on their anchor residues, defined as the three residues on either side of the loop. **(a)** shows pooled RMSD values for each loop relative to the crystal structure (the histograms were smoothened using a kernel density estimation), **(b)** shows the fraction frames where loops were found in their starting conformations for each antibody, defined as being within 1 Å of the crystal conformation, and **(c)** shows the RMSF of loops for each structure, defined as the time average of the RMSD.

In the simulations, the minor loops (all loops apart from CDR-H3) generally showed little flexibility. In order to quantify this, we aligned backbone atoms of each loop on three flanking residues on each side and calculated the RMSD and RMSF values relative to the starting structure. The CDR-H3 loop only appears in the starting conformation in a fraction of MD frames, with deviations ranging up to 3 Å (Fig 4.5a). The peak that is within the 1 Å cutoff used to define antibody

loop conformational differences (Wong et al. 2019) is shifted further to the right for CDR-H3 compared to the major peaks found for other CDR loops (especially when compared to CDR-L3, as this loop was also biased in the simulations) (Fig 4.5b), suggesting that even in its native conformation, the CDR-H3 loop is more flexible and poised for conformational shifts. In line with previous studies (Fernández-Quintero et al. 2019b; Fernández-Quintero et al. 2020c; Fernández-Quintero et al. 2020b), we therefore conclude that the CDR-H3 loop is distinctly flexible. In the context of antigen binding, where CDR-H3 is highly dominant, this might be an adaptation to allow a multitude of loop conformations in order to generate surfaces complementary to the vast range of possible epitope topologies. In our simulated systems, the least flexible CDR-H3 loop is found in non-crystal conformations in roughly 45% of frames, whereas the most flexible loop immediately adopts a new conformation without revisiting the native form during the simulation (Fig 4.5a).

Conversely, the minor loops do not show similar levels of flexibility (Fig 4.5a). In all cases, the major peak of the distribution is within the 1 Å cutoff and lies around 0.5 Å. As is expected based on its length and observed conformational diversity in crystal structures, CDR-L2 is the least flexible loop, with no instances of a conformational shift being observed. The CDR-L3 loop, although biased using the same protocol as CDR-H3, is relatively rigid in its conformation. We hypothesise that this is due to a combination of factors including the length of the loop, the conformational space allowed by CDR-L3 sequences, and the tighter packing of CDR loops in the V_L domain. It has been shown that CDR-L3 loops tend to be shorter than CDR-H3 loops on average (Marks et al. 2017a; Wong et al. 2019). Further, a number of studies have shown that CDR-L3 loops are both less conformationally diverse and less flexible than CDR-H3 loops (Wong et al. 2019; Regep et al. 2017; Weitzner et al. 2015; Fernández-Quintero et al. 2020a). The former is presumably due to the fact that CDR-L3 loops are generated only by sequence contributions from V and J genes, whereas the D gene also contributes to CDR-H3 sequence, thereby likely increasing the conformational space by virtue

of greater sequence space. The presence of a second junction during the CDR-H3 recombination event also leads to greater sequence diversity by virtue of an additional site for nontemplated nucleotide addition (Alt et al. 1982). The reduced flexibility of CDR-L3 compared to CDR-H3, on the other hand, is likely to be driven by a number of factors including loop length, loop sequence, and the presence or lack of stabilising lateral contacts. In line with this, a much smaller fraction of collected frames of CDR-L3 are found in non-crystal conformations (Fig 4.5b), though there is a significant range, with the most flexible CDR-L3 loop being found in a different conformation in 80% of the trajectories. The CDR-L1 loop shows a much broader distribution in Figure 4.5b, especially when compared to CDR-H1, suggesting the presence of highly flexible CDR-L1 loops. However, examination of the underlying distribution shows that, on average, the CDR-L1 loop is found in a conformation much closer to its native pose compared to CDR-H1, as demonstrated by the shift of the major peak towards lower RMSD compared to CDR-H1 (Fig 4.5a).

To further quantify the observed relative flexibility of the CDR loops, we calculated RMSF values. RMSF values are the time average of RMSD relative to the starting conformation and provide a measure of flexibility. We found that in all cases, the V_L CDR loops show decreased flexibility relative to their V_H counterparts. However, our biasing scheme was highly aggressive, with biases roughly 10 times the values usually used in metadynamics simulations. This provides a high energetic incentive for conformational space exploration of the biased loops. The lack of significant conformational changes in the minor loops despite this approach suggests that conformational changes in the CDR-H3 loop are not sufficient to induce conformational changes in the minor loops. We hypothesise that the minor loops form a scaffold around the CDR-H3 loop, which then explores feasible conformations available to it based on the available space within this scaffold, as evidenced by the fact that CDR-H3 is the only loop exhibiting RMSF values beyond the 1 Å cutoff for conformational change (Fig 4.5c).

Our results do not indicate that the minor loops are unable to change conformation, a phenomenon that has been well documented in the literature (Fernández-

Quintero et al. 2019b; Fernández-Quintero et al. 2020a). Instead we discuss only the flexibility of the CDR-H3 loop when the minor loops are mainly in their crystal pose conformations. These conformations are thought to be the lowest energy shape for these loops (Fernández-Quintero et al. 2019b; Fernández-Quintero et al. 2020a). Our results show that, despite significant packing of the CDR-H3 loop against the minor loops, the CDR-H3 loop retains flexibility, exploring non-crystal conformations more than 40% of the time, even in the least flexible antibody in our simulation data set. Further, we show that this movement is possible without significant conformational rearrangement in the minor loops. Taken together, our results point to at least part of the conformational ensemble of CDR-H3 loops being accessible without any movement in the minor loops.

4.4.4 Minor CDR loops create steric boundaries within which CDR-H3 samples conformational space

Based on our observation that CDR-H3 explores the conformational landscape in the simulation data without concurrent movement in the minor loops, we examined whether the observed packing of CDR loops in crystal structures (as shown using the RDFs in Fig 4.4) changes with CDR-H3 conformation. To this end, we calculated RDFs for CDR loops (Fig B.3) and inter-loop minimum distances on our simulation data set (Fig 4.6a-j). The RDFs show that the patterns observed in the crystal data set RDFs are conserved in simulations, with CDR-H3 making a wider range of contacts than CDR-L3 (Fig B.3c, B.3f) and the V_L loops showing greater inter-loop packing than their V_H counterparts. This is supported by our minimum distance calculations, where in the vast majority of pairs, no significant change in the average minimum distance is observed (Fig 4.6a-c,e,f,h,j).

The RDFs for CDR-3 loops showed minor differences in interactions with CDR-2 loops (Fig 4.7). In the simulation data, we did not observe the presence of CDR-L2-CDR-H3 interaction that was found in the crystal data set (Fig 4.7c 4.7d). Indeed, an examination of the minimum distance distribution for this pair (Fig 4.6d) shows an average increase of 2.47 Å, with the majority of the distribution shifting

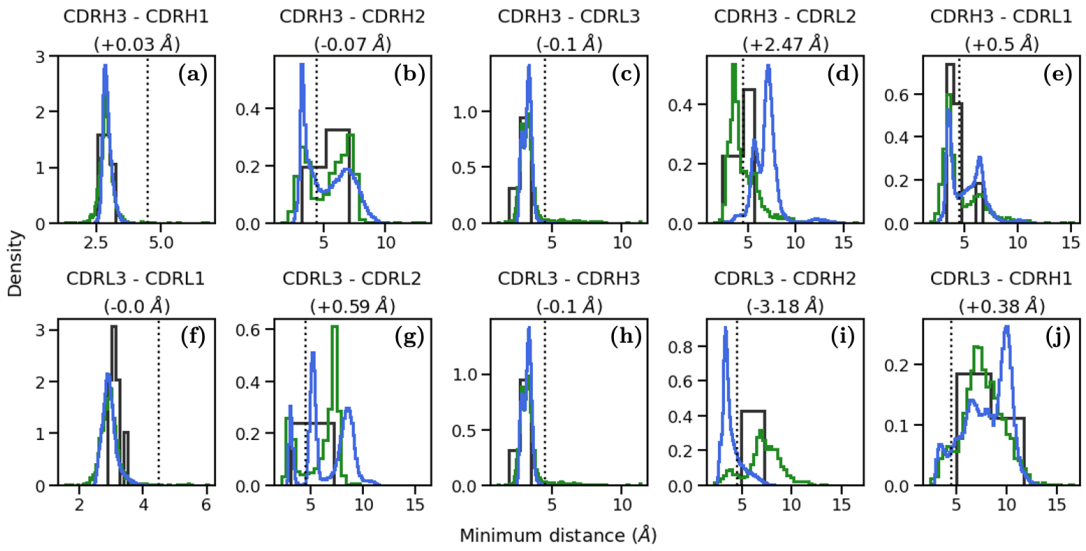


Figure 4.6: Conformations explored by CDR-H3 through CDR loop contacts. **(a-j)** show minimum heavy atom distances between pairs of loops. Green histograms refer to all crystal structures, blue histograms refer to simulation data, and black histograms refer to crystal structures of systems used to perform simulations.

out of the 0-4.5 Å heavy atom distance range used to define contacts (Yuan et al. 2012). This confirms the loss of interactions between CDR-L2 and CDR-H3 in the simulations, as shown in Figures 4.8c and 4.8d). This is not always the case, Figure 4.6d exhibits a minor shoulder within the 4.5 Å cutoff and Figures 4.8a and 4.8b confirm that conformations of CDR-H3 that allow interactions with CDR-L2 still exist, especially when the CDR-H3 loop is heavily oriented towards the V_L domain. The observation that the CDR-H3 loop generally pulls away from the CDR-L2 loop during exploration of conformational space indicates that the CDR-H3-CDR-L2 interaction might be important in the stabilisation of the crystal pose of antibodies.

We also identified a deviation from the crystal data set in the case of the CDR-H2-CDR-L3 interaction. In this case, the opposite effect was observed, with the average minimum distance between these loops decreasing by ∼3 Å on average and a concurrent shift in the simulation frame distribution to the minor peak of the crystal data distribution. This is confirmed by the appearance of density in the RDF (Fig 4.7c, 4.7d), showing that, in simulations, CDR-L3 and CDR-H2 close the gap between themselves, thereby providing steric bulk in the V_H /V_L interface. Based on the fact that the minor loops do not undergo any major conformational

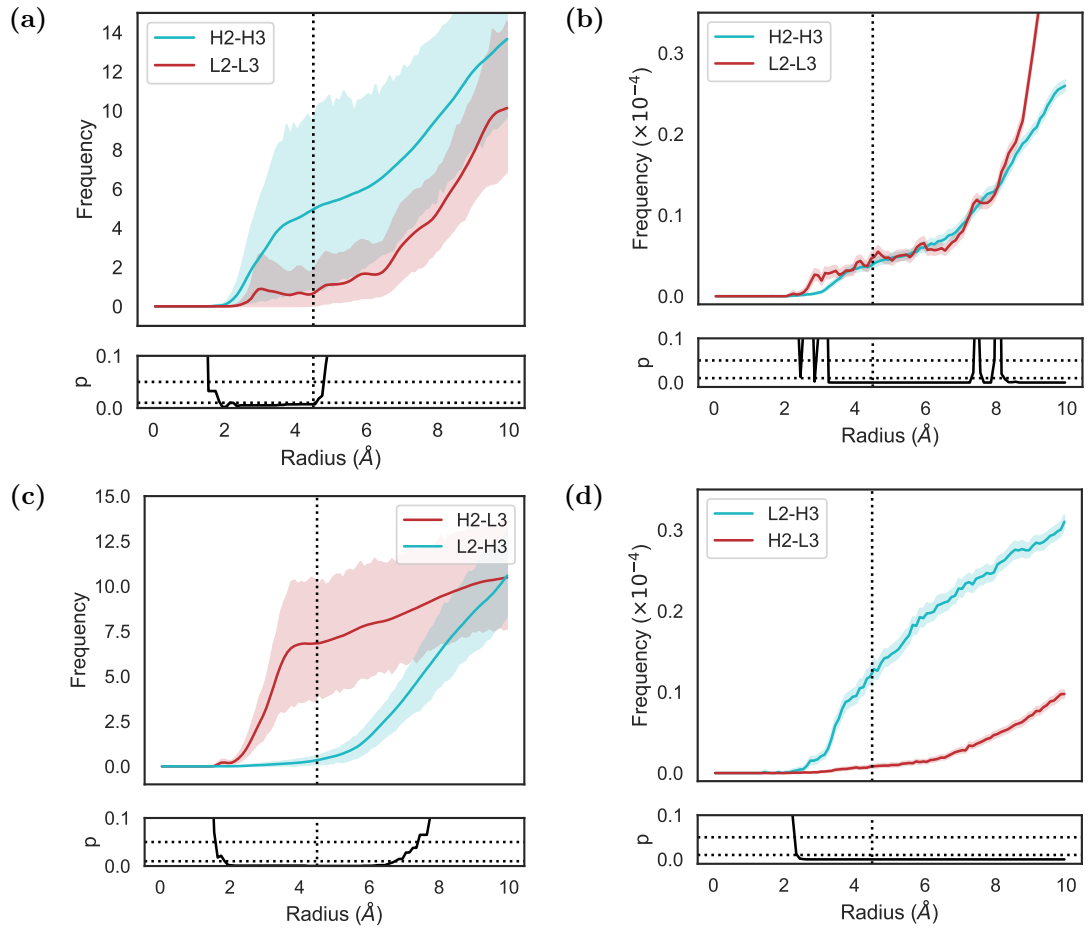


Figure 4.7: Changes in loop heavy atom packing patterns in **(a, c)** simulation and **(b, d)** crystal structure data sets. For simulation data **(a, c)**, we calculated RDFs using the mdtraj implementation with a bin width of 0.1 Å. For crystal structure data **(b, d)**, we used an analogous method without the unit cell correction term; that is, we generated histograms for all pairwise CDR loop heavy atom distances using a bin width of 0.1 Å. Each bin was then normalized by dividing by the product of the volume of a spherical shell centered at 0 and located on the edges of the bin and the number of possible pairs of heavy atoms. *p*-values are reported based on t-tests conducted at each bin. Shaded areas correspond to 95% confidence intervals.

changes in our simulations, this could be explained by independent changes in the framework regions of V_H and V_L domains. We found no correlation between the increase in CDR-H3-CDR-L2 and decrease in CDR-L3-CDR-H2 minimum distances.

The superimposed density maps of representatives from clustered MD trajectories show that the CDR-H3 loop samples conformations to exhaustively fill the space left by the minor loops (Fig 4.8). This suggests that out of all possible conformations given the CDR-H3 sequence and anchor positions, the loop explores only those that

are allowed within the steric mass of the minor loops. This is analogous to a "cage" formed around CDR-H3 by these loops, within which the CDR-H3 loop is able to adopt new conformations. Thus, the conformation of the other loops fine-tunes the conformational landscape of CDR-H3, influencing the potential orientations of the paratope residues provided by the CDR-H3 loop. The CDR-L3 loop, being the main V_L loop that CDR-H3 packs against (Fig 4.4c, 4.4f) would be a major contributor to this, as the CDR-H3 and CDR-L3 loops are oriented such that they cross each other diagonally in F_V structures. Following from the idea of the CDR-H3 loop being "caged" by the minor loops, we propose that different minor loop canonical forms would lead to increases or decreases in CDR-H3 conformational diversity if a canonical form provides less or more steric bulk, thereby making the cage larger or smaller, respectively. Further, the packing between the CDR-H3 loop and minor loops provides another avenue for the modulation of CDR-H3 conformation through the presence or absence of specific interactions, in line with the observation that affinity maturation often targets residues not directly involved in contacts with the antigen, but rather those that make contact with other loops forming the paratope (Mishra et al. 2018).

Based on our observations that CDR loops are packed against each other and that CDR-H3 explores conformational space available to it within the constraints of the minor loops, we propose a model of a "caged" CDR-H3. This cage is constructed of the minor loops, with the precise orientation of these loops forming the boundaries. The CDR-H3 loop is able to undergo conformational shifts that are permitted by the exact shape and dimensions of the cage.

4.4.5 Antibody subtypes, genes, and CDR canonical forms show preference for CDR-H3 conformations with different levels of diversity

Based on our "caged" CDR-H3 model, we predict that minor loops that lead to more steric bulk around the CDR-H3 loop should also lead to a reduction in conformational diversity of CDR-H3 loops. Conversely, we predict that structures with minor loop attributes that widen the physical space available to the CDR-H3

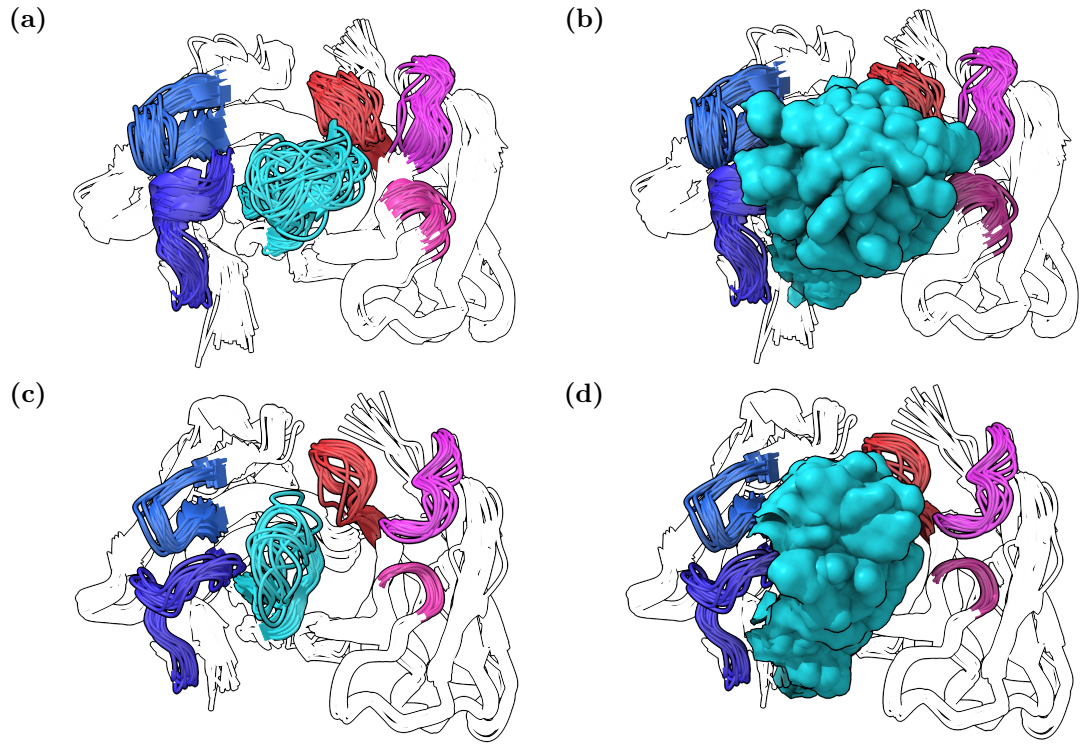


Figure 4.8: Conformations of CDR loops in metadynamics simulations. Show are the conformations of CDR-H3 loops in cluster representatives extracted from simulation data clustered on CDR RMSD after alignment on the framework regions. CDR-H3 and CDR-L3 are shown in cyan and red, respectively. CDR-H1 and CDR-H2 are shown in dark and light blue, respectively. CDR-L1 and CDR-L2 are shown in magenta and pink, respectively. **(a,b)** shows results for 1e6j, whereas **(c,d)** shows results for 3l5x. CDR-H3 loops are visualised as densities in **(b,d)**.

loop should lead to the broadening of conformational diversity. These predictions are in line with our results, which show that CDR-H3 explores conformational space within the boundaries allowed by the minor loops in simulations. Since it is expected that the crystal pose will be part of this conformational ensemble, it follows that the steric bulk provided by minor loops should also influence the conformational diversity of CDR-H3 loops. To test these predictions, we examined whether the subtypes, genes, or minor loop canonical forms have an observable effect on CDR-H3 conformational space in crystal structures using dynamic time warping (DTW). DTW allows for the structural comparison of loops of different lengths and reduces to RMSD when two loops have the same number of residues. This has previously been used to effectively compare antibody CDR loops of different lengths (Nowak et al. 2016) and allows us to embed the CDR-H3 conformations in our data set

on a manifold in a length-independent manner. We calculated the average DTW CDR-H3 distance for the entire data set, and then stratified by subtype, gene, and canonical form before calculating average DTW CDR-H3 distances for each of these subgroups, which we then compared to baseline CDR-H3 diversity. To test for significance, we used a bootstrapping procedure.

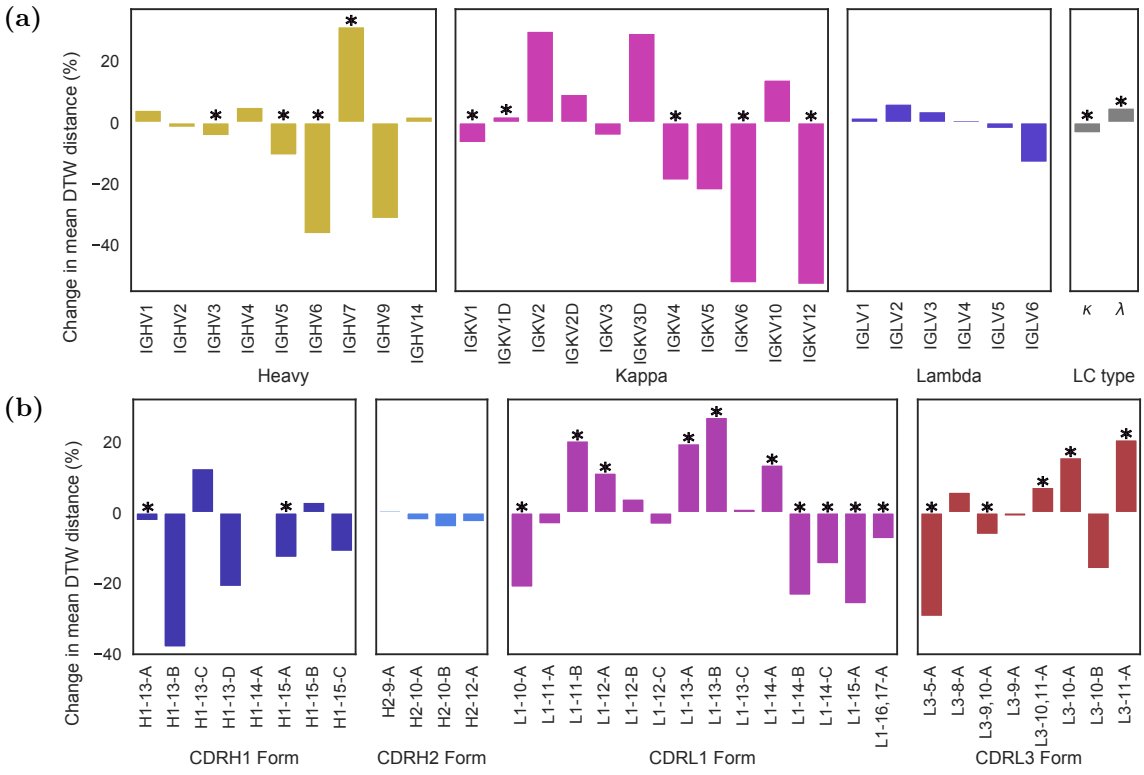


Figure 4.9: Effects of genetically encoded antibody attributes on CDR-H3 conformational diversity. Shown are changes in conformational diversity of CDR-H3 loops compared to baseline after stratification on **(a)** V_H genes, V_L genes, or light chain type, and **(b)** CDR canonical forms based on dynamic time warping distances. Data in panel **(a)** are based on a set of non-redundant human antibody structures whereas data in panel **(b)** are based on a set of non-redundant antibody structures from any species. We aligned CDR-H3 loops on anchoring residues and generated a pairwise distance matrix which we then stratified based on the examined attributes. The difference between the mean diversity of CDR-H3 loops compared to baseline (all CDR-H3 loops in the set) is reported, with * denoting statistical significance based on bootstrapping.

We began by testing for differences based on light chain subtype, and found that κ subtypes lead to a roughly 3.41% reduction in the average DTW distance compared to baseline, whereas λ subtypes lead to a 4.81% increase (Fig 4.9a), leading to an overall difference in CDR-H3 conformational diversity of 8.22% between κ

and λ light chains. This suggests that κ light chains may be more likely to restrict the CDR-H3 loop. In both of these cases, we found the effect to be statistically significant with $p < 0.05$. However, we found no significant differences for inter-loop packing using RDFs between κ and λ antibodies (Fig B.2), suggesting that κ antibodies might be more likely to position minor loops in a way that restricts the CDR-H3 and that λ antibodies provide a larger "cage" in which the CDR-H3 loop may explore conformations. In order to examine whether specific genes drive this observation, we further stratified our data by genes. The general trend found for light chain subtypes was recapitulated in the case of κ genes (Fig 4.9a), but we found that that certain genes have a higher impact on CDR-H3 conformational diversity than others. The effect was most pronounced for IGKV6 and IGKV12, whereas the effect was not statistically significant in the cases of IGKV2/3/5/10 (Fig 4.9a). In the case of IGLV genes, we found no single gene that had a predictable effect on CDRH3 conformational diversity in either direction (Fig 4.9a). We argue that this might be due to a lack of gene-specific "cage" shapes in these antibodies or due to the more limited number of λ antibodies in our data set. As such, the only conclusion we draw for this subclass is that λ antibodies show a significant increase in CDR-H3 conformational diversity when compared to baseline.

Lastly, we examined the effect of IGHV genes. The majority led to a modest reduction in CDR-H3 diversity, with IGHV7 as an exception (Fig 4.9a). This result is expected, as the IGHV gene contributes to the CDR-H3 sequence. A less diverse sequence space for the CDR-H3 loop would lead to a less diverse space of possible conformations for the loop due to residue and position-specific preferences of backbone dihedral angles.

We also examined the effect of specific canonical forms on CDR-H3 diversity. For CDR-H1 and CDR-H2, we observed that CDR-H3 diversity is limited by specific canonical forms, with H1-13-C being the single significant outlier (Fig 4.9b). Since CDR-H1 and CDR-H2 canonical forms are determined by the IGHV gene, which also makes sequence contributions to the CDR-H3, we do not draw conclusions about the effect of CDR-H1-2 canonical forms on CDR-H3 conformational diversity. We

also did not extend our analysis include to CDR-L2, due to its lack of diverse canonical forms.

In the case of CDR-L1 canonical forms, we observe that longer forms (14-17 residues) tend to lead to a reduction in diversity (Fig 4.9b), consistent with our previous observations of loop packing. Longer CDR-L1 loops provide more steric bulk, therefore constraining the CDR-H3 loop. Using our previously outlined framework of the minor loop 'cage' around the CDR-H3, this would lead to a smaller space for the CDR-H3 loop. Conversely, shorter CDR-L1 canonical forms generally lead to an increase in conformational diversity compared to baseline. For CDR-L3, we find the opposite effect: longer CDR-L3 canonical forms tend to increase the diversity of CDR-H3 conformations compared to baseline (Fig 4.9b). We hypothesise that this observation is due to the relative orientation of CDR-H3 and CDR-L3. Usually in antibody crystal structures, CDR-H3 and CDR-L3 cross each other (Fig 1.3, 4.8), with CDR-H3 pointing slightly towards the V_L domain and CDR-L3 pointing towards the V_H domain. Therefore, a shorter CDR-L3 loop would enable a groove to open in the V_H /V_L interface. As CDR-H3 makes significant contacts with CDR-L3 (shown in both our crystal data set and our simulations), this could lead to CDR-H3 packing down towards the rest of the antibody, rather than extending further up, causing the observed decrease in CDR-H3 diversity.

Further, the observed preferences of different canonical forms and CDR-L3 lengths for differing levels of CDR-H3 conformational diversity provide an explanation for the difference observed between κ and λ light chains. It is well-established that human κ antibodies have a preference for shorter CDR-L3 loops compared to their λ counterparts (Rock et al. 1994; Mitsunaga et al. 2020; Townsend et al. 2016). Our results show that longer CDR-L3 loops have a preference for higher CDR-H3 diversity compared to baseline, and we posit that the lower CDR-H3 diversity observed in κ antibodies may be due to the shorter CDR-L3 loops.

These results are consistent with our model of a "caged" CDR-H3 loop, whose conformational diversity partially depends on the minor loops surrounding it,

showing that CDR-H3 structural diversity is significantly modulated by V_L attributes as granular as CDR canonical forms.

4.4.6 Identical V_H sequences show differences in CDR-H3 conformation when paired with different V_L sequences

To confirm our results from the DTW analysis of antibodies that canonical loop forms lead to a change in CDR-H3 conformational diversity, we examined a set of antibody pairs with identical V_H but dissimilar V_L sequences, yielding a data set of 86 pairs of crystal structures, called the pairing data set (see Methods section). Briefly, we aligned each pair on the V_H framework region and calculated RMSD values of the V_H CDRs and framework region before calculating Pearson's correlation coefficients (r) (Fig 4.10). It should be noted that at the 99% V_L sequence identity cutoff which we used (see Methods section), all pairs that we identified had at least one difference in the V_L CDR loops. We found that the V_L sequence identity between these pairs was correlated significantly with CDR-H3 RMSD ($r = -0.49$, $p < 5 \times 10^{-6}$), but not with CDR-H1 ($r = -0.14$, $p = 0.21$), CDR-H2 ($r = -0.13$, $p = 0.22$), or V_H framework ($r = -0.09$, $p = 0.40$) RMSD. We also note one outlier where V_L sequence identity was found to be below 55%. In order to ensure that our observed correlation between sequence identity and CDR-H3 conformation is not driven by this pair, we calculated the correlation coefficient after removing it and found the correlation still to be significant ($r = -0.46$, $p < 5 \times 10^{-}5$).

Next, we examined the canonical forms of these pairs. We separated our data set based on whether all V_L canonical forms remain the same or whether at least one changes and found that changing V_L canonical forms lead to a greater average CDR-H3 RMSD, with an increase of 1.17 Å compared to pairs where V_L canonical forms remain identical. We did not observe effects of similar magnitude for CDR-H1, CDR-H2, or V_H framework regions, with mean RMSD increases of 0.16 Å, 0.28 Å, and 0.08 Å, respectively (Fig 4.11). We found that the correlation between CDR-H3 RMSD and V_L sequence identity disappeared when we examined only pairs where all V_L canonical forms remain the same ($r(56) = -0.11$, $p = 0.43$). This suggests

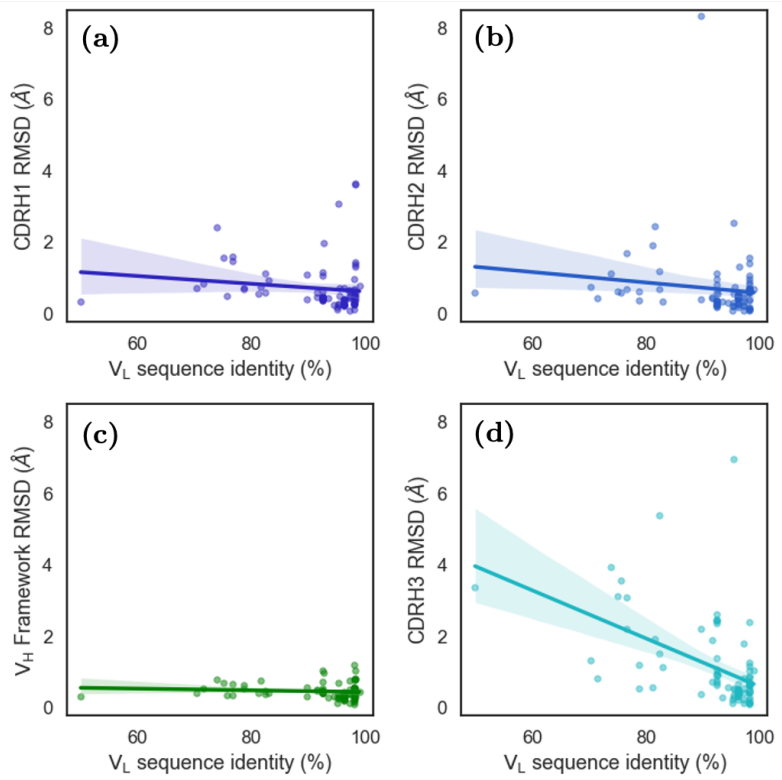


Figure 4.10: Correlations between V_L sequence identity and V_H RMSD. We extracted pairs of structures that have identical V_H but different V_L sequences from SAbDab and aligned structures on the V_H framework regions before calculating RMSD values of **(a)** CDR-H1, **(b)** CDR-H2, **(c)** VH framework, and **(d)** CDR-H3 regions. Lines of best fit are reported with shaded areas corresponding to 95% confidence intervals.

that even small changes in the V_L sequence are able to induce conformational changes in CDR-H3 without accompanying changes in the rest of the V_H domain. Further, we show that this is dependent on the canonical forms of V_L CDR loops. This supports our "caged" CDR-H3 model and provides evidence for the influence of V_L CDR loop steric bulk on the structure of CDR-H3 in crystal structures.

We selected the pair that showed <55% V_L sequence identity from this data set for further analysis: 5Y2K (V_H: chain A, V_L: chain B) (Fig 4.12a) and 5Y2L (V_H: chain I, V_L: chain J) (Fig 4.12b). 5Y2K has one additional residue on CDR-L3 compared to 5Y2L, and one less residue on CDR-L1. The effect of this is that the space available to CDR-H3 is much broader in 5Y2K, whereas the short CDR-L3 coupled with the long CDR-L1 forms the hypothesised groove into which the CDR-H3 loop packs. Indeed, when both F_V domains are aligned on the V_H

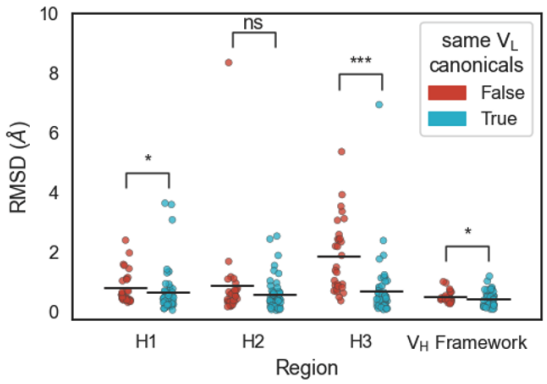


Figure 4.11: Dependence of V_L sequence identity influence on V_H structure on V_L canonical forms. Shown is the comparison of region RMSDs in the V_H -identical pairs when all V_L loops have the same canonical forms, or when at least one loop is has a different canonical form. Statistical significance was calculated using Mann-Whitney U tests (ns: not significant, *: $0.005 < p < 0.05$, ***: $p < 5 \times 10^{-7}$)

domain, we observe that, in the case of the shorter CDR-L3 loop, the CDR-H3 loop is folded down into the cleft and that this conformation of the loop would clash with the longer CDR-L3 loop (Fig 4.12c).

4.5 Discussion and conclusions

It has long been hypothesised that the conformational space of CDR-H3 loops depends not only on their sequences, but also on structural features in their vicinity. Several studies have confirmed this to be the case in specific F_V regions (Fernández-Quintero et al. 2020a; Teplyakov et al. 2016). A question that remains is whether the presence of the V_L domain, or indeed specific attributes of the V_L domain, can generally be shown to modulate CDR-H3 conformation in a predictable manner.

In this study, we show that V_L domains influence CDR-H3 structure by demonstrating that genetic attributes of the V_L domain influence both the structure and structural diversity of CDR-H3 loops. We propose a model to provide a physical explanation for this phenomenon, namely that the minor loops of the V_L domain can be arranged in different forms to physically restrain CDR-H3 loops, thereby modulating the conformational space that the latter are able to occupy. We argue that this is likely mediated through CDR-L loops providing steric bulk and scaffolding with which CDR-H3 can make interactions. We show that the CDR-H3

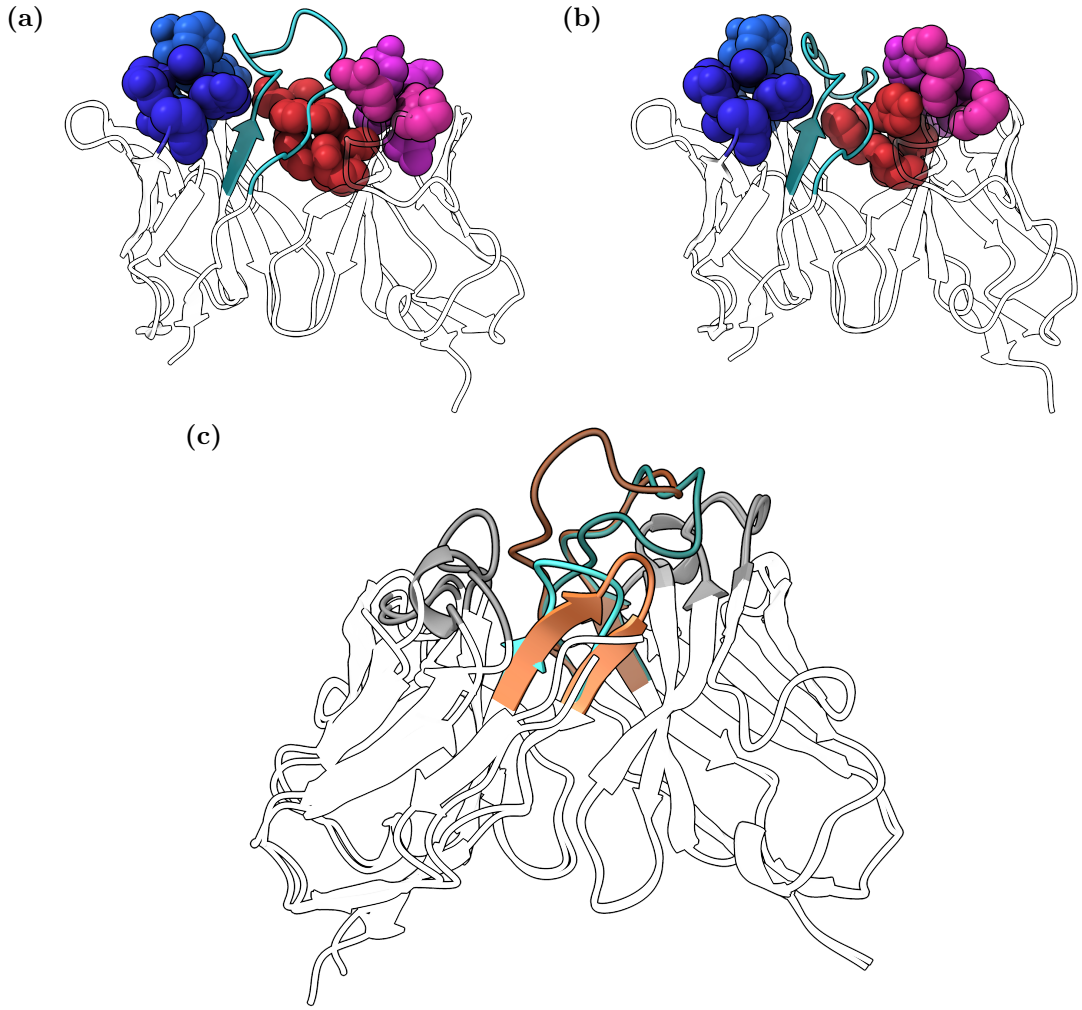


Figure 4.12: Case study demonstrating the caged CDR-H3 model. **(a)** shows 5Y2K with a shallower groove occupied by CDR-H3 due to the longer CDR-L3 and shorter CDR-L1 loop, whereas **(b)** shows 5Y2L with a deeper groove occupied by CDR-H3 due to the shorter CDR-L3 and longer CDR-L1 loop. Minor loops are shown using spheres, with CDR-H3 shown in cyan in **(a, b)**, with CDR-H3 and CDR-L3 shown in cyan and red, respectively. CDR-H1 and CDR-H2 are shown in dark and light blue, respectively. CDR-L1 and CDR-L2 are shown in magenta and pink, respectively. **(c)** shows the change in CDR-H3 conformation for V_H -identical pair 5Y2K (orange) and 5Y2L (cyan), with the longer CDR-L3 loop of 5Y2K resulting in CDR-H3 being forced out of the groove that is present in 5Y2L. CDR-H3 loops are shaded using darker colors, and CDR-L3 loops are shown in lighter shades.

loop forms contacts with the minor loops by packing against them and therefore hypothesise that, if these minor loops were to be found in different conformations, this would be reflected in the conformation of the CDR-H3 loop. We found that in simulations where minor loop conformations do not change, the packing effects are recapitulated, in that CDR-H3 efficiently explores the conformational space that is available based on the minor loop bulk. Further, we show that inter-loop packing is tighter in the case of the V_L , suggesting increased rigidity in the CDRs of this domain compared to the V_H domain. This is further supported by our observation of reduced flexibility of V_L loops compared to V_H loops.

Based on this, we show that predictable effects of minor loop attributes on CDR-H3 loop conformational diversity can be identified. We also identify the seemingly paradoxical observation that while longer CDR-L1 loops appear to reduce CDR-H3 diversity, the opposite is true for longer CDR-L3 loops, and we explain this observation within the context of our "caged" CDR-H3 model.

The effects observed here are one of many CDR-H3 conformation modulators, with the most significant being the specific sequence of the CDR-H3 loop. As is well documented, the CDR-H3 loop does not lend itself to sequence-based clustering approaches (Regep et al. 2017), suggesting that a plethora of factors are important in determining its exact structure, among which is the steric bulk provided by the minor loops. We note that residues that are traditionally considered to be part of the FW region, such as the DE loop, have also been shown to be structural modulators of CDRs (Kelow et al. 2020) and hypothesise that this may therefore lead to modulation of the proposed steric cage. This highlights the presence of multiple possible sources of CDR-H3 structural diversity apart from the sequence of the loop.

Our results also explain the observation that deep learning-based CDR-H3 loop modelling algorithms perform significantly better when trained using data that contains information on what is in the vicinity of the CDR-H3 loop (Abanades et al. 2021). Furthermore, our observations also provide a strong case for the formulation of antibody design as a multi-objective optimisation problem where all loops must

be considered together, rather than the traditional approach of modular design, where minor loops are clicked into place before subsequent modelling of CDR-H3.

In the context of B-cell immunity, our model is in line with previous studies suggesting that conformational diversity of antibodies generally decreases over the course of affinity maturation (Julian et al. 2017; Wong et al. 2011; Fernández-Quintero et al. 2020c; Ovchinnikov et al. 2018). A germline combination begins with a sub-space of possible CDR-H3 conformations due to the initial F_V pairing (Briney et al. 2018) and, over the course of the maturation process, accumulates more and more mutations that home in on and stabilise the required conformation of the CDR-H3 loop for epitope complementarity.

This might then explain the observation that individual repertoires show enrichment of similar V_L genes upon similar antigenic insult and in functional antibodies (Jaffe et al. 2022), as these genes may be conducive to the necessary CDR-H3 conformational space being explored to target the antigen. However, this is only one driver of V_L selection, as the V_L CDR loops, for example, also make contact with the antigen in many cases, placing emphasis again on antibody design and maturation as a multi-objective problem, both *in silico* and in natural B-cell immunity contexts.

In summary, we demonstrate the role of the V_L CDR loops as modulators of the CDR-H3 loop through modifications of the conformational subspace of the loop. Further research into this role and its examination in dynamical studies of antibodies where minor loop flexibility is also explored will allow us to more accurately understand the interconnected nature of CDR loop conformations. This will be important to fully understand the effect of this on rational antibody design. The results in this chapter also inform our approach in the next chapter, where we aim to refine predicted antibody structures and approach the problem by considering both the V_H and V_L domains at once.

5

Refinement of Predicted Antibody Structures Using Graph Neural Networks

Contents

5.1 Chapter abstract 126
5.2 Introduction 126
5.3 Methods 129
5.3.1 Quality assessment of structures predicted by GNNRefine 129
5.3.2 Training, validation, and test set curation 130
5.3.3 Graph neural network architecture 131
5.3.4 Training schemes 136
5.3.5 Calculation of predicted distances and comparison to ABodyBuilder2 137
5.3.6 Refinement of structures using OpenMM 138
5.4 Results 138
5.4.1 GNNRefine is unable to consistently improve antibody models generated by ABodyBuilder2 138
5.4.2 An expanded data set of full-length Fvs can be generated from the PDB by modelling terminal residues 142
5.4.3 Graph neural networks trained on antibody structures predict improved inter-Cβ distances 144
5.4.4 Trained graph neural networks improve structures predicted to be of lower quality 147
5.4.5 Energy minimisation using modified molecular mechanics force fields is unable to translate predicted distances to structural modifications 149
5.4.6 Inter-Cβ distances predicted by our models obey the triangle inequality 151
5.5 Discussion and conclusions 153

5.1 Chapter abstract

Computer-aided design of antibody therapeutics often relies on the availability of structural data, but it is currently impractical to solve the structure of every candidate antibody experimentally. In lieu of this, structural modelling allows for predictions of structure to be made. Although recent methods such as ABodyBuilder2 (ABB2) and AlphaFold-Multimer (AFM) have made significant advances in the prediction of antibody structures, the need for further improvement persists. This can be achieved either by better "naive" predictions or by refinement methods that improve an already existing structural model. In this chapter, we train a graph neural network specifically on antibody models from ABB2 in order to predict refined inter-Cβ distances. Our architecture is inspired by GNNRefine, a state-of-the-art tool for general protein refinement, with several additions to tune the architecture to the domain of antibody structure prediction. We show that our models improve inter-Cβ distances in low-quality starting models and that the quality estimates of the initial model are an excellent indicator of which models are candidates for refinement with our method. We then go on to shown that although our model is able to generate physically plausible distances, energy minimisation using OpenMM with predicted distances as restraints is unable to translate this information into improvements of the structure. This is in line with recent work on physics-based structure refinement. Our results demonstrate the learnability of mistakes made by naive structure predictors and highlight the need for architectures that will refine starting models without the need for molecular mechanics-based engines.

5.2 Introduction

A significant challenge of working with structural data in experimental settings is that most antibody discovery campaigns start with a large number of sequences from which candidates are chosen and elaborated upon during later stages, with

initial libraries routinely exceeding scales of 10^4 sequences. The Observed Antibody Space database (Kovaltsuk et al. 2018; Olsen et al. 2021) currently houses more than 2 million paired sequences, and next-generation sequencing technologies are allowing the evaluation of several million sequences per campaign (Rouet et al. 2018; Marks et al. 2020; Goldstein et al. 2019). The speed of sequence data generation is outstripping the speed of structure determination by several orders of magnitude, as the latter requires a much more time-consuming and expensive set of techniques. Indeed, this is reflected by the relatively small number of known antibody structures in SAbDab (Dunbar et al. 2014), currently placed at 7005[1].

Knowledge of the structure of an antibody facilitates the understanding of properties not accessible by sequence information alone, such as antibody-antigen docking to understand interactions relevant to complex formation (Weitzner et al. 2017; Schneider et al. 2021), structure-based clustering (Robinson et al. 2021; Spoendlin et al. 2023), or paratope similarity measurements (Wong et al. 2020) to identify sequence-dissimilar antibodies targeting a single epitope.

The lack of experimental structural data for the vast majority of candidate antibodies can be overcome by the use of structural models. For example, in Chapter 3 we examined developability risks of antibodies by performing MD simulations using structural models. However, approaches such as these require starting models to be accurate in order to yield correct information. For example, Schneider et al. 2021 showed that the quality of antibody-antigen docked poses is heavily dependent on the quality of the starting models. As such, structural modelling of antibodies is a key step in most antibody development pipelines.

Historically, antibody structural modelling has been performed using template-based approaches (Schoeder et al. 2021; Marks et al. 2017b; Leem et al. 2016), which perform well on the structurally conserved FW regions and minor CDR loops. However, such approaches fail to accurately model the CDR-H3 loop, which, as previously discussed, is often the major site of antigenic recognition. Recent work using deep learning approaches (Abanades et al. 2021; Evans et al. 2021;

[1]As of Nov 22nd, 2023.

Ruffolo et al. 2022; Lee et al. 2023) has resulted in considerable improvements in the accuracy of the prediction of all regions of the antibody, and this is especially evident in the CDR-H3 loop. For instance, AbodyBuilder2 (ABB2) (Abanades et al. 2021) is able to model the CDR-H3 loop of the Rosetta Antibody Benchmark (Guest et al. 2020) at an average RMSD of 2.81 Å, compared to ABodyBuilder (Leem et al. 2016; Marks et al. 2017b), a state-of-the-art homology modelling method, which results in a CDR-H3 RMSD of 3.46 Å using the same set of sequences.

In general protein structural modelling, a common approach to further increase the accuracy of generated structures is model refinement, which aims to modify a "naive" prediction. This can be achieved using a number of approaches including molecular mechanics (Heo et al. 2019) and fragment assembly (Lee et al. 2019), though recent successful approaches have also included deep learning, with DeepAccNet (Hiranuma et al. 2020) and GNNRefine (Jing et al. 2021) being state-of-the-art examples. While both perform similarly, DeepAccNet is based on convolutional neural networks, whereas GNNRefine relies on a graph neural network-based architecture and requires less computational time to refine a structure compared to the former. However, both of these methods are designed to work on structural models containing only single chains.

Since deep learning-based methods can benefit from domain-specific training (Abanades et al. 2022), we aimed to develop a antibody-specific structure refinement method. Here, we present results obtained from the training of models relying on a deep learning architecture similar to that of GNNRefine, which first uses a GNN to predict inter-Cβ atom distances in a starting model, followed by energy minimisation using Rosetta. We begin by showing that GNNRefine performs poorly on antibody structures, and then train a GNN to specifically refine models generated by ABB2. We show that our model performs well on CDR-H3 loops and that poor predictions are made mainly in cases where the starting structure has an RMSD of less than 1 Å. We then show that the uncertainty estimates of ABB2 can be used to identify structures which would benefit from refinement using our model. Lastly, we show that energy minimisation in OpenMM using these distances

as restraints does not result in improved RMSD values, even though the refined distances predicted by our model are physically feasible. Our results indicate that alternate architectures that explicitly update coordinates may be able to translate the improved predictions into structural models.

5.3 Methods

5.3.1 Quality assessment of structures predicted by GNNRefine

To refine structures with GNNRefine (Jing et al. 2021), we used ABB2 (Abanades et al. 2022) models of a set of 32 random antibodies (we used a resolution cutoff of 5 Å and ensured that no missing residues were found in the F_V region) deposited in the PDB (Berman et al. 2000) after the publication date of GNNRefine to ensure that these would not have been included in the training procedure. We used default settings where possible, but made several modifications to the source code in order to allow for multi-chain structures to be accepted by the software. Namely, functions in GNNRefine that extract features and nodes from input structures are hard-coded to only consider the first chain in an input model. We modified these functions such that both the V_H and V_L chains were considered by the software. We ensured that the order of the two chains was identical in all cases, with the V_L domain following the V_H domain. We then ensured that nodes were being generated for all residues, and that all relevant edges were being established prior to inference by the pre-trained network. Additionally, we generated four variants of GNNRefine by modifying the step in which the predicted inter-Cβ distances are used to move residues of the input model using PyRosetta (Chaudhury et al. 2010). These are outlined below:

- GNNR: Original implementation. Runs refinement on both chains, but ignores all but the first chain when predicting Cβ distances.
- Rosetta only: Ignores predicted Cβ distances when running refinement on both chains.

- Multichain: Predicts Cβ distances for all chains in the input structure rather than the first chain only (In the case of our Ab models, these are V_H chains). Runs refinement on all residues.
- CDR only: Predicts Cβ distances for all chains in the input structure rather than the first chain only. Runs refinement only on CDR loops by removing FW residues from the Rosetta MoveMap.
- CDR and FW only: Predicts Cβ distances for all chains in the input structure rather than the first chain only. Runs refinement only on CDR loops by removing FW residues other than two residues on either side of any CDR loop from the Rosetta MoveMap.

In order to compare GNNRefine and its variants to ABB2, we aligned the output models from all methods with their respective crystal structure backbone atoms and calculated RMSD values. For V_H and V_L CDRs or FW regions, we aligned of the V_H and V_L FW regions, respectively. When calculating V_H or V_L RMSD, we aligned on V_H or V_L, respectively. When calculating all CDR RMSD, we aligned on all FW residues. For whole F_V calculations, we aligned on the whole F_V.

5.3.2 Training, validation, and test set curation

In order to generate a diverse set of structures to train and test our models on, we used all structures deposited in SAbDab (Dunbar et al. 2014) as of 22.03.2023. We filtered the 13887 Fab/F_V structures found based on whether they were solved by X-ray crystallography or cryo-electron microscopy, and filtered out those that had a resolution of less than 5 Å, or had missing heavy atoms other than in the first two or last two residues of their variable domains. As our validation set, we used the Rosetta Antibody Benchmark (Guest et al. 2020), and as the test set, we used all antibodies deposited between 22.11.2022 and 22.03.2023 with no missing F_V residues and resolutions of 5 Å or better. All other entries were assigned to the training set. We ensured that there were no duplicate sequences across the three sets, but did allow duplicate sequences in the training set, as these can be

found in different conformations. This resulted in a set of 7927 structures for the training set, 41 structures for the validation set, and 32 structures for the test set. The full sets are provided in Appendix C.

We examined sequence identities between the train and test sets using cd-hit (Li et al. 2006), and found that 22 of the 32 structures had a sequence identity of 90% or lower across the F_V region.

Next, we prepared the identified F_V structures for model training. We first extracted the full sequence of the F_V from each PDB file and used ANARCI (Dunbar et al. 2015) to identify structures with missing termini. If the complete sequence had been deposited with the PDB, we used this sequence to model in the missing residues. If the sequence was not directly recoverable from the PDB entry, we used AbLang (Olsen et al. 2022) to predict the missing residue identities. We then modelled in missing termini residues using MODELLER (Webb et al. 2016), generating five models for each structure using the `very slow` refinement setting and picking the decoy with the lowest energy as the final structure. Next, we generated ABB2 models (Abanades et al. 2022) of each structure in the data set. ABB2 generates four unrefined models of the structure, picks the medoid based on RMSD as the most likely, and refines this using OpenMM. We saved both unrefined and refined structures generated by ABB2.

5.3.3 Graph neural network architecture

We trained our models using the same architecture that was used to train GNNRefine (Jing et al. 2021). We implemented the architecture from the ground up in pytorch (Paszke et al. 2019) and pytorch-geometric (Fey et al. 2019). The input features include residue-level features that are embedded in the nodes of the graph and residue pair-level features embedded in the edges of the graph. We retained the majority of features from GNNRefine, but added several new features to better define the antibody-specific context in the input data.

5.3.3.1 Input Features

As outlined, the input features of the graphs are separated into node- and edge-level features. The node features include a 38-dimensional vector, where the first 21 dimensions are one-hot encodings of the residue identity, with an additional category for modified or unknown residues. One dimension contains a relative positional encoding for each residue at a chain-level given by i/L where i is the index of the residue and L is the total number of residues in the chain. Two dimensions are one-hot encodings of whether the residue is part of the V_H or V_L domain. Seven dimensions are one-hot encodings of which fragment the residue belongs to in the variable domain (FW-1, CDR-1, FW-2, CDR-2, FW-3, CDR-3, FW-4) according to the IMGT numbering scheme. One dimension gives the ABB2 root mean squared predicted error (RMSPE) scaled by 0.1. The last six dimensions contain the backbone torsion angles in radians, relative solvent-accessible surface area, and one-hot encoding of DSSP (Kabsch et al. 1983) secondary structure labels (helix, strand, or coil, as described in Section 3.3). A 7-dimensional atom feature set is also calculated in a manner identical to that used in GNNRefine for each non-Cα heavy atom by computing the relative position of each atom to the Cα atom of the residue.

Edges for two residues in the graph are established if the Cβ (Cα for glycine residues) distance between them is less than or equal to 10 Å. The features for each edge are identical to those in GNNRefine, with three dimensions representing Cα-Cα, Cβ-Cβ, and N-O distances, scaled by a factor of 0.1. Three further dimensions contain angular terms that describe the relative orientation of the two residues as described in trRosetta (Yang et al. 2019). For residue pair i, j, ω is the dihedral angle given by the atoms Cα_i, Cβ_i, Cβ_j, Cα_j, ϕ is the angle given by Cα_i, Cβ_i, Cβ_j, and θ is the dihedral angle given by N$_i$, Cα_i, Cβ_i, Cβ_j. Lastly, a discretised embedding of the absolute sequential distance of the two residues is provided using 9 bins: [1, 2, 3, 4, 5, 6-10, 11-15, 16-20, >20]. We established self-edges as we found that these resulted in more stable training.

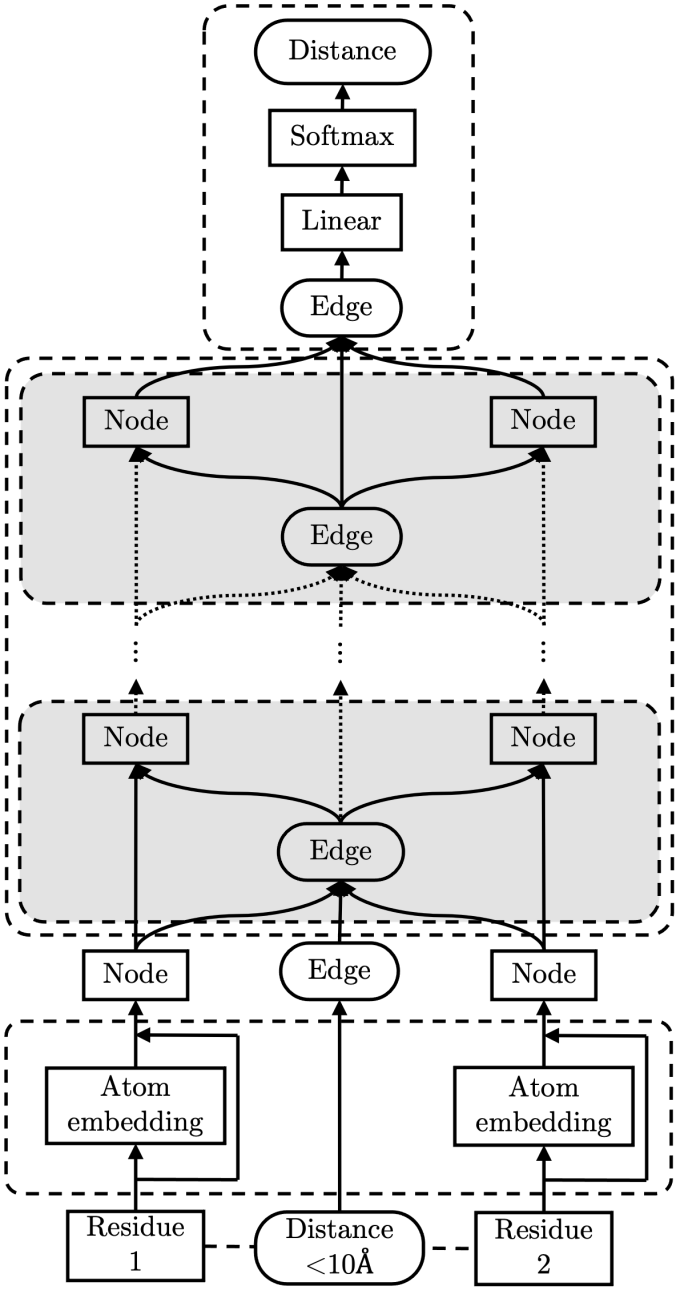


Figure 5.1: GNNRefine architecture. An edge is established between two residues if the distance between their Cβ (Cα for glycine residues) atoms is less than or equal to 10 Å. Atom embeddings are calculated using the atom embedding layer and concatenated to residue features. Each grey box shows a single message passing layer, of which there are 10 in total. After message passing, the edge features are passed through a linear layer and a softmax calculation is used to determine the distance probabilities. Adapted from Jing et al. 2021.

5.3.3.2 Architecture

We used a similar architecture to the original GNNRefine model to train our antibody-specific predictors. The model consists of 12 layers (Fig 5.1). The first of these is an atom embedding layer, which embeds the previously described residue-level atom features in a 256-dimensional space. This layer uses instance normalisation, followed by a linear layer. Non-linearity is then added using a rectified linear unit (ReLU) layer, followed by a second instance normalisation pass and a second linear layer. The resultant atom embeddings are then concatenated to the residue-level features.

The atom embedding layer is followed by a series of 10 message-passing layers (MPL) (Fig 5.2; algorithmic details for the MPLs is provided in Appendix C). The MPL architecture consists of the message calculation block, the aggregation block, and the update block.

Messages are calculated using edge features and node features of the connected nodes for each edge (Fig 5.2a). These are concatenated and normalised using an instance normalisation layer and then passed through a linear layer, followed by a leaky ReLU unit with the negative slope set to 0.01. During training, we added a 10% dropout layer after concatenation. The resultant activation values are then passed through a long short-term memory (LSTM) cell (Hochreiter et al. 1997), which takes the current activations as input and the edge features from the previous layer as the hidden state, with cell states being passed forward. Lastly, the resultant edge features are passed through a linear layer to calculate attention values for message aggregation. Messages for each node are then aggregated (Fig 5.2b) by multiplying sparsely evaluated softmax attention values with messages to each node and summing over the resultant values. Messages calculated in layer l also become edge features for layer $l + 1$.

Node updates are conducted in a manner similar to message calculation (Fig 5.2c). We first concatenate the current node features with the aggregated message, adding a dropout of 10% here during training. The concatenated input is then instance normalised and passed through a linear layer. Non-linearity is added

by a leaky ReLU unit with a negative slope of 0.01. Similarly, an LSTM cell is used with the node features from the previous layer as hidden states, generating the updated node features.

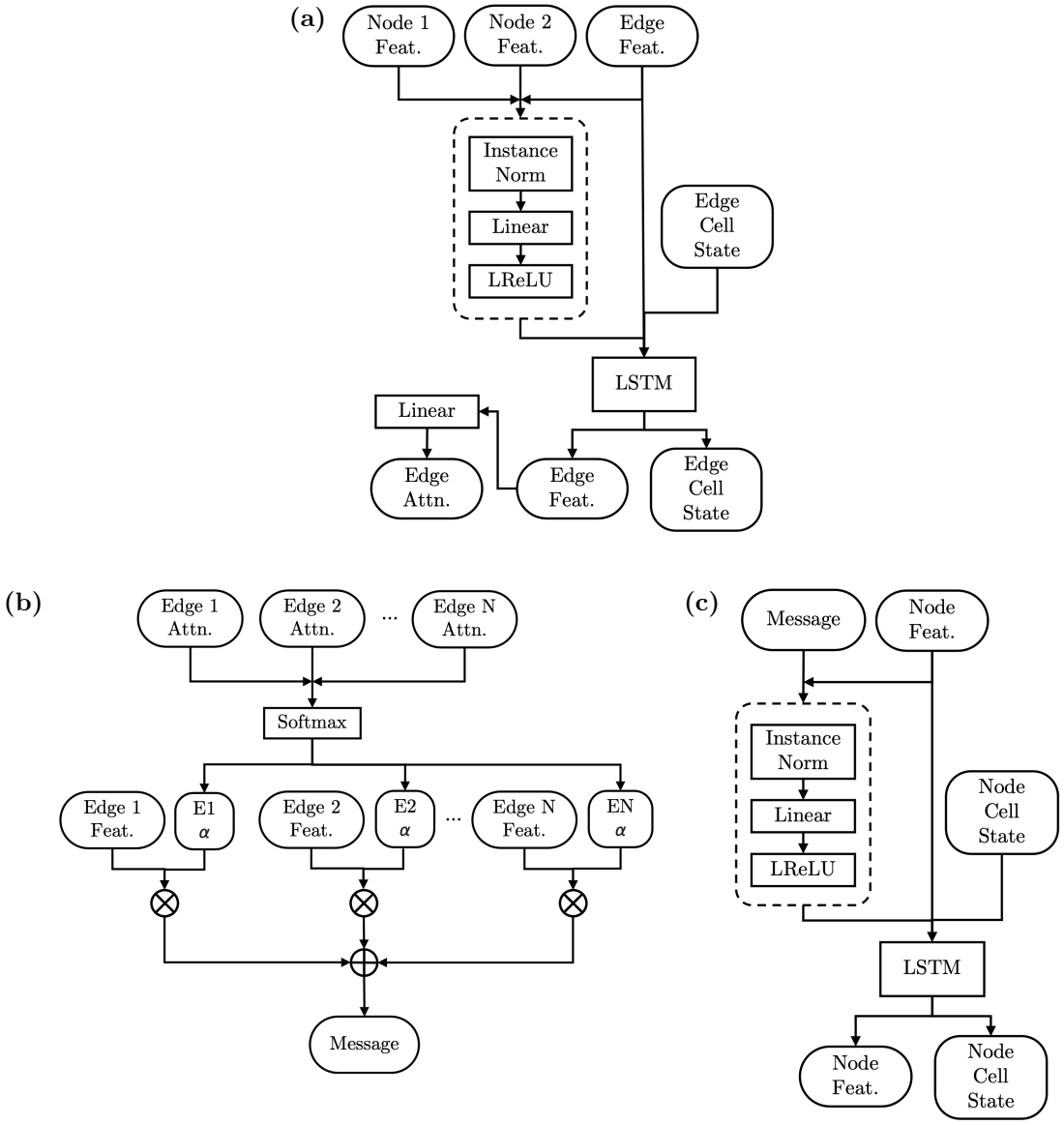


Figure 5.2: Details of the GNNRefine message passing layers. **(a)** shows the calculation of edge features and edge attentions based on the edge feature and features of the connected node from the previous layer. **(b)** shows how the message for a node is calculated using edge features and attention values of its edges. **(c)** shows the update procedure for a node using the calculated message and node features from the previous layer. Adapted from Jing et al. 2021

Lastly, the output is calculated in the form of a 37-dimensional vector for each edge using a linear layer followed by a softmax layer. The output is the probability

distribution for the predicted distance between Cβ (Cα in the case of glycine) atoms of the two residues connected by the edge. The first 36 dimensions are based on 36 bins standing for distances between 2 and 20 Å, using a 0.5 Å interval, and the last dimension represents distances >20 Å.

5.3.4 Training schemes

We trained all models using the Adam optimiser (Kingma et al. 2014) with default settings, and an initial learning rate of 0.01, which was halved every 10 epochs of training, for a total of 30 epochs. We also trained models using a hidden layer dimension of either 256 or 512 (with a corresponding change in batch size from 16 to 8 in order to avoid computer memory errors) and found that training using the smaller models with larger batch sizes generated better results.

As the prediction task is formulated as an ordinal regression problem, we used the ordinal log loss (OLL) (Castagnos et al. 2022) function over the cross-entropy loss function, which is given by

$$\mathcal{L}(P, y, \alpha) = -\sum_{i=1}^{N} \log(1 - p_i)\ |y - i|^{\alpha} \tag{5.1}$$

for a given output probability distribution $P = (p_1, ..., p_N)$ with N classes where the true label is y. α is a penalty factor, increasing the loss exponentially for high probabilities for classes further away from the true label. During model training, we found that OLL with $\alpha = 1.5$ performed much better than cross-entropy loss, as predictions are penalised more the further they are from the true label, as opposed to cross-entropy loss, where distance away from the true label has no effect on the magnitude of the loss.

ABB2 shows near-perfect accuracy across the FW regions, with the lowest accuracy in the CDR-H3 loop. To this end, we weighted the loss across different regions of the antibody:

$$\mathcal{L} = \frac{w_{\mathrm{FW}}\mathcal{L}_{\mathrm{FW}} + w_{\mathrm{CDR}}\mathcal{L}_{\mathrm{CDR}} + w_{\mathrm{H3}}\mathcal{L}_{\mathrm{H3}}}{w_{\mathrm{FW}} + w_{\mathrm{CDR}} + w_{\mathrm{H3}}} \tag{5.2}$$

where w_X is the weight and $\mathcal{L}_\mathrm{X}$ is the base loss of the FW, the minor CDRs, or CDR-H3 respectively. We found that using a weight of 1 for the FW, a weight of 5 for the minor CDRs, and a weight of 10 for CDR-H3 provided the most accurate results.

We also used two upsampling schemes to better balance the quality of ABB2 models that the models were exposed to during training. In the first scheme, the number of appearances of a model in a single epoch was calculated as the ratio of the CDR RMSD of the model to its crystal structure and the smallest CDR RMSD of any model in the training set to its crystal structure, rounded up:

$$n_i = \left\lceil \frac{\mathrm{RMSD}_{\mathrm{CDR},i}}{\min\{\mathrm{RMSD}_{\mathrm{CDR},1}, ..., \mathrm{RMSD}_{\mathrm{CDR},N}\}} \right\rceil \tag{5.3}$$

where n_i is the number of times model i will appear in a single epoch, $\mathrm{RMSD}_{\mathrm{CDR}}$ is the CDR RMSD of the model to its crystal structure, and N is the total number of training examples. We chose a more conservative second scheme, calculating n_i using the 25^{th}percentile, P_{25}, instead:

$$n_i = \left\lceil \frac{\mathrm{RMSD}_{\mathrm{CDR},i}}{P_{25}\{\mathrm{RMSD}_{\mathrm{CDR},1}, ..., \mathrm{RMSD}_{\mathrm{CDR},N}\}} \right\rceil \tag{5.4}$$

We found that neither upsampling scheme had an appreciable effect on the validation set loss.

We trained a total of 36 models using different combinations of the described hyperparameters, identified the most effective combination based on loss across the validation set, and then trained two additional models using this configuration.

5.3.5 Calculation of predicted distances and comparison to ABodyBuilder2

We calculated the distances between two residue Cβ atoms based on the predictions using a weighted average of the probability distributions output by our models:

$$d = \frac{\sum_{d\in 2.25,...,19.75} dp_d}{\sum_{d\in 2.25,...,19.75} p_d} \tag{5.5}$$

We excluded the first and last bins, as the first corresponds to nonphysical distances and the last corresponds to any distance greater than 20 Å. To compare

ABB2 distances to the refined distances predicted here, we calculated the mean absolute error (MAE):

$$\text{MAE} = \frac{\sum_{i=1}^{N} |\hat{d}_i - d_i|}{N} \tag{5.6}$$

where, for edge i of N total edges, $\hat{d}_i$ is the predicted distance by either ABB2 or our model, and d_i is the true distance of the edge in the crystal structure.

5.3.6 Refinement of structures using OpenMM

We used OpenMM (Eastman et al. 2016) to incorporate the distance predictions and refine the ABB2 output PDBs. We modelled the PDBs using the Amber 14 force field (Maier et al. 2015) without solvating them. Next, we applied positional harmonic restraints to heavy atoms of non-CDR-H3 residues using a force constant of $1 \times 10^4\,\text{kJ}\,\text{mol}^{-1}\,\text{nm}^{-2}$. We then enumerated the edges involving CDR-H3 residues and added the Cβ-Cβ distance as a harmonic restraint. To examine the effect of the strength of these restraints, we tested distance spring constant strengths of 1×10^1, 1×10^2, 1×10^3, and 1×10^4 $\text{kJ}\,\text{mol}^{-1}\,\text{nm}^{-2}$. Additionally, we tested a second scheme wherein distance restraints for all CDR loops were used and only FW residues were restrained. This scheme is identical to the one described where CDR-H3 loops are allowed to move and a positional harmonic restraint is added to all other residues, except that the positional restraint is only applied to FW residues, with all CDR loops being allowed to move.

5.4 Results

5.4.1 GNNRefine is unable to consistently improve antibody models generated by ABodyBuilder2

We began our examination of antibody model refinement by testing the performance of a state-of-the-art tool on the task, GNNRefine. This is a tool developed by Jing and Xu (Jing et al. 2021), and involves the prediction of Cβ-Cβ distances across residues of a protein structural model using a GNN. These restraints are then fed into the Rosetta suite in order to minimise the structure, thereby generating a refined

model. We began by using the published version of GNNRefine as is, and used ABB2 models of antibodies that GNNRefine would not have seen during training. We found that GNNRefine was unable to improve modelled antibody structures from ABB2 (Fig 5.3). All loops tended to be degraded. Across the heavy chain, CDR-H1 (Fig 5.3a) and CDR-H2 (Fig 5.3b) were degraded to a greater extent than CDR-H3 (Fig 5.3c). We hypothesize that this is due to the fact that these two loops adopt canonical forms (Chothia et al. 1987; North et al. 2011; Nowak et al. 2016), and would benefit more from domain specific training schemes compared to the more diverse CDR-H3. Across the V_L CDRs (Fig 5.3d, 5.3e, 5.3f), we found much higher rates of degradation. This is expected, as no restraints were calculated for these residues using GNNRefine, thereby allowing the Rosetta minimisation procedure to freely move the residues according to the internal scoring function. The framework residues are generally correctly predicted by ABB2, and we unsurprisingly found that these were generally degraded. Once again, we found the degradation across the V_H domain (Fig 5.3g) to be lower than across the V_L domain (Fig 5.3h).

Considering that GNNRefine only uses restraints for the V_H domain and performs better on this than the V_L domain, we next modified the code to examine whether generating distance restraints for the V_L domain would improve prediction. In order to allow for treatment of the second chain, we modified the code so that both chains are parsed and considered. Since the FW residues were also consistently degraded due to their having been modelled correctly using ABB2 with an average RMSDs of 0.54 Å and 0.57 Å for the V_H and V_L domains, respectively (Abanades et al. 2022), we also included changes to restrict movement to the CDR loops only. The modifications are described in Methods Section 5.3.

We found that our modified version of GNNRefine performed better than the originally published code, though we found the behaviours to be dependent on which loop was being considered (Fig 5.4). For all loops other than CDR-L1 (Fig 5.4a, 5.4b, 5.4c, 5.4e, 5.4f), we found that restricting movement to the CDR loops only provided the same or slightly improved qualities when considering the median values, but we consistently found degradation of high quality structures.

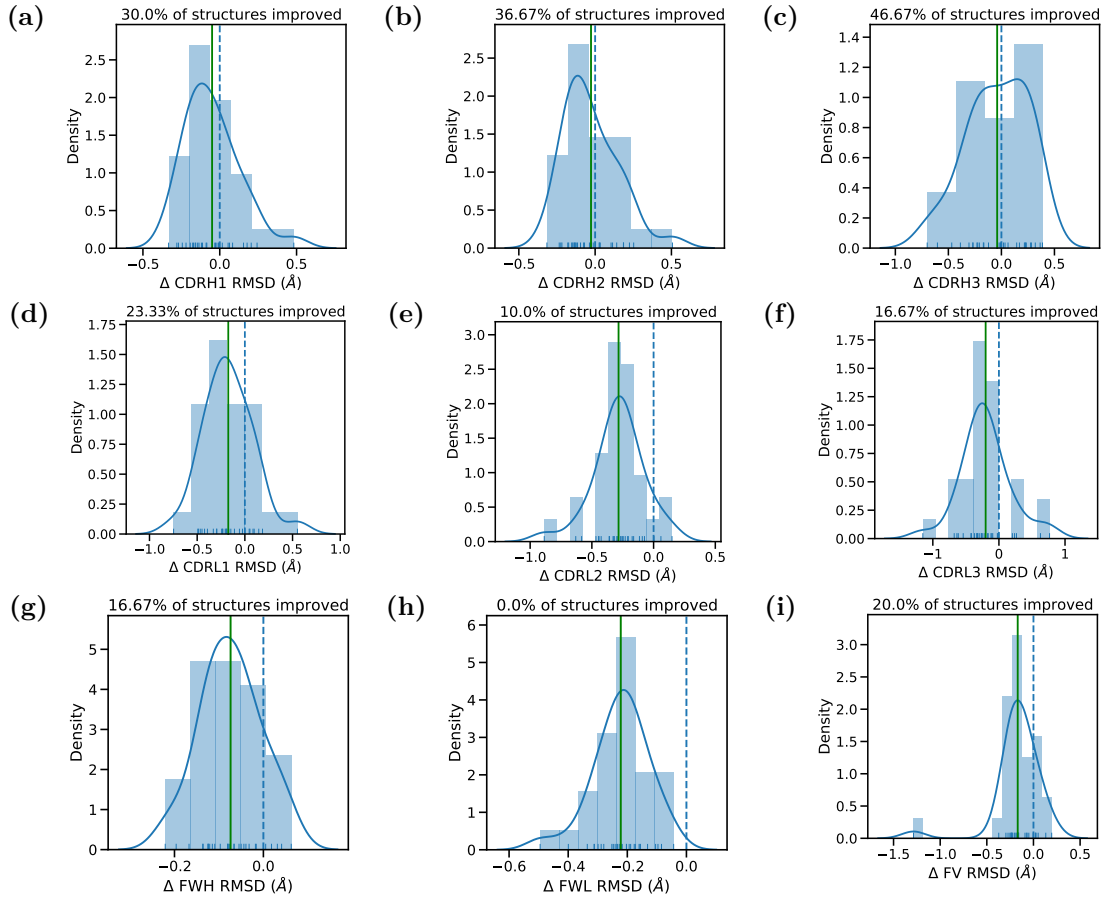


Figure 5.3: Performance of unmodified GNNRefine on ABB2 models. Antibody structures unseen by GNNRefine were modelled using ABB2 followed by refinement using GNNRefine. Shown are the changes in RMSD to the crystal structure upon refiement using GNNRefine, compared to ABB2 RMSD. The average is shown using a green line, and the percentage value shows the number of structures where GNNRefine improves the RMSD to the crystal structure. **(a-c)** show heavy chain CDR loops, **(d-f)** show light chain CDR loops, **(g,h)** show performance on FW residues, and **(i)** shows performance on the whole F_V region.

As FW residues were restricted, these did not show appreciable changes (Fig C.1). Further, we observed that removing distance restraints and minimising using only the Rosetta scoring function resulted in significant degradation of all regions (Fig 5.4), suggesting that the generated restraints do indeed contain information on how to improve the structure, and that the hybrid physical and statistical scoring function used during the minimisation finds a local optimum that is further away from the crystal structure than the input model. Based on these results, we conclude that information to aid in the structural model refinement process can be learned

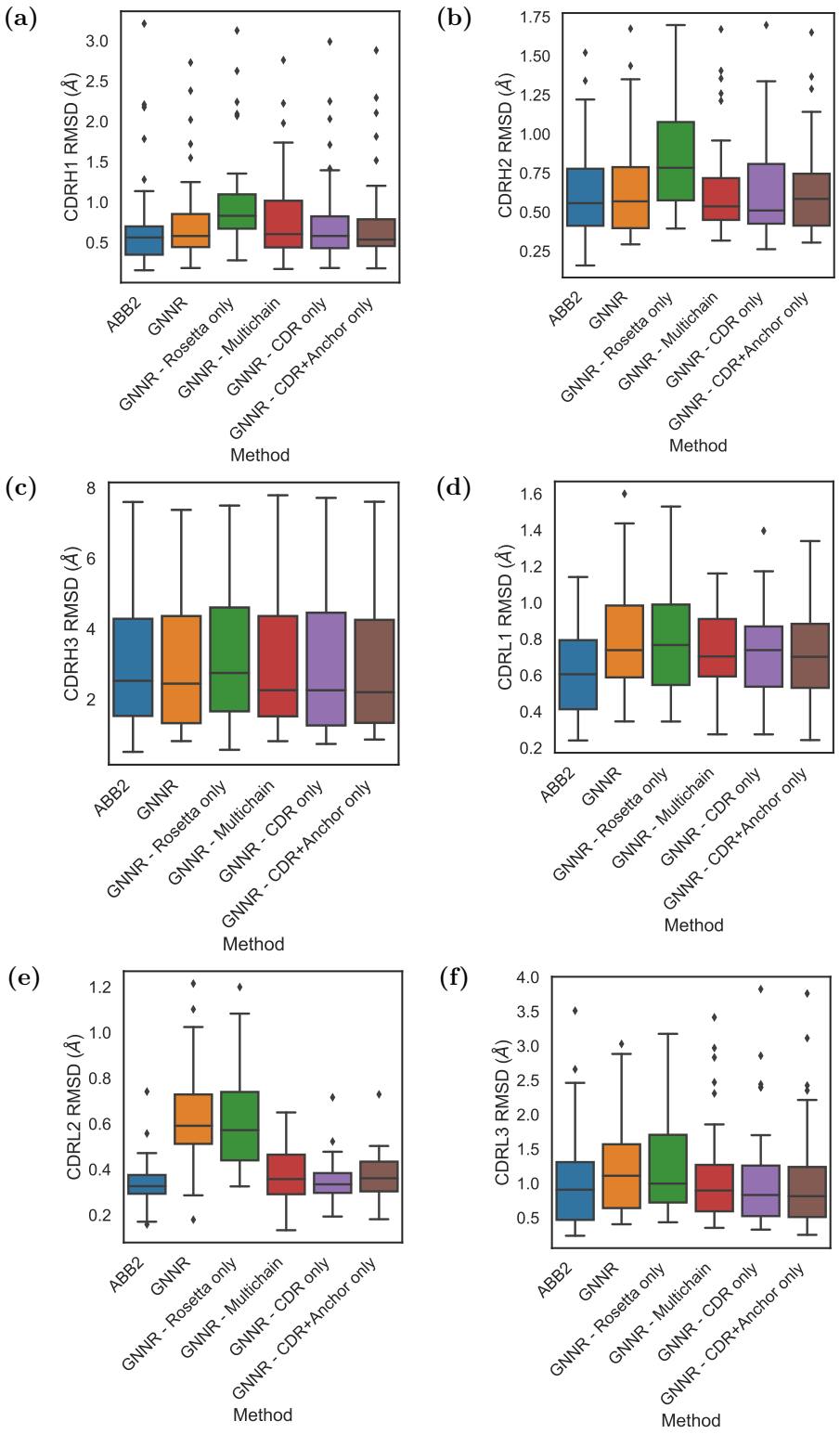


Figure 5.4: Performance of modified GNNRefine versions on ABB2 models. Antibody structures unseen by GNNRefine were modelled using ABB2 followed by refinement using GNNRefine and variants, as described in Section 5.3. Shown are the distributions of RMSD values compared to crystal structures for the different versions of GNNRefine, and for ABB2. The boxplots are drawn using the interquartile ranges, with the middle line denoting the median. **(a-c)** show heavy chain CDR loops, and **(d-f)** show light chain CDR loops.

using a GNN, and that classical scoring functions are likely to push antibody structural models generated by ABB2 further away from the crystal structure.

5.4.2 An expanded data set of full-length Fvs can be generated from the PDB by modelling terminal residues

Based on our observations on the performance of GNNRefine on ABB2 models, we hypothesized that training a similar GNN using only antibody structures and providing additional, domain-specific input information may result in better performance in determining inter-residue distances. We first constructed a data set of known antibody structures from SAbDab.

We began data set curation by examining the number of structures within SAbDab where only terminal residues on either the V_L or V_H domains according to the IMGT scheme were missing, and found that the majority of structures only miss up to three residues at any terminus (Fig 5.5).

As we had previously determined that the terminal residues are highly mobile (see Chapter 4) we hypothesised that these could be modelled into the structures, and that the benefits stemming from the expansion of training examples through this approach would outweigh the relatively small amount of inaccuracy added through the possible misplacement of termini. Since the C-terminus is on the opposite end of the domain relative to the CDR loops, we argue that misplaced C-terminal residues would have no impact on edges involving CDR loops, due to the 10 Å distance cutoff for edge establishment. Further, since the number of edges is proportional the the number of residues, we argue that up to three potentially misplaced N-terminal residues would only make up a small number of edges involving CDR loops, and that the vast majority of established edges would originate from residues resolved in the structure.

We also examined whether changing the resolution cutoff has an effect on how many structures with only missing terminal residues are added to the data sat, and found that, at higher cutoffs, the data set was expanded more dramatically by inclusion of such structures compared to lower resolution cutoffs (Fig 5.6).

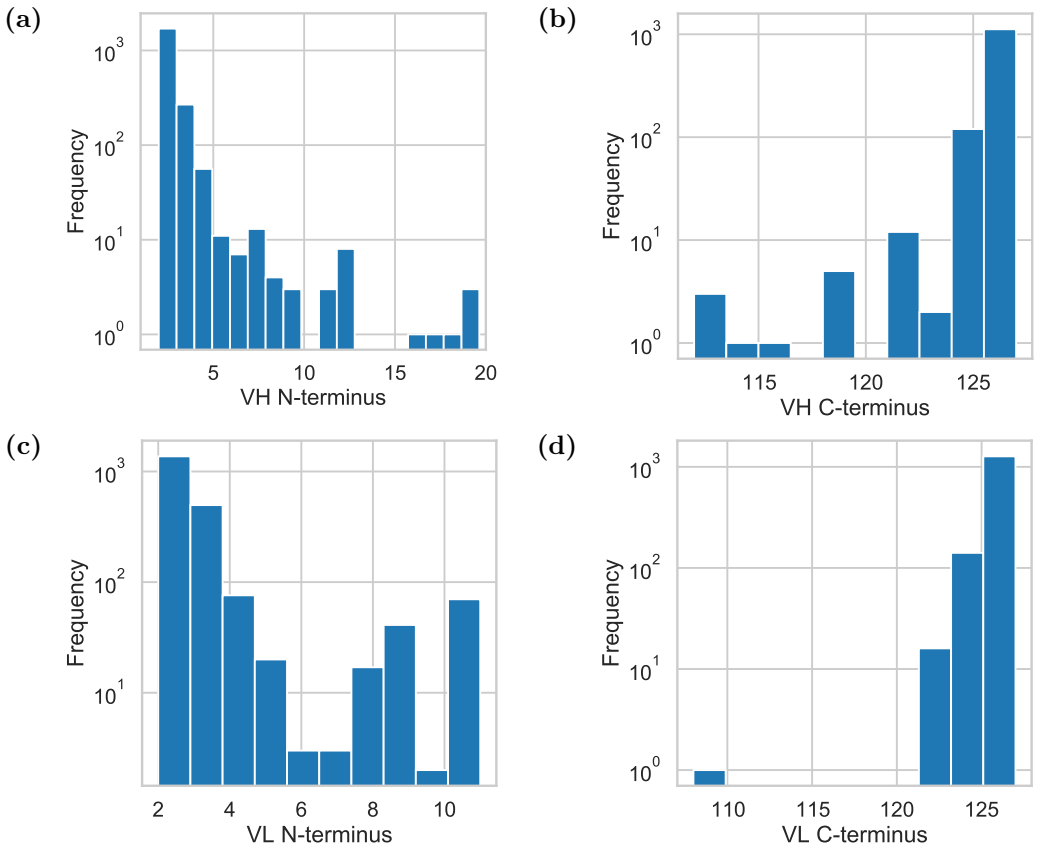


Figure 5.5: Distribution of first residues resolved in F_V crystal structures. We extracted all F_V structures from SAbDab, removing those with missing residues in either the V_H or the V_L domains. We numbered these structures using IMGT scheme and tabulated the number of missing residues at any terminus. **(a)** and **(b)** show distributions resolved of N- and C-termini in the V_H domains, respectively, and **(c)** and **(d)** show distributions resolved of N- and C-termini in the V_H domains, respectively. *NB:* for clarity of visualisation, domains with completely resolved termini are excluded in these plots and y-axes are given in logarithmic scale.

At our chosen cutoff of 5 Å, this resulted in a 45.9% increase in training examples. Based on these observations, we decided to create an expanded data set of antibody structures by modelling up to three terminal residues into F_V regions. We conclude that this approach would enable incorporation of more information for machine learning tasks requiring full F_V structures.

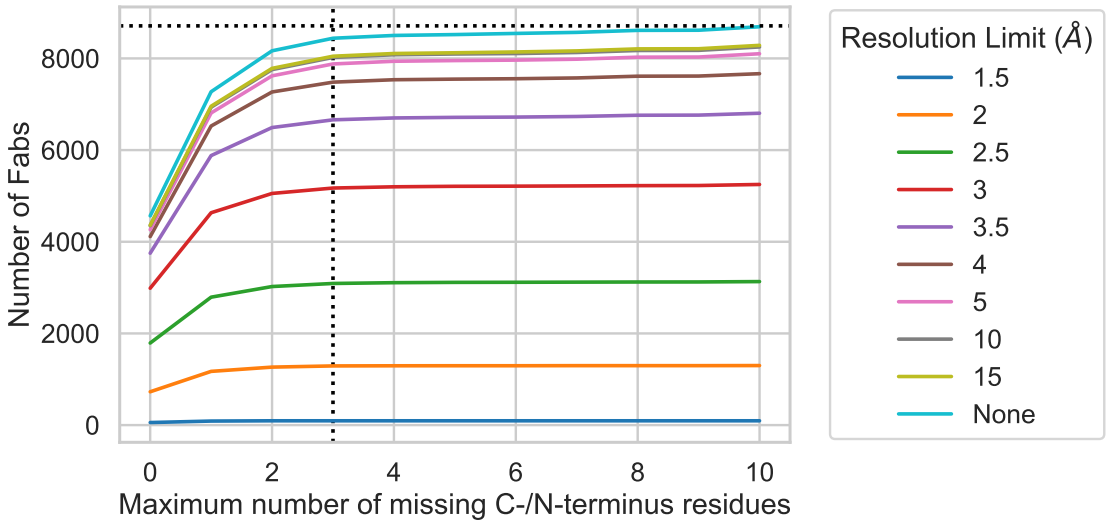


Figure 5.6: Cumulative distributions of structures with missing terminal residues. The plot displays the cumulative count of structures with up to a certain number of missing residues at any terminus. Each line represents a different resolution threshold, with structures filtered to include only those with equal or better resolution than the specified cutoff.

5.4.3 Graph neural networks trained on antibody structures predict improved inter-Cβ distances

Based on our observations using GNNRefine, we trained GNNs using a similar architecture on our expanded data set, making modifications to input features and the architecture to better enable learning specifically for antibody structures (see Methods Section 5.3). We trained a total of 36 models and found that the best performance was achieved by training on unrefined ABB2 models using a batch size of 16, without the use of upsampling, and with loss weights set to 1, 5, and 10 for FW, minor CDR loops, and CDR-H3, respectively. We then trained two additional models using the same hyperparameters and found that two of the three models performed similarly, while the other performed significantly better (Fig 5.7). We found that the best model was able to improve inter-Cβ predictions relative to ABB2 for CDR loops, but not for FW residues. Further, we found the improvement to be best when considering the CDR-H3 loop, and we hypothesise that this may be due to the relatively high weight of CDR-H3 edge accuracy in the calculation of the loss during the training procedure, since models that were trained without

this reweighting showed inferior performance.

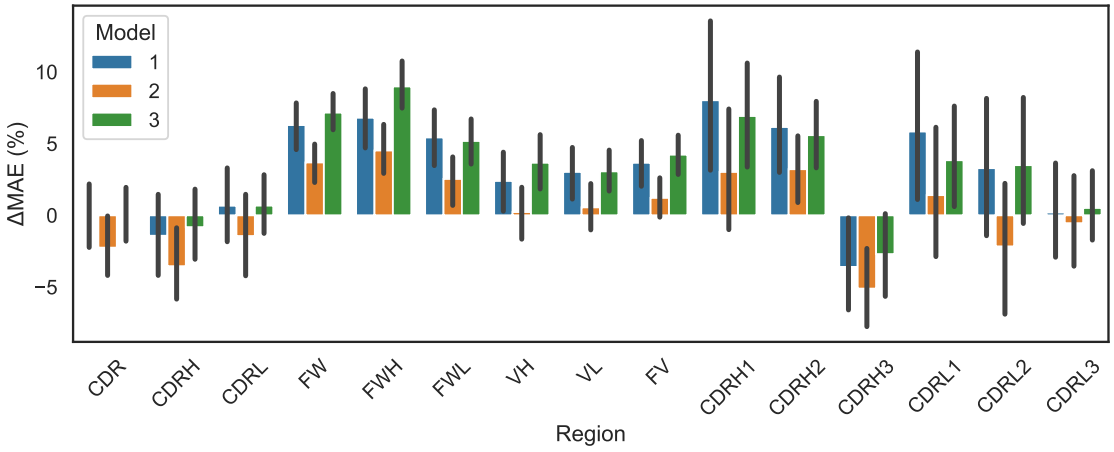


Figure 5.7: Performance of our models across different regions of antibodies in predicting inter-Cβ distances. The plot shows average changes in mean absolute error (MAE) for edges involving at least one residue in the regions given on the x-axis. Error bars represent 95% confidence intervals. A decrease in MAE indicates an improved average inter-Cβ distance prediction compared to ABB2 (the baseline).

In order to better understand the degree to which inter-Cβ distances were being degraded or improved in individual ABB2 structural models, we examined the distribution of MAE changes upon prediction of refined distances using our method (Fig 5.8). Using this metric, a negative ΔMAE represents improvement, while a positive value is indicative of degradation. We found that for all loops except CDR-H2 (Fig 5.8a, 5.8c, 5.8d, 5.8e, 5.8f), at least half of the ABB2 models were improved, and we found the greatest improvement in CDR-H3 (Fig 5.8c), with improvements in 25 of the 32 structures in this region. Further, we found that improvements for this loop could be as large as 15% compared to the inter-Cβ distances predicted by ABB2. We found this to be the case for other loops as well, with improvements up to around 15% being observed. However, we also note that degraded structures are often degraded significantly, with degradation in excess of 20% being observed at least once for most loops (Fig 5.8). Based on this, we hypothesised that the models where inter-Cβ distances were degraded might be models of high starting quality in the first place. If the ABB2 structural models show a very low initial RMSD, then any change in inter-Cβ distances will be more likely to be detrimental.

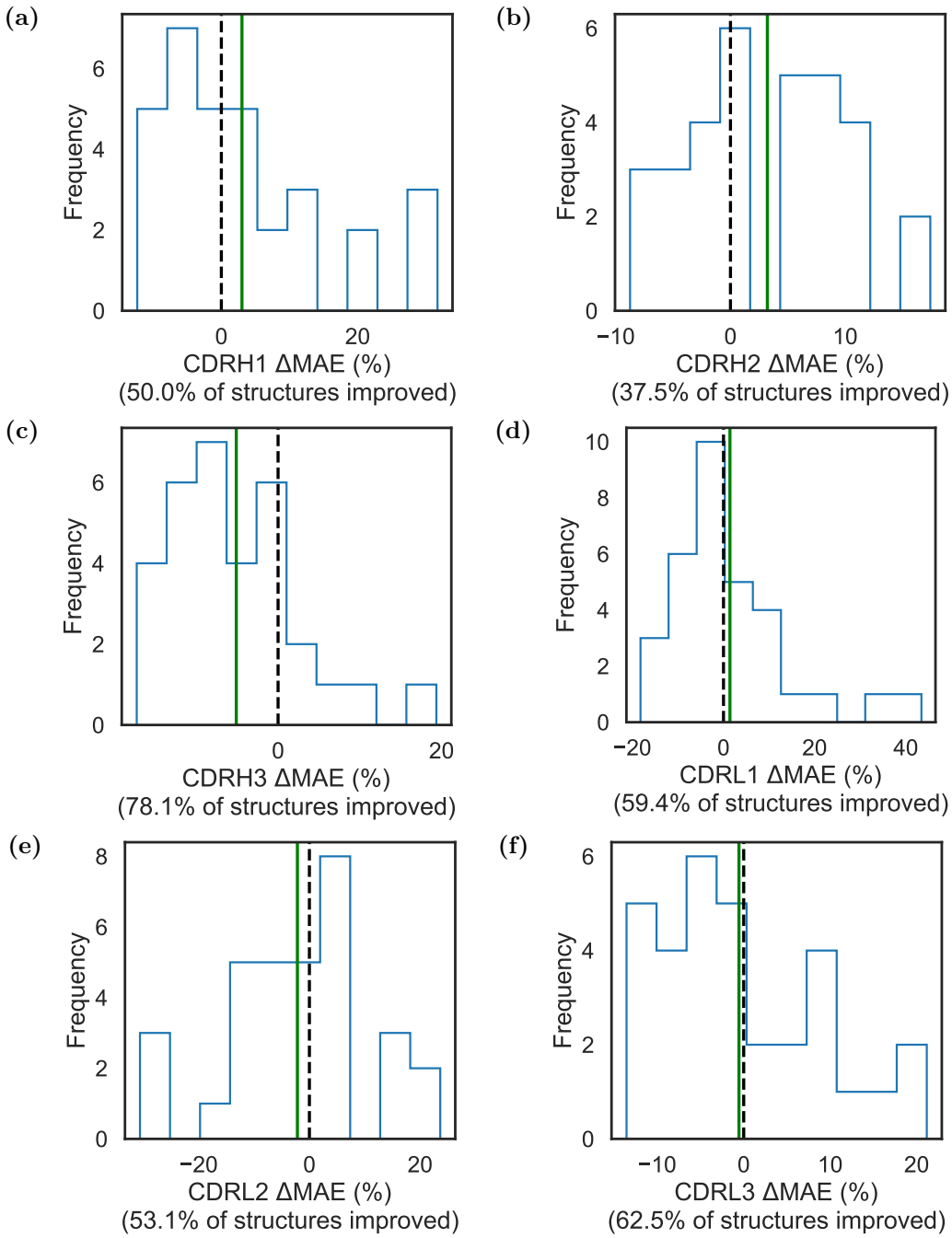


Figure 5.8: Changes in MAE of inter-Cβ distances compared to ABB2. Shown are the percentage changes in MAE relative to ABB2 when refined inter-Cβ distances were inferred using our model. The green lines represent averages across all test set structures, and the percentage value shows the number of structures where our model improves the average MAE compared to ABB2. **(a-c)** show V_H CDR loops, **(d-f)** show V_L CDR loops. Framework residues and the entire F_V are shown in Fig C.2.

To test this hypothesis, we investigated the relationship between the initial MAE in inter-Cβ distances of ABB2 models and the change induced by calculating refined distances using our model (Fig 5.9). We found significant depedence between the quality of the initial structures and the absolute ΔMAE metric in all loops, with this being most pronounced in the case of CDR-H3 (Fig 5.9c). This confirms that when the initial models are of high quality, our models are unable to improve the inter-Cβ distances. All structures were improved in all loops (with the exception of one structural model in the CDR-L3 loop) if the average initial MAE for edges of the loop was greater than 1 Å. Further, we found that the absolute ΔMAE for the maximally improved test structure for each loop is higher than the average ΔMAE for the example that was degraded the most. We observed this to be evident in the cases of CDR-H3, CDR-L1, CDR-L2, and CDR-L3 (Fig 5.9c, 5.9d, 5.9e, 5.9f). Based on these results, we conclude that our model, trained specifically on antibody structures and ABB2 models, is able to rectify errors made by ABB2, especially in cases where the structural model generated by ABB2 is of low quality. The ability to differentiate between high- and low-quality starting structures would enable us to select which ABB2 models would present candidates for refinement using our model.

5.4.4 Trained graph neural networks improve structures predicted to be of lower quality

Having demonstrated the ability of our model to improve inter-Cβ distances in ABB2 structural models and having shown the relationship between refined distances and input model quality, we next sought to identify a method to test which structures might benefit from refinement. When a sequence is modelled using ABB2, four individual models are produced using four sets of weights and biases. These four models are then compared to one another by calculating pairwise RMSD values, and the medoid is chosen as the final structure. Further, the average pairwise RMSD per residue across all four predictions is used as the root mean square predicted error (RMSPE). Abanades *et al.* show that the RMSPE correlates well with the actual

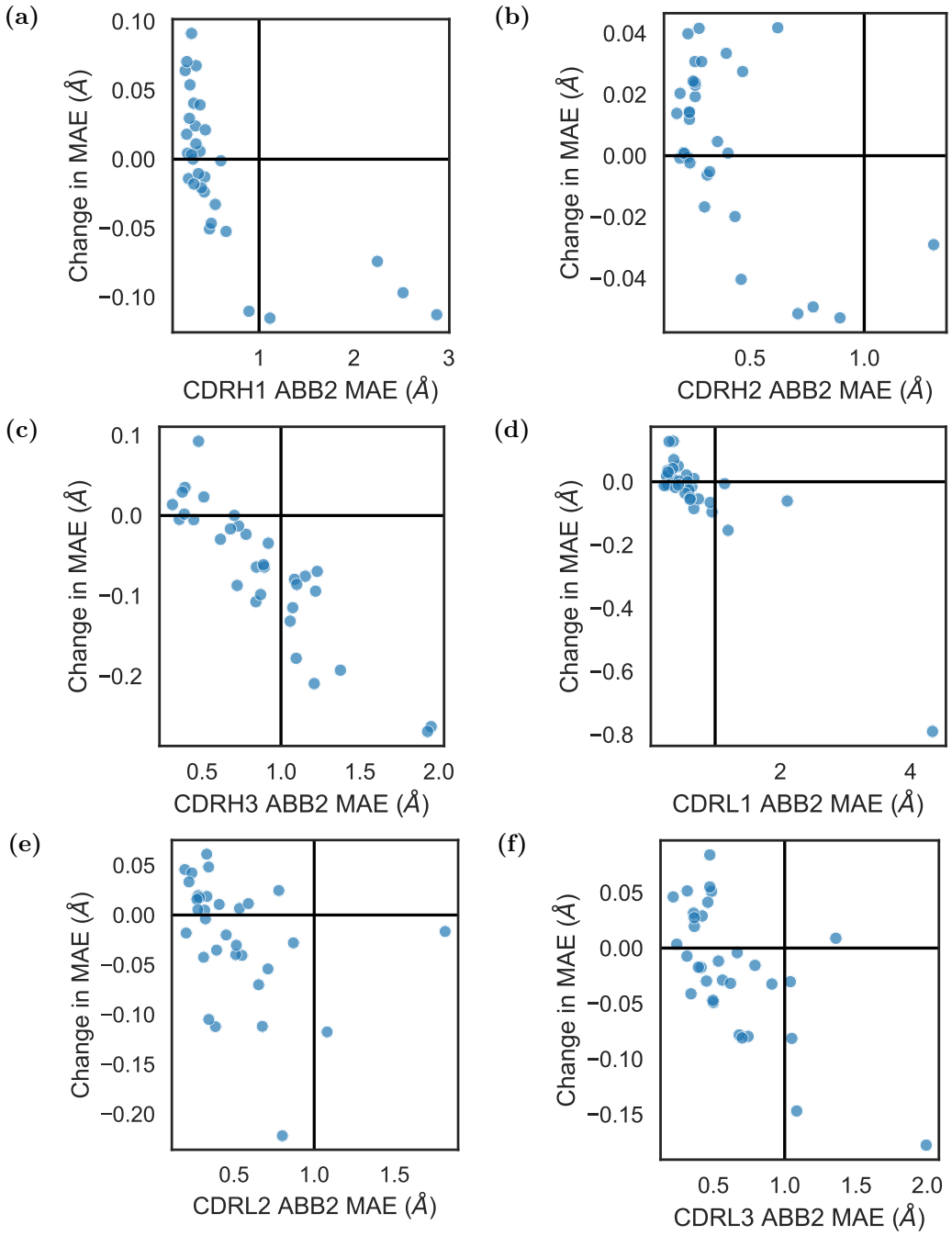


Figure 5.9: Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 error. Shown are the absolute changes in MAE relative to ABB2 when refined inter-Cβ distances were inferred using our model, plotted against the original mean MAE for each ABB2 model. **(a-c)** show V_H CDR loops, **(d-f)** show V_L CDR loops. Framework residues and the entire F_V are shown in Fig C.3.

error of the structural model when compared to the crystal structure (Abanades et al. 2022), especially in the CDR-H3 loop.

In order to adapt the ABB2 RMSPE to our method, we calculated the predicted error per edge by averaging the RMSPE of the two residues that are connected by an edge. We then analysed correlations of the average per-edge RMSPE of a region and our predicted inter-Cβ distances (Fig 5.10), and found that greater RMSPE is correlated with greater levels of improvement of predicted distances using our models, though we note that this effect is loop-specific. In the case of CDR-H2 and CDR-L2, the correlations are weak ($r = -0.26$ and $r = -0.33$, respectively) and not statistically significant (Fig 5.10b, 5.10e), whereas for the other minor loops RMSPE is a significant predictor of the degree of improvement of inter-Cβ distances (Fig 5.10a, 5.10d, 5.10f). This effect was recapitulated in the case of CDR-H3 with a moderate Pearson correlation coefficient of -0.54 (Fig 5.10c). The RMSPE values for CDR-H3 are greater than those of the minor loops, mirroring the observation that CDR-H3 is generally less accurately modelled by ABB2 (Fig 5.9).

From these results, we conclude not only that our model can be used to obtain more accurate inter-Cβ distances, but also that the estimated error of ABB2 can be used as a filtering method to choose which structures would benefit most from additional refinement after the initial model is generated using ABB2.

5.4.5 Energy minimisation using modified molecular mechanics force fields is unable to translate predicted distances to structural modifications

Based on the performance of our model, we aimed to produce refined ABB2 structural models using the predicted inter-Cβ distances as restraints and performing an energy minimisation using OpenMM. Since FW residues in ABB2 models tend to have very low RMSD values compared to ground truth structures, we chose to perform minimisation on CDR loops only, with one of two tested schemes allowing movement in all CDR loops and the other scheme restricting movement to CDR-H3 only (see Section 5.3). We found that neither of these schemes was able to consistently improve the RMSD of ABB2 models to their ground truth structures.

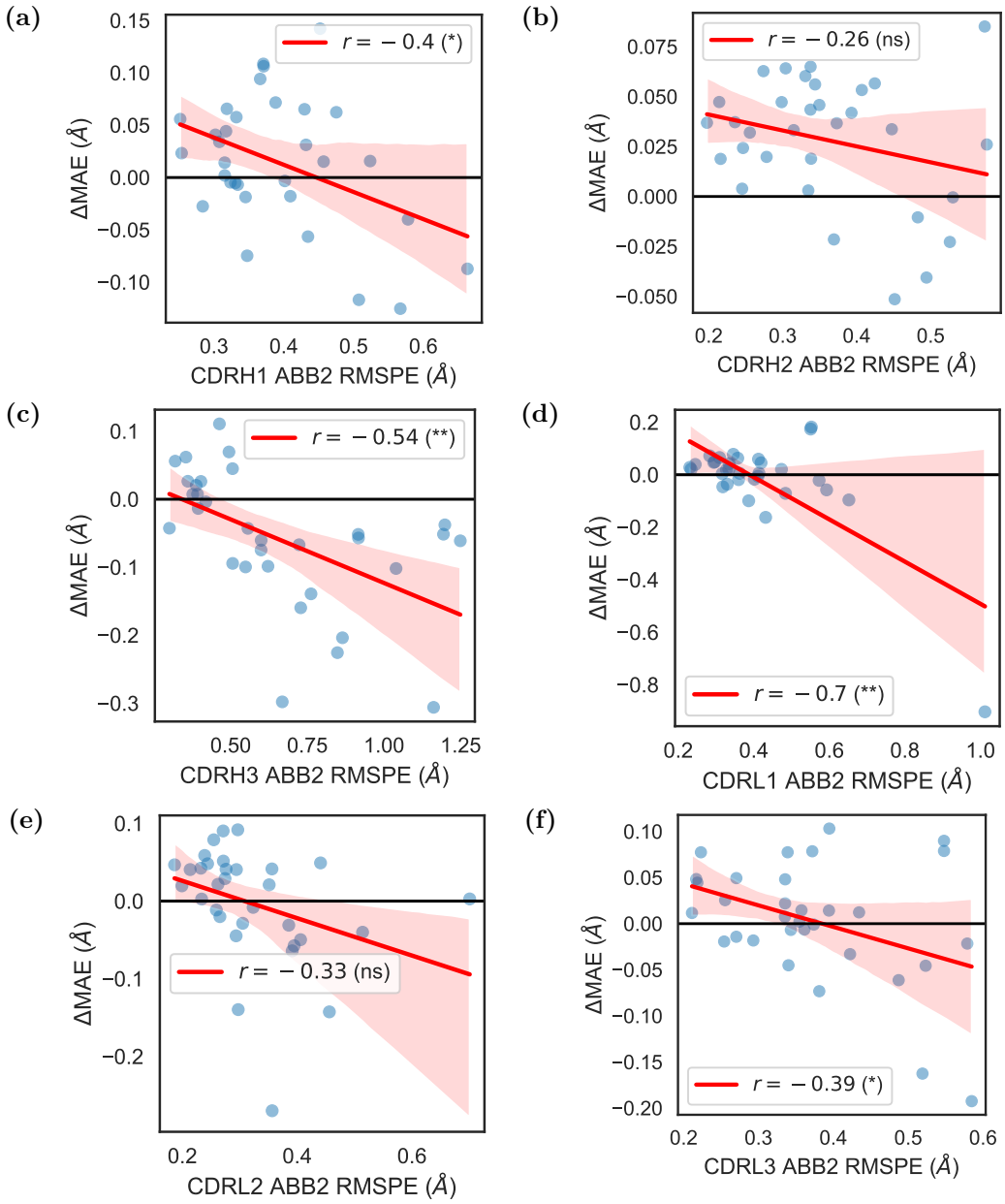


Figure 5.10: Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 predicted error. Shown are the absolute changes in MAE relative to ABB2 when refined inter-Cβ distances were inferred using our model, plotted against the ABB2 predicted root mean squared error of residues. The red lines show lines of best fit, with the shaded regions representing the 95% confidence interval for the regressions. Pearson correlation coefficients are reported for the linear regressions. (ns: not significant, *: $0.005 \leq p < 0.05$, **: $p < 0.005$) **(a-c)** show V_H CDR loops, **(d-f)** show V_L CDR loops. Framework residues and the entire F_V are shown in Fig C.4.

Figure 5.11 shows the change in RMSD using a restraint force of 100 kJ mol^{-1} nm^{-2} and using all edges involving CDR residues as distance restraints.

Refinement of the structures using OpenMM resulted in degradation in almost all minor CDR loops and improvement in 16 of 32 CDR-H3 structures, with an average change in CDR-H3 RMSD of -0.05 Å (Fig 5.11c). Our previous results suggest that structures with low RMSD values will not benefit from refinement. However, in the case of CDR-H3, our predicted distances improve the ABB2 MAE in almost all cases (Fig 5.9c), with strong dependence on the initial error. Based on these results, we conclude that refinement using a simple energy minimisation scheme using molecular mechanics and predicted distance as restraints is unable to improve starting structural models.

5.4.6 Inter-Cβ distances predicted by our models obey the triangle inequality

Based on our results using energy minimisation procedures to refine ABB2 structural models, we examined whether our model was generating nonphysical distances that violate the triangle inequality, which states that for any triangle the sum of the lengths of two sides must be less than the length of the remaining side. The triangle inequality has found use in protein structure prediction methods previously, with both AFM and ABB2 relying on it to place residues. Since any three residues which are all connected to each other via edges in our graph structures form a triangle, we tested the physical plausibility of our distance predictions by testing whether these predictions violate the inequality. Across the entire test set (a total of more than 64,000 edges) we found only one triangle that results in a violation of the inequality. Based on this, we conclude that our predicted distances are likely to be physically plausible, and that refinement using a technique other than simple energy minimisation using a modified molecular mechanics force field may be beneficial in translating these distance predictions into structural changes.

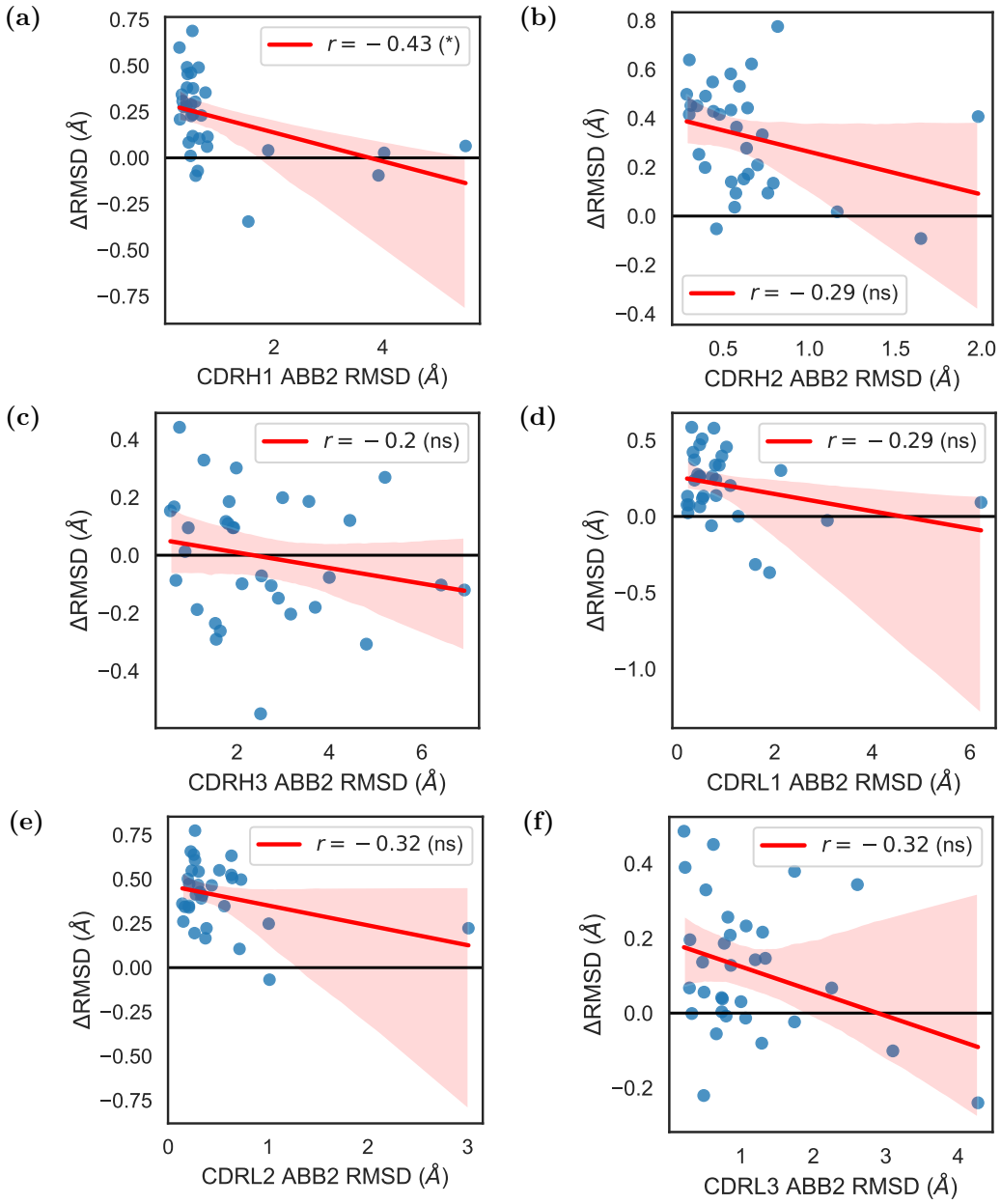


Figure 5.11: Changes in RMSD of refined structures compared to ABB2. We refined the ABB2 structural models using OpenMM by placing harmonic positional restraints on framework heavy atoms and adding harmonic distance restraints on CDR loop residues centered at the predicted distance for each edge. The red lines show lines of best fit, with the shaded regions representing the 95% confidence interval for the regressions. Pearson correlation coefficients are reported for the linear regressions. (ns: not significant, *: $0.005 \leq p < 0.05$, **: $p < 0.005$) **(a-c)** show V_H CDR loops, **(d-f)** show V_L CDR loops. Framework residues and the entire F_V are shown in Fig C.5.

5.5 Discussion and conclusions

The availability of accurate structural models is paramount to computer-aided discovery of antibody therapeutics. To this end, we have developed a method to rectify errors made by ABB2 during structural modelling of an antibody sequence. Our model is inspired by the recent work of Jing and Xu (Jing et al. 2021), but allows for the consideration of multiple chains and is specifically trained on antibody structures. Our results show that errors made by ABB2 are learnable and, therefore, can theoretically be rectified using a refinement approach.

We have shown that the error estimates provided by ABB2 can be used to identify structures that would benefit from refinement as well as structures that are likely to be degraded by it. Since millions of sequences are now routinely modelled during antibody discovery campaigns, being able to identify models where refinement should be added as an additional step is critical. It is also important that the refinement process is rapid and can feasibly be applied to entire sequence data sets. Our method predicts inter-Cβ distances for an antibody in under two seconds on average, which is similar to GNNRefine. However, the refinement pipeline of GNNRefine relies on conformational sampling using Rosetta, resulting in the entire process taking roughly 15 minutes for a protein of comparable size to an antibody (Jing et al. 2021) using eight CPUs. The Rosetta-based method is therefore not feasible for repertoire-scale modelling of antibodies. In order to decrease the amount of time needed for structure realisation from predicted inter-Cβ distances, we chose to implement the conformational sampling step in OpenMM, yet were unable to generate refined structures this way.

ABB2 produces models at comparable accuracy to AFM. Work by Heo *et al.* has shown that refinement using molecular mechanics approaches is unable to refine Alphafold2 structures (Heo et al. 2019). We argue that this may be due to simplifications in the force fields that are used (as discussed in Chapter 2) and that this simply may not allow the level of granularity needed to refine models beyond a certain accuracy. This is further supported by the results we obtained using a modified version of GNNRefine, where distance restraints were not included and

input structures were simply minimised using the internal force field. Here, we found that structures were significantly degraded across all regions of the antibody. Based on this, we hypothesise that when deep learning approaches are coupled with molecular mechanics approaches to predict crystal structure conformations, the assumptions made during force field parameterisation may prevent information from the former from being fully incorporated into the structural model. Further, structure prediction methods using deep learning are mainly trained on crystal structures and are therefore subject to crystal environments and cryogenic conditions, which have the potential to alter both conformation and dynamics (Faraggi et al. 2018; Fernández-Quintero et al. 2020b). On the other hand, force fields used for molecular mechanics are commonly designed to mimic solution-state properties. As such, we hypothesise that alternative deep learning architectures that explicitly modify and refine atomic positions may be useful in future approaches to antibody structural model refinement. The bottlenecks in current approaches to structure refinement are conformational searches that are performed during the process, and refinement algorithms that do not rely on this step would also benefit from the improved speeds necessary for repertoire-scale modelling.

6

Discussion, Conclusions and Future Work

Contents

6.1 Discussion and conclusions 155
6.2 Future work 158
6.3 Concluding remarks 159

6.1 Discussion and conclusions

In this DPhil thesis, we have reported work pertaining to both the physical and structural properties of antibodies. In Chapter 3, we have shown that the binding rate of antibodies is controlled by the flexibility of the CDR-H3 loop. Further, we have shown that the surface properties of antibodies can exhibit drastic ranges within conformational ensembles, showing that insight into these ensembles and consideration of antibodies as dynamic rather than static structures is instrumental for a complete understanding of their biophysical properties that impact binding affinity and developability. This raises important considerations for the antibody design process, as a shift towards an understanding of antibody structure and dynamics, rather than just structure, will likely be necessary to enable rational engineering of a number of biopysical properties. Indeed, we have found that static

analyses of antibodies, for example, do not allow us to understand their binding rates. This will likely be complicated by the fact that the epitope should then also be considered as a dynamic entity in many cases, especially as antibodies preferentially target coil-based surfaces.

In Chapter 4, we provide evidence in support of a long-standing hypothesis in the field, namely that the CDR-H3 conformation is influenced by the minor loops that surround it. We found that this is due to the formation of a steric cage around the CDR-H3 loop, which in turn restricts its configurational freedom to certain sub-regions, thereby fine-tuning its structure and flexibility. Interestingly, our results indicate that this is primarily mediated by the V_L minor loops, rather than by the V_H minor loops, and that specific canonical forms have preferences for specific sub-regions of CDR-H3 conformational space and levels of CDR-H3 flexibility.

Interpreting the results of Chapter 3 in light of the caged CDR-H3 model, it should then follow that the canonical forms of CDR-L1 and CDR-L3 should be modulators of binding rates. We argue that minor loop conformations, V-gene lineages, and light-chain subtypes may all exert control on the binding rate of antibodies via their effects on CDR-H3 flexibility. A smaller steric cage, which leads to a less flexible CDR-H3 loop by constricting it, should also result in a faster binding rate and more stable surface properties due to the smaller number of possible conformations of the CDR-H3 loop. We also note that methods used to probe conformational ensembles in Chapter 3 were concerned mainly with sampling CDR-H3 conformational diversity. As such, we hypothesise that techniques that allow more complete probing of flexibility of the antibody may reveal direct links between minor loop conformational ensembles and binding rates.

Currently, conformational ensembles of a protein are mainly generated by deriving physically plausible conformations of the protein, beginning from a starting structure. As such, a crucial step in understanding the dynamics of an antibody for which no experimental structure is available is to first make an accurate prediction of its static structure. In Chapter 5 we detail the development of a tool to improve the accuracy of ABB2, a current state-of-the-art tool in antibody

structure prediction. We base our model heavily on the GNNRefine architecture and show that training a GNN using antibody structures allows us to learn the errors that ABB2 makes, which in turn allows us to improve the predicted inter-residue distances within an antibody. While our refined predictions improve upon inaccurate naive predictions made by ABB2 (which can be identified *a priori*), we were unable to translate these into structural refinements using modified molecular mechanics force fields, and hypothesise that alternate model architectures may improve upon our approach by explicitly updating the atomic coordinates of the input structure during model inference.

Lastly, we pose a question regarding how the accuracy of antibody structure prediction algorithms to model F_V regions are evaluated: given that antibodies are dynamic molecules and that a number of antibodies are demonstrably stable in multiple conformations even in crystal structures, should the per sequence accuracy be evaluated on the target conformation or any conformation of this sequence? When training such models, the same antibody can appear in the training set multiple times in different conformations. Similarly, a structure in the test set presents only one conformation of an antibody that may appear multiple times in the PDB, as discussed in Chapter 5. The model being evaluated may well be modelling a physically feasible conformation (although the ultimate goal would be to rapidly predict the conformational ensemble of the antibody along with their relative probabilities), given that the task is to predict any conformation of this antibody, yet this prediction may not be the one associated with the target crystal structure in this case. As such, we propose that it may be a more biologically accurate approach to label a prediction as being correct if it matches any conformation of this antibody in the PDB. Considering the free energy landscape within which these conformations reside (which was discussed in Chapter 1), we propose that the aim should be to identify any pose that is highly stable in the given context. Such a paradigm shift will likely evolve naturally in the future, as the success of AF2 has enabled the structure prediction community to focus on the next frontier: the prediction of multiple conformations. We note that a current potential flaw

in this way of evaluating predictions is coverage in the PDB. With the current amount of antibody structural data in the PDB, it is likely that many antibodies, of which there possibly exist multiple stable crystal conformations, have been sampled only once, in which case a prediction of an alternate but physically feasible pose would still be marked as wrong. Consequently, we argue that genuine evaluation of multi-pose antibodies will require both algorithmic developments that allow multiple conformations and the expansion of available structural data in the PDB.

6.2 Future work

In Chapter 3, the major challenge we encountered was the lack of publicly available data. Therefore, the obvious first extension to this project would be to repeat analyses of binding rates using larger data sets. Since our approach using ABB2 (Abanades et al. 2022), AbLooper (Abanades et al. 2021), and Sphinx (Marks et al. 2017b) does not rely on accurate knowledge of the CDR-H3 loop, and since the remainder of the antibody can be reliably modelled using state-of-the art methods, further data need not have both a structure and kinetic measurements, but would only require kinetic measurements and the sequence of the antibody. This data is much easier to generate, and can be scraped from publicly available sources in much higher numbers. Further, as discussed in Section 6.1, it is likely that the flexibility of the minor loops may influence the binding rates of antibodies directly. Probing this is likely to be much more challenging than probing CDR-H3 flexibility, but could be approached using enhanced sampling simulations such as bias exchange metadynamics (Piana et al. 2007), wherein multiple loops can be biased.

Similar bias exchange metadynamics simulations, where each loop is biased in parallel simulations that exchange configurations at random, would provide insights that would expand upon the results presented in Chapter 4. In our work, we limit ourselves to interpretation of the effect of the steric cage when minor loops do not change conformation, since we do not observe these changes in our simulations. However, the ability of minor loops to change conformation, given enough sampling, has been documented in the literature (Fernández-Quintero et al.

2020b). As such, a natural extension to this work would be to analyse the change in CDR-H3 conformation and diversity in response to changes in the shape and size of the steric cage caused by shifts in minor loop conformation.

Lastly, we propose two extensions to the project detailed in Chapter 5. A first extension would be a study to understand why our minimisation procedure failed to produce refined structural models, given that the harmonic restraints introduced were based on demonstrably improved inter-Cβ distances. Based on the hypothesis that this may be due to inaccuracies in the force field used, due to necessary approximations during its parameterisation, an obvious first step would be to explore alternate force fields that capture the underlying physics of protein structure more accurately, such as the Amoeba force field (Ponder et al. 2010). Further, the harmonic restraint functions may not present the optimal way of incorporating these distance restraints, and alternate formulations such as that used in tools like trRosetta (Du et al. 2021), which also rely on inter-residue distance potentials that resemble Lenard-Jones potentials in shape, may be tested. The second extension to this project would be to employ an architecture that foregoes the restraint-based minimisation entirely, such as an EGNN (Satorras et al. 2021), or an extension thereof. This approach would involve explicitly updating the positions of atoms in the structure during inference. Based on the conclusions that we draw from our work in Chapter 4, we argue that any antibody structure prediction tool will necessitate that the entire F_V region, rather than single parts thereof, be considered during the modelling process if such a tool is to generate as accurate a prediction as possible.

6.3 Concluding remarks

An understanding of antibody structure and function is necessary for both antibody and immunogen design. Traditionally, antibodies have been regarded as a class of modular, static proteins, with CDR loops being considered as rigid structures existing in isolation. Our work contributes to an ongoing paradigm shift, in which we re-imagine antibodies both as dynamic molecules, with multiple states of CDR

loops being occupied, and as molecules where the exact structures and structural ensembles of an antibody are results of the protein as a holistic unit. We predict that future work in this line will further demonstrate the richness of this approach in being able to predict and rationally engineer their properties and hope that this will allow us to rapidly produce both next-generation therapeutics and vaccines.

Appendices

A

Supplementary Material for Chapter 3

A.1 Kinetic data set

Table A.1: Rate constants of the kinetic data set, along with PDB accession codes. An antibody is marked as paired, whereas a single-domain antiboyd is marked as being unpaired.

PDB Code	Paired?	$\log(k_{on})$	$\log(k_{off})$	k_{on} ($M^{-1}\,s^{-1}$)	k_{off} (s^{-1})
4zpt	TRUE	5.6911	-2.3116	4.91E+05	4.88E-03
4liq	TRUE	5.8513	-3.8861	7.10E+05	1.30E-04
4f9l	TRUE	5.0828	-2.0980	1.21E+05	7.98E-03
2xtj	TRUE	4.8579	-3.5003	7.21E+04	3.16E-04
5w08	TRUE	4.4942	-3.0000	3.12E+04	1.00E-03
4dke	TRUE	6.5315	-3.3768	3.40E+06	4.20E-04
6a0z	TRUE	4.3674	-2.9172	2.33E+04	1.21E-03
2h9g	TRUE	5.9542	-2.3979	9.00E+05	4.00E-03
3mxw	TRUE	5.7050	-3.4535	5.07E+05	3.52E-04
5vag	TRUE	4.8519	-4.7077	7.11E+04	1.96E-05
4yk4	TRUE	4.7427	-1.6716	5.53E+04	2.13E-02
4eiz	FALSE	6.4771	-2.0458	3.00E+06	9.00E-03
6fe4	FALSE	5.2900	-2.7212	1.95E+05	1.90E-03

5zxv	TRUE	5.1903	-2.0487	1.55E+05	8.94E-03
2cmr	TRUE	6.3560	-3.9208	2.27E+06	1.20E-04
4m8q	TRUE	5.6590	-2.7447	4.56E+05	1.80E-03
6flb	TRUE	5.9523	-3.0830	8.96E+05	8.26E-04
6b0g	TRUE	4.3424	-3.4202	2.20E+04	3.80E-04
4zff	TRUE	5.5224	-2.7696	3.33E+05	1.70E-03
4lu5	TRUE	6.2068	-3.9747	1.61E+06	1.06E-04
5xj4	TRUE	5.6314	-3.5452	4.28E+05	2.85E-04
5sx5	TRUE	6.8195	-2.0000	6.60E+06	1.00E-02
5sx4	TRUE	6.3222	-3.0000	2.10E+06	1.00E-03
2vxq	TRUE	5.4713	-4.4711	2.96E+05	3.38E-05
5i5k	TRUE	6.1139	-4.6383	1.30E+06	2.30E-05
3p0y	TRUE	5.2648	-2.4101	1.84E+05	3.89E-03
1p2c	TRUE	6.8567	-3.1549	7.19E+06	7.00E-04
5w0k	TRUE	4.3428	-2.8210	2.20E+04	1.51E-03
1sy6	TRUE	5.2945	-0.3279	1.97E+05	4.70E-01
3t2n	TRUE	4.2718	-3.1290	1.87E+04	7.43E-04
4k2u	TRUE	5.6503	-5.1662	4.47E+05	6.82E-06
5xj3	TRUE	5.5832	-2.1574	3.83E+05	6.96E-03
4jkp	TRUE	4.1761	-3.0969	1.50E+04	8.00E-04
5c0s	TRUE	5.5563	-2.5686	3.60E+05	2.70E-03
3rkd	TRUE	5.5843	-2.6440	3.84E+05	2.27E-03
5whk	TRUE	5.2330	-3.5114	1.71E+05	3.08E-04
2xra	TRUE	5.3424	-3.1385	2.20E+05	7.27E-04
4f9p	TRUE	5.0334	-2.7878	1.08E+05	1.63E-03
3l5x	TRUE	6.8109	-2.7595	6.47E+06	1.74E-03
4dkf	TRUE	6.1761	-2.3468	1.50E+06	4.50E-03
5xwd	TRUE	6.3962	-1.8327	2.49E+06	1.47E-02
3ru8	TRUE	6.4771	-1.0000	3.00E+06	1.00E-01
3ux9	TRUE	5.3927	-3.7471	2.47E+05	1.79E-04

3nh7	TRUE	4.9445	-2.7696	8.80E+04	1.70E-03
5fcu	TRUE	5.2041	-3.0250	1.60E+05	9.44E-04
3idy	TRUE	4.0828	-4.3107	1.21E+04	4.89E-05
5vob	TRUE	5.6405	-4.1972	4.37E+05	6.35E-05
5kvf	TRUE	5.4983	-1.9431	3.15E+05	1.14E-02
4f37	TRUE	5.2553	-3.0000	1.80E+05	1.00E-03
6ii8	TRUE	4.1461	-2.2840	1.40E+04	5.20E-03
5o2u	FALSE	5.5682	-3.5376	3.70E+05	2.90E-04
4jpk	TRUE	4.1951	-3.1653	1.57E+04	6.83E-04
3w2d	TRUE	4.0286	-5.2088	1.07E+04	6.18E-06
4zfg	TRUE	6.0531	-2.2111	1.13E+06	6.15E-03
6ejm	TRUE	5.7404	-3.3279	5.50E+05	4.70E-04
5usl	TRUE	5.8089	-4.9788	6.44E+05	1.05E-05
4d9r	TRUE	7.9138	-3.4815	8.20E+07	3.30E-04
4ffz	TRUE	5.4082	-0.9731	2.56E+05	1.06E-01
4k3j	TRUE	5.0414	-3.7212	1.10E+05	1.90E-04
6flc	TRUE	5.7952	-0.9586	6.24E+05	1.10E-01
4hj0	TRUE	5.2014	-2.4330	1.59E+05	3.69E-03
3wd5	TRUE	5.6794	-4.2588	4.78E+05	5.51E-05
5hys	TRUE	7.4362	-2.9747	2.73E+07	1.06E-03
6b08	TRUE	5.3222	-2.4318	2.10E+05	3.70E-03
1tzi	TRUE	6.0000	-2.0000	1.00E+06	1.00E-02
4f3f	TRUE	5.4928	-3.6517	3.11E+05	2.23E-04
5kw9	TRUE	5.3096	-2.3969	2.04E+05	4.01E-03
5nmv	TRUE	5.3838	-3.6308	2.42E+05	2.34E-04
5sy8	TRUE	6.5527	-4.3979	3.57E+06	4.00E-05
3vg9	TRUE	4.4216	-3.9355	2.64E+04	1.16E-04
2vwe	TRUE	6.3560	-3.5867	2.27E+06	2.59E-04
2b2x	TRUE	5.5682	-3.1549	3.70E+05	7.00E-04
4k9e	TRUE	6.3802	-4.8239	2.40E+06	1.50E-05

5o0w	FALSE	6.2718	-3.8601	1.87E+06	1.38E-04
4qci	TRUE	7.1038	-3.4413	1.27E+07	3.62E-04
5l6y	TRUE	6.7482	-3.4949	5.60E+06	3.20E-04
2uzi	TRUE	6.5276	-2.9318	3.37E+06	1.17E-03
6mlk	TRUE	5.4771	-2.3188	3.00E+05	4.80E-03
3lev	TRUE	6.2222	-3.0000	1.67E+06	1.00E-03
6cmg	TRUE	4.6599	-2.9031	4.57E+04	1.25E-03
3gjf	TRUE	4.8675	-2.3307	7.37E+04	4.67E-03
4qti	TRUE	5.1492	-2.6421	1.41E+05	2.28E-03
3k2u	TRUE	6.0334	-3.7570	1.08E+06	1.75E-04
5vyf	TRUE	5.4099	-3.6289	2.57E+05	2.35E-04
2xqb	TRUE	6.6345	-3.7033	4.31E+06	1.98E-04
5ezo	TRUE	5.4829	-3.5157	3.04E+05	3.05E-04
5c6t	TRUE	5.4698	-4.4260	2.95E+05	3.75E-05
4cni	TRUE	5.8976	-5.1135	7.90E+05	7.70E-06
2ybr	TRUE	5.7559	-3.9101	5.70E+05	1.23E-04
3ngb	TRUE	4.5832	-3.6576	3.83E+04	2.20E-04
5t6l	TRUE	6.5011	-4.0438	3.17E+06	9.04E-05
5u4m	FALSE	5.8893	-3.5186	7.75E+05	3.03E-04
5lqb	TRUE	4.6656	-4.1884	4.63E+04	6.48E-05
5hgg	FALSE	6.0000	-4.2676	1.00E+06	5.40E-05
5t5b	TRUE	6.1139	-3.7959	1.30E+06	1.60E-04
4wem	FALSE	3.6628	-1.5031	4.60E+03	3.14E-02
3lhp	TRUE	6.5011	-4.0590	3.17E+06	8.73E-05
1iqd	TRUE	5.8692	-5.0000	7.40E+05	1.00E-05
2dd8	TRUE	5.4771	-2.2147	3.00E+05	6.10E-03
3etb	TRUE	6.2041	-4.3188	1.60E+06	4.80E-05
4iof	TRUE	5.3385	-2.2314	2.18E+05	5.87E-03
6eyo	TRUE	6.6990	-4.2218	5.00E+06	6.00E-05
4g3y	TRUE	4.5441	-3.5229	3.50E+04	3.00E-04

5igx	TRUE	4.1461	-2.2518	1.40E+04	5.60E-03
4jlr	TRUE	6.6010	-3.9245	3.99E+06	1.19E-04
1e6j	TRUE	5.5441	-2.9208	3.50E+05	1.20E-03
5utz	TRUE	6.5775	-2.1481	3.78E+06	7.11E-03
4z5r	TRUE	6.3054	-4.4306	2.02E+06	3.71E-05
3idx	TRUE	4.0828	-4.3107	1.21E+04	4.89E-05
6ii4	TRUE	5.4472	-2.3279	2.80E+05	4.70E-03
4m62	TRUE	5.7709	-2.8182	5.90E+05	1.52E-03
6mhr	TRUE	5.7993	-1.8539	6.30E+05	1.40E-02
5kvg	TRUE	5.0607	-3.1079	1.15E+05	7.80E-04
5t29	TRUE	5.7474	-3.6757	5.59E+05	2.11E-04
5tfw	TRUE	6.4472	-4.3372	2.80E+06	4.60E-05
6bit	TRUE	6.0043	-2.5850	1.01E+06	2.60E-03
3mj9	TRUE	5.1139	-3.0000	1.30E+05	1.00E-03
3n85	TRUE	6.1139	-3.0088	1.30E+06	9.80E-04
5jq6	TRUE	5.7738	-3.3098	5.94E+05	4.90E-04
3hae	TRUE	4.9079	-2.4413	8.09E+04	3.62E-03
5kvd	TRUE	5.4298	-1.4868	2.69E+05	3.26E-02
3hi1	TRUE	4.8808	-3.0716	7.60E+04	8.48E-04
5sv3	FALSE	5.7419	-3.4609	5.52E+05	3.46E-04
5bvp	TRUE	6.0414	-4.4815	1.10E+06	3.30E-05
5my6	FALSE	5.3444	-3.0560	2.21E+05	8.79E-04
6b9j	TRUE	5.8621	-3.6108	7.28E+05	2.45E-04
6b0e	TRUE	4.3617	-3.0315	2.30E+04	9.30E-04
4zs7	TRUE	6.3181	-4.5575	2.08E+06	2.77E-05
5b71	TRUE	6.1584	-3.6615	1.44E+06	2.18E-04
6ii9	TRUE	5.1461	-2.5686	1.40E+05	2.70E-03
6azz	TRUE	5.4771	-2.6778	3.00E+05	2.10E-03
4weu	FALSE	4.7202	-0.5214	5.25E+04	3.01E-01
1cz8	TRUE	4.7324	-5.2291	5.40E+04	5.90E-06

1bj1	TRUE	4.6128	-3.8539	4.10E+04	1.40E-04
5kov	TRUE	5.8162	-2.9101	6.55E+05	1.23E-03
4k94	TRUE	6.0000	-3.1805	1.00E+06	6.60E-04
3ogo	FALSE	5.8854	-3.7595	7.68E+05	1.74E-04
3l95	TRUE	4.9395	-3.6576	8.70E+04	2.20E-04
5e8e	TRUE	5.6721	-2.7212	4.70E+05	1.90E-03
4wen	FALSE	4.1461	-2.8539	1.40E+04	1.40E-03
5bo1	TRUE	5.3598	-3.7447	2.29E+05	1.80E-04
5ush	TRUE	5.8561	-2.9706	7.18E+05	1.07E-03
3qwo	TRUE	5.8943	-1.1681	7.84E+05	6.79E-02
5boz	FALSE	5.5441	-3.6198	3.50E+05	2.40E-04
5gzo	TRUE	4.1467	-2.6396	1.40E+04	2.29E-03
5ip4	FALSE	5.2900	-1.0550	1.95E+05	8.81E-02
6ejg	TRUE	5.6435	-3.4437	4.40E+05	3.60E-04
3p9w	FALSE	5.4914	-2.3188	3.10E+05	4.80E-03
3eba	FALSE	6.2122	-2.4145	1.63E+06	3.85E-03
4hkx	TRUE	5.3747	-1.0655	2.37E+05	8.60E-02
4uu9	TRUE	7.2405	-3.6003	1.74E+07	2.51E-04
5kve	TRUE	5.1703	-2.4202	1.48E+05	3.80E-03
4dtg	TRUE	6.6513	-2.9957	4.48E+06	1.01E-03
4pgj	FALSE	4.0000	-2.5850	1.00E+04	2.60E-03
4lqf	TRUE	6.2529	-3.9393	1.79E+06	1.15E-04
4ffy	TRUE	5.2967	-2.4486	1.98E+05	3.56E-03
4u6h	TRUE	5.7097	-4.3188	5.13E+05	4.80E-05
5dwu	TRUE	6.3010	-1.0177	2.00E+06	9.60E-02
3zkq	FALSE	6.4914	-2.4318	3.10E+06	3.70E-03
3fku	TRUE	5.8142	-4.0757	6.52E+05	8.40E-05
6mi2	TRUE	6.1461	-1.0269	1.40E+06	9.40E-02
4leo	TRUE	5.2529	-4.0545	1.79E+05	8.82E-05
4odx	TRUE	5.8000	-2.8861	6.31E+05	1.30E-03

4xi5	TRUE	5.6294	-3.6799	4.26E+05	2.09E-04
5wux	TRUE	6.2945	-4.2676	1.97E+06	5.40E-05
5w06	TRUE	5.1959	-4.6576	1.57E+05	2.20E-05
2ypv	TRUE	6.0128	-4.5243	1.03E+06	2.99E-05
6b0a	TRUE	5.3424	-3.0757	2.20E+05	8.40E-04
4m1g	TRUE	6.0374	-1.8069	1.09E+06	1.56E-02
5lhn	FALSE	5.0000	-2.9586	1.00E+05	1.10E-03
6ehg	FALSE	5.3222	-4.2924	2.10E+05	5.10E-05
5n88	FALSE	4.0828	-1.9101	1.21E+04	1.23E-02
4etq	TRUE	6.0334	-3.7167	1.08E+06	1.92E-04
3p11	TRUE	6.8802	-2.8477	7.59E+06	1.42E-03
4np4	TRUE	6.1818	-4.5482	1.52E+06	2.83E-05
5j56	FALSE	4.9590	-3.5376	9.10E+04	2.90E-04
5toj	FALSE	5.4771	-2.1675	3.00E+05	6.80E-03
4i3r	TRUE	4.4886	-3.4353	3.08E+04	3.67E-04
2qr0	TRUE	5.9542	-2.1549	9.00E+05	7.00E-03
5bk1	TRUE	6.0792	-2.8861	1.20E+06	1.30E-03
3skj	TRUE	5.7853	-3.3028	6.10E+05	4.98E-04
6fgb	TRUE	5.5682	-4.6576	3.70E+05	2.20E-05
3l5w	TRUE	6.9899	-3.2958	9.77E+06	5.06E-04
5j57	FALSE	5.3424	-5.3768	2.20E+05	4.20E-06
3v6z	TRUE	3.7709	-3.1612	5.90E+03	6.90E-04
3kr3	TRUE	6.2553	-4.0506	1.80E+06	8.90E-05
5j13	TRUE	5.5185	-4.7212	3.30E+05	1.90E-05
4d9q	TRUE	8.4472	-2.0706	2.80E+08	8.50E-03
1tzh	TRUE	5.4771	-3.3010	3.00E+05	5.00E-04
5yy5	TRUE	4.5575	-2.9101	3.61E+04	1.23E-03
4qex	TRUE	5.3711	-3.3645	2.35E+05	4.32E-04

A.2 Flags used to assess antibody-antigen interactions using Arpeggio

```
python ./arpeggio.py <pdb_file>
                     -mh -mf MMFF94
                     -mm ConjugateGradients -v
```

A.3 Flags used to model CDR-H3 loops using Sphinx

```
Sphinx --sphinx_version h3
       --ranking_method=ddfire
       --ignore <pdb_code>
       <path_to_CDRH3_fragment_library>
       <pdb_file>
       <loop_start> <loop_end>
       <loop_chain> <loop_sequence>
       <output_dir>
```

A.4 Supplementary figures

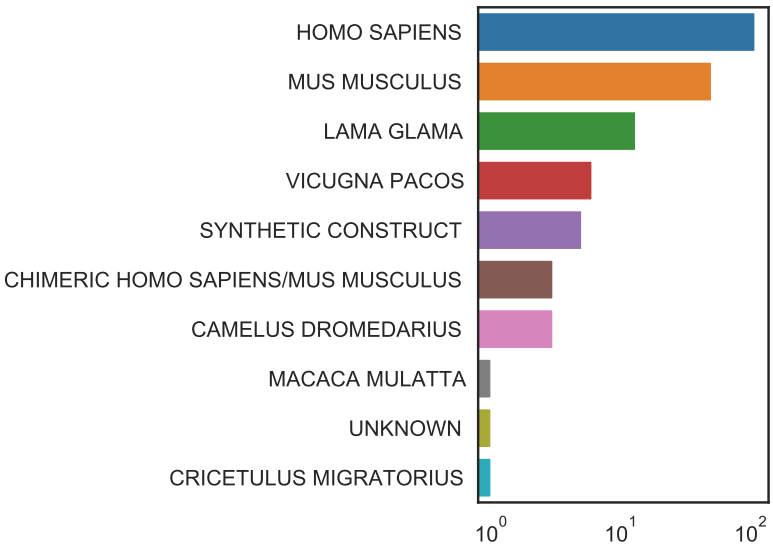


Figure A.1: Distribution of species in the kinetic data set according to their labels within SAbDab

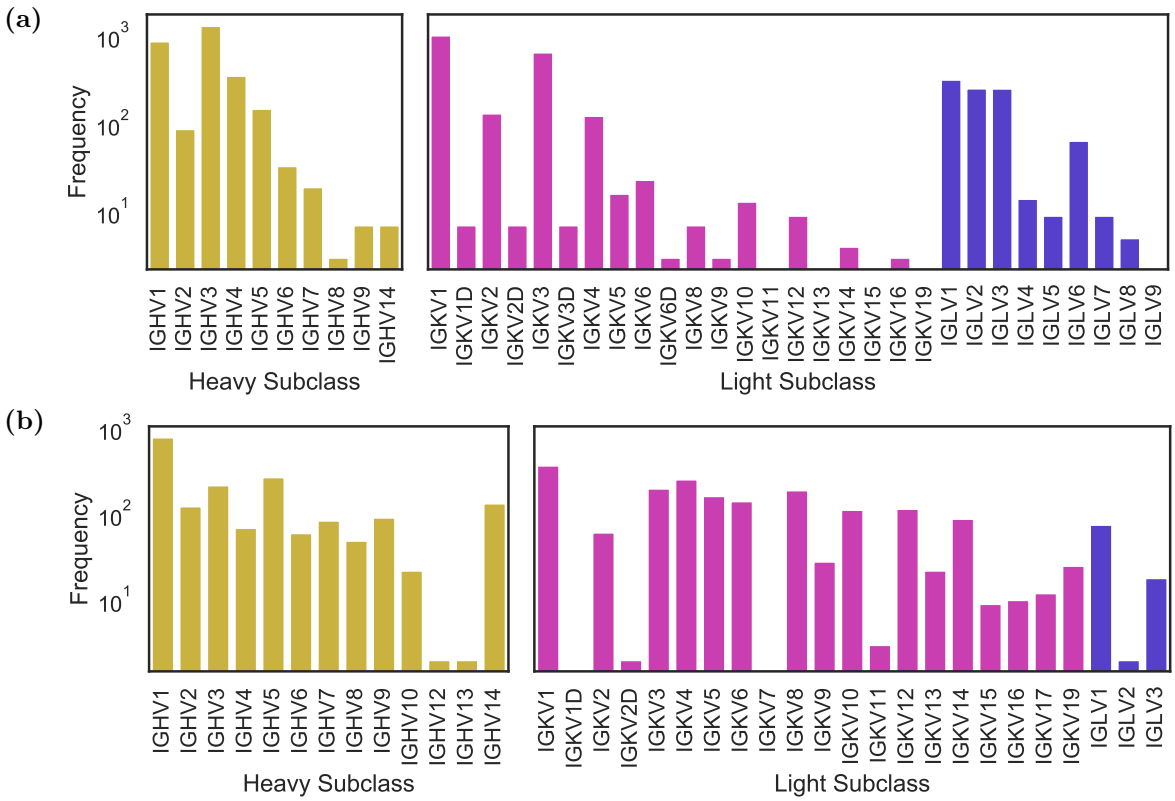


Figure A.2: Background distribution of genetic attributes across SAbDab. **(a)** shows the germline genes for all human antibodies in SAbDab and **(b)** shows the germline genes for all mouse antibodies in SAbDab.

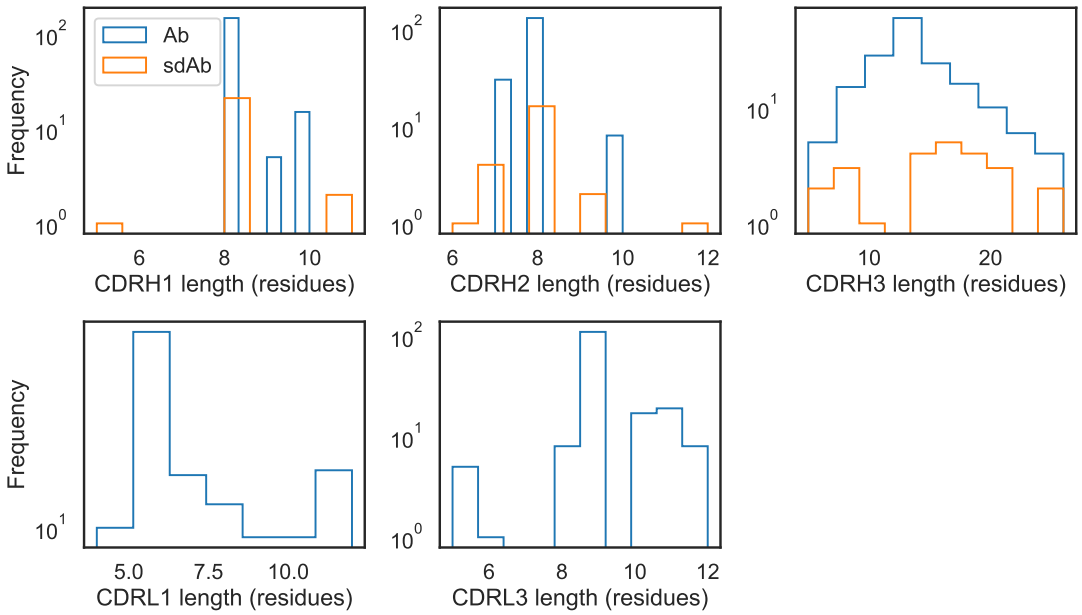


Figure A.3: Distribution of CDR loop lengths in the kinetic data set. Ab: paired antibodies, sdAb: single-domain antibodies/nanobodies.

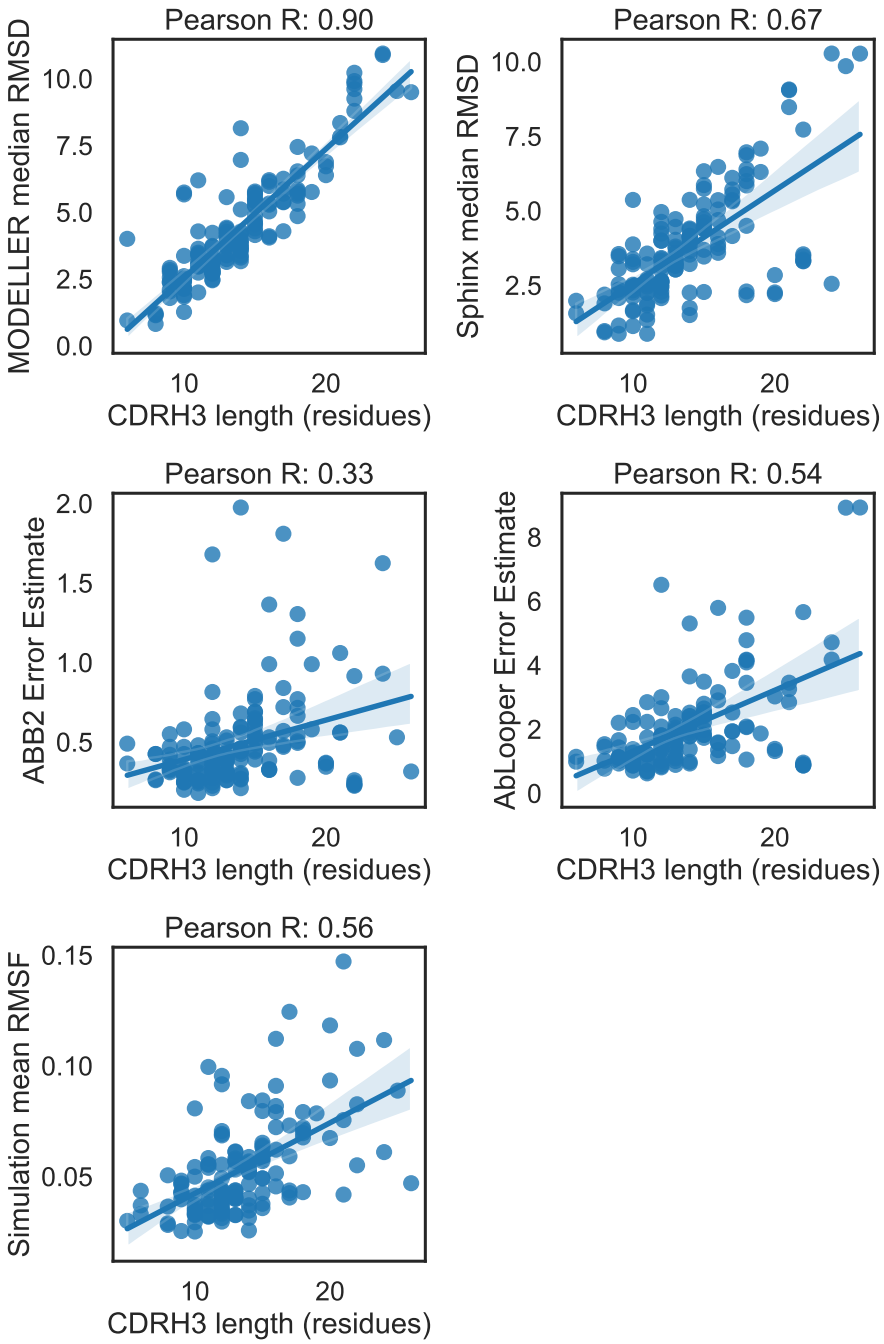


Figure A.4: Pearson correlation coefficients and scatter plots showing the relationship between predicted CDR-H3 flexibility by different methods and the length of the modelled loop.

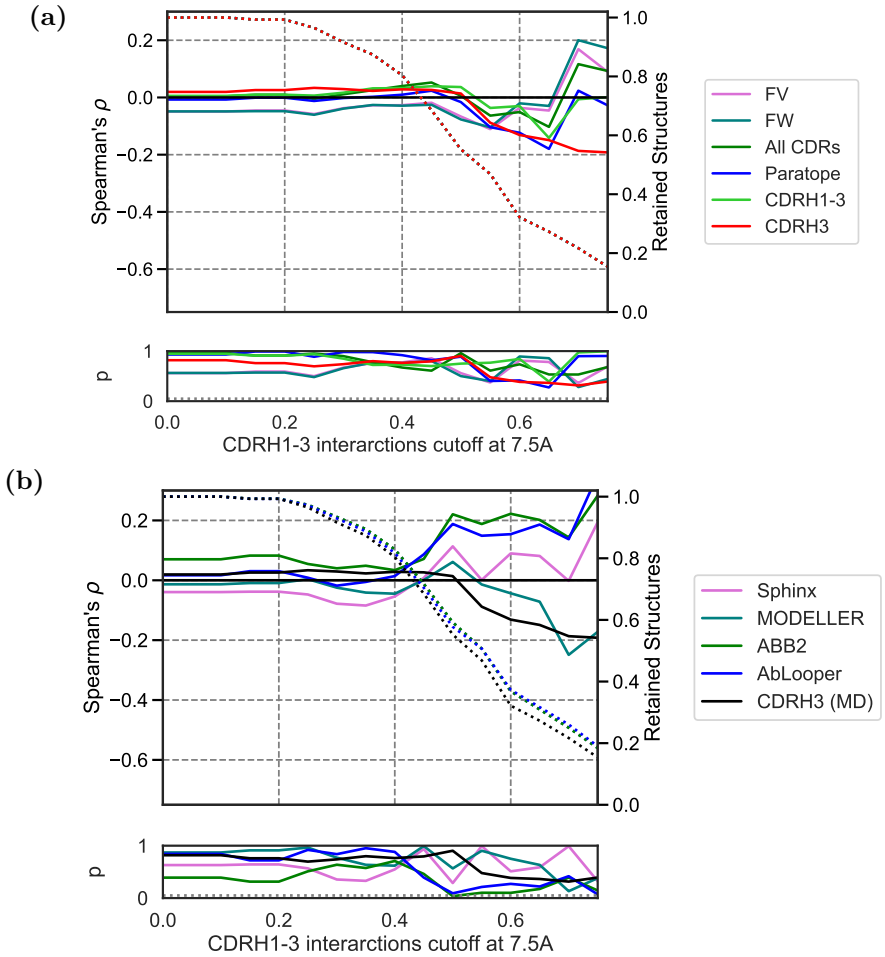


Figure A.5: Correlations between antibody flexibility as predicted through MD simulations and antibody modelling methods and unbinding rates. At each fnat_{VH} cutoff (CDR-H1-3 interactions cutoff at 7.5 Å), we filtered the kinetic data set and correlated loop lengths and $\log(k_{on})$. The fraction of structures from the original data set remaining at each cutoff is shown using dotted lines (right y-axis) while the Spearman rank correlation coefficient (ρ) is shown in cyan for each cutoff (left y-axis). **(a)** shows the correlations between unbinding rates and RMSF values of labelled regions of the antibody, with p-values for each correlation shown below (gray dashed line: 0.05). **(b)** shows the correlations between predicted flexibility of the CDR-H3 loop and binding rates using different methods. For Sphinx and MODELLER, the median pairwise RMSD is used as the predictive variable. For ABodyBuilder2 and AbLooper, the average predicted error across the loop is used for correlations. p-values for each correlation shown below (gray dashed line: 0.05)

CST	PDB Structure for Constant Domain Grafting (ID, (C_H1 chain, C_L chain))
Belimumab	5Y9K (HL)
Boserolimab	8DS5 (CB)
Burosumab	7VEN (BA)
Casdozokitug	7ZXK (HL)
Dostarlimab	7WSL (HL)
Erenumab	6UMH (HL)
Fepixnebart	5KN5 (AB)
Glenzocimab	7R58 (HL)
Golocdacimab	7R8U (HL)
Imalumab	6FOE (HL)
Quilizumab	3HR5 (HL)
Regdanvimab	7CM4 (HL)
Secukinumab	6WIO (AB)
Simaravibart	7SBU (HL)

Table A.2: Protein Data Bank IDs and chain IDs mapping to the coordinates used for C_H1/C_L domain grafting for each clinical-stage therapeutic (CST) used for assessment of TAP properties in conformational ensembles.

TAP Property	Amber Flag Region	Red Flag Region
PSH	$95.58 \leq \text{PSH} \leq 109.51$ $162.15 \leq \text{PSH} \leq 195.68$	$\text{PSH} < 95.58$ $\text{PSH} > 195.68$
PPC	$1.28 \leq \text{PPC} \leq 4.22$	$\text{PPC} > 4.22$
PNC	$1.94 \leq \text{PNC} \leq 4.42$	$\text{PNC} > 4.42$
SFvCSP	$-30.6 \leq \text{SFvCSP} \leq -6.0$	$\text{SFvCSP} < -30.6$

Table A.3: Values used as thresholds for TAP flags

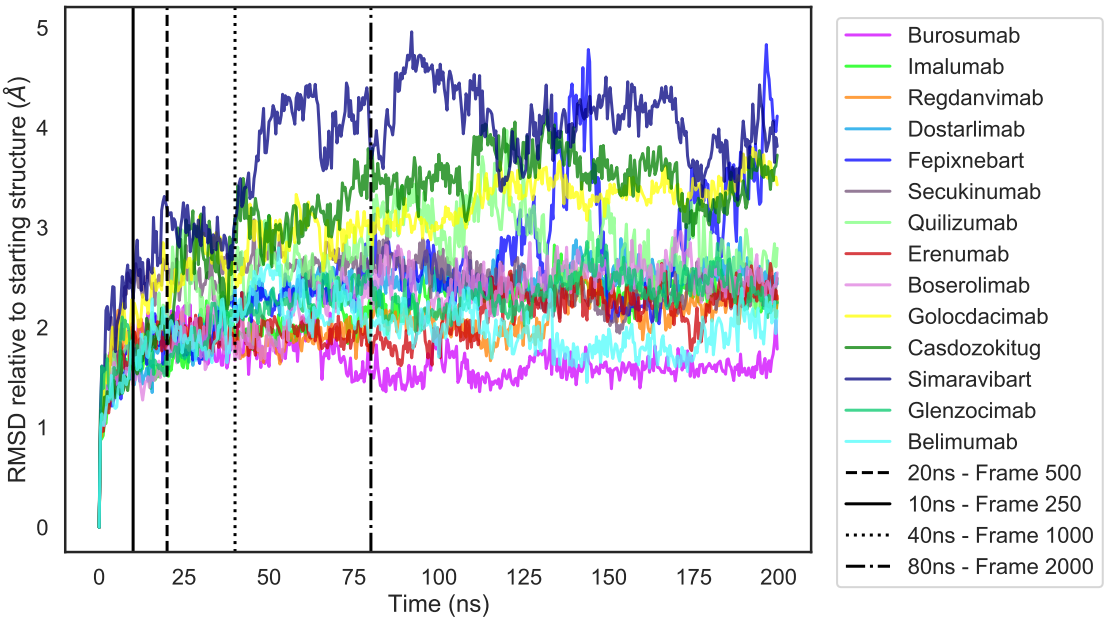


Figure A.6: RMSDs of assessed CSTs from the starting structure over the course of MD simulations. F_V heavy atoms in each frame were aligned to the starting structure, and RMSD values calculated using mdtraj. TAP properties were calculates for frames after 80 ns, as this was determined to be the timeframe required for full equilibration of all systems.

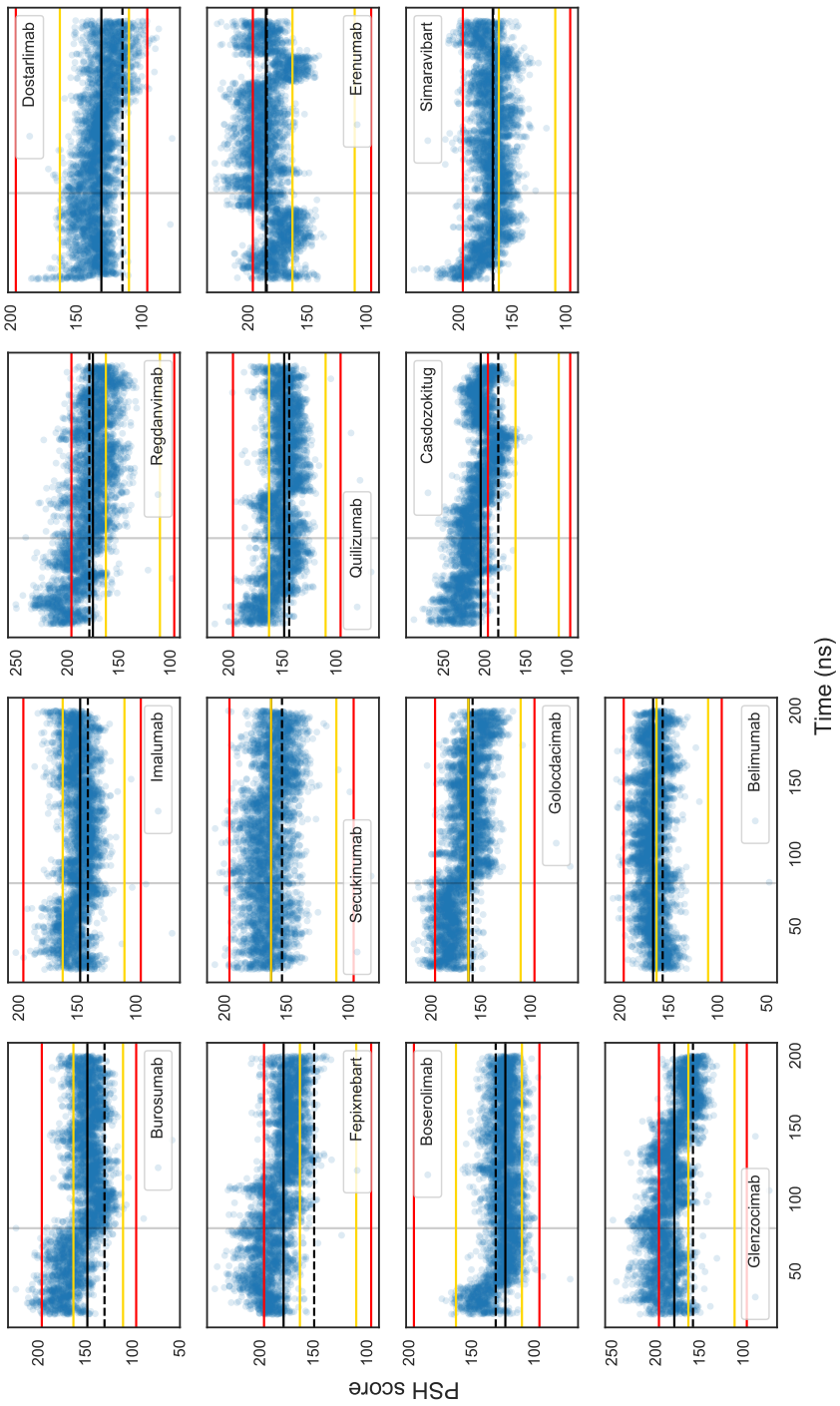


Figure A.7: Changes in surface hydrophobic patches of antibodies during simulation. The region between the amber lines corresponds to scores that would be assigned a green flag by TAP and the regions outside the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average score from three ABB2 models and the solid black line shows the average score over the final 120 ns of the trajectory (marked by the grey line).

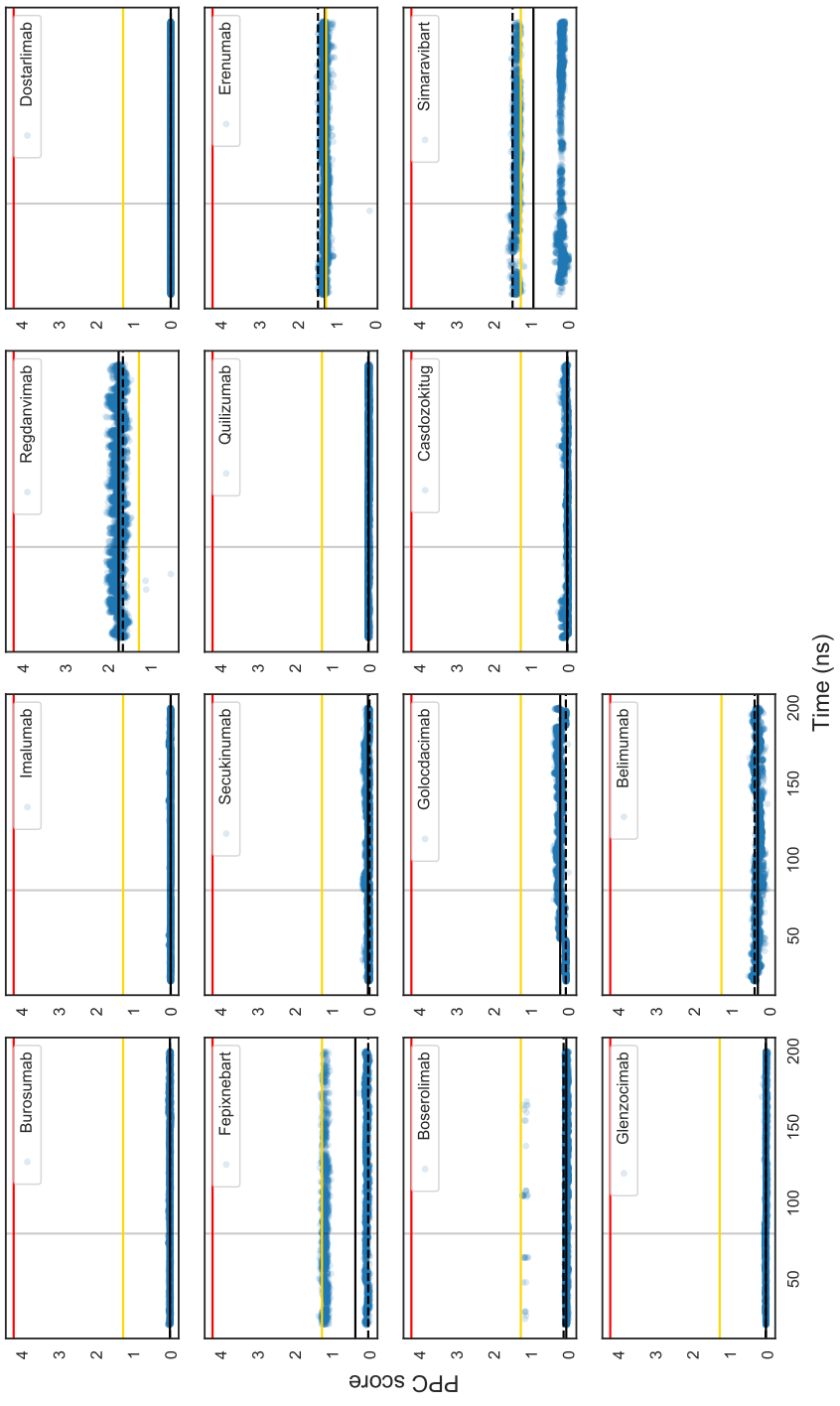


Figure A.8: Changes in surface positive patch (PPC) scores in of antibodies during simulation. The region under the amber lines corresponds to PPC scores that would be assigned a green flag by TAP and the regions above the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average PPC score from three ABB2 models and the solid black line shows the average PPC score over the final 120 ns of the trajectory (marked by the grey line).

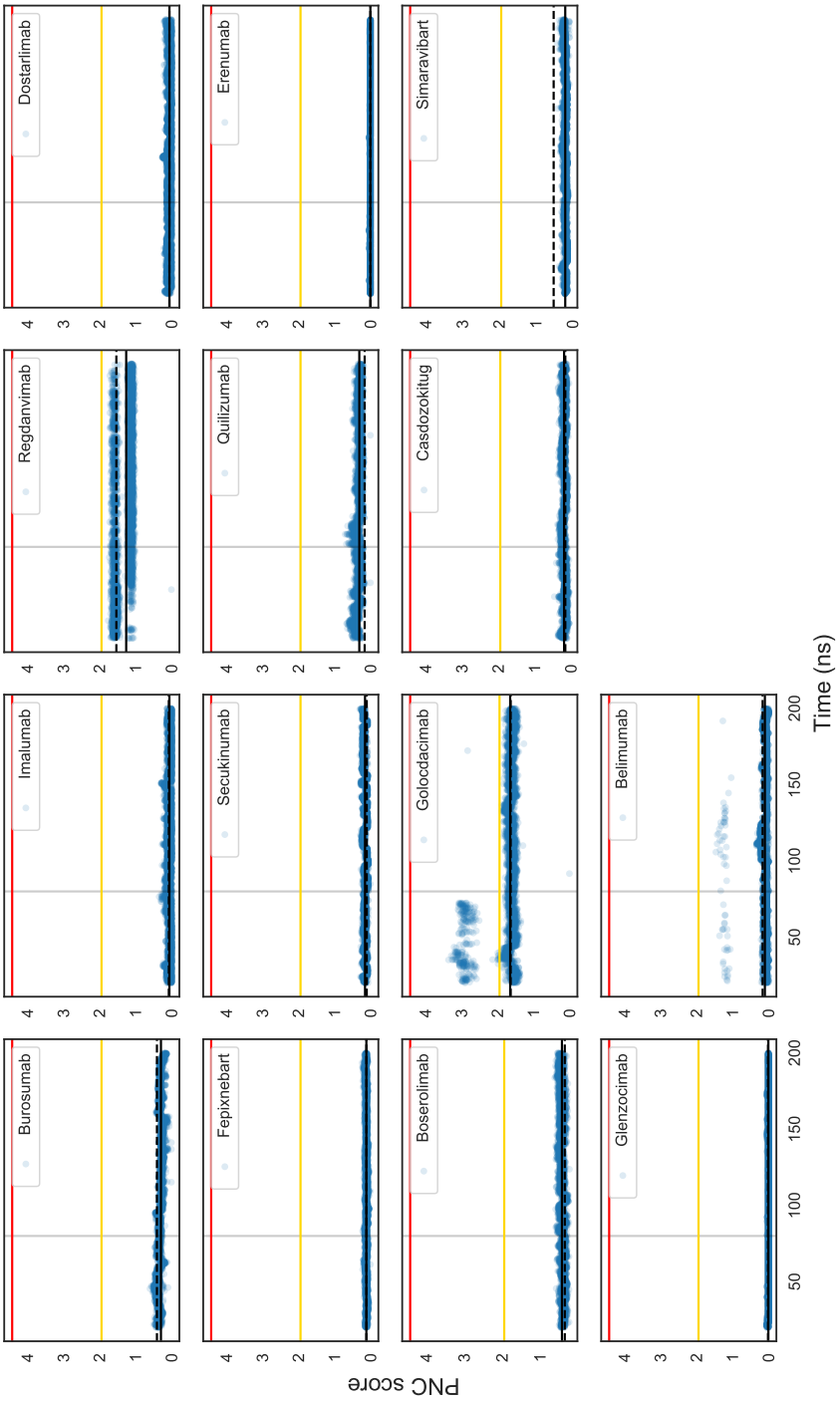


Figure A.9: Changes in surface negative patch (PNC) scores in of antibodies during simulation. The region under the amber lines corresponds to PNC scores that would be assigned a green flag by TAP and the regions above the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average PNC score from three ABB2 models and the solid black line shows the average PNC score over the final 120 ns of the trajectory (marked by the grey line).

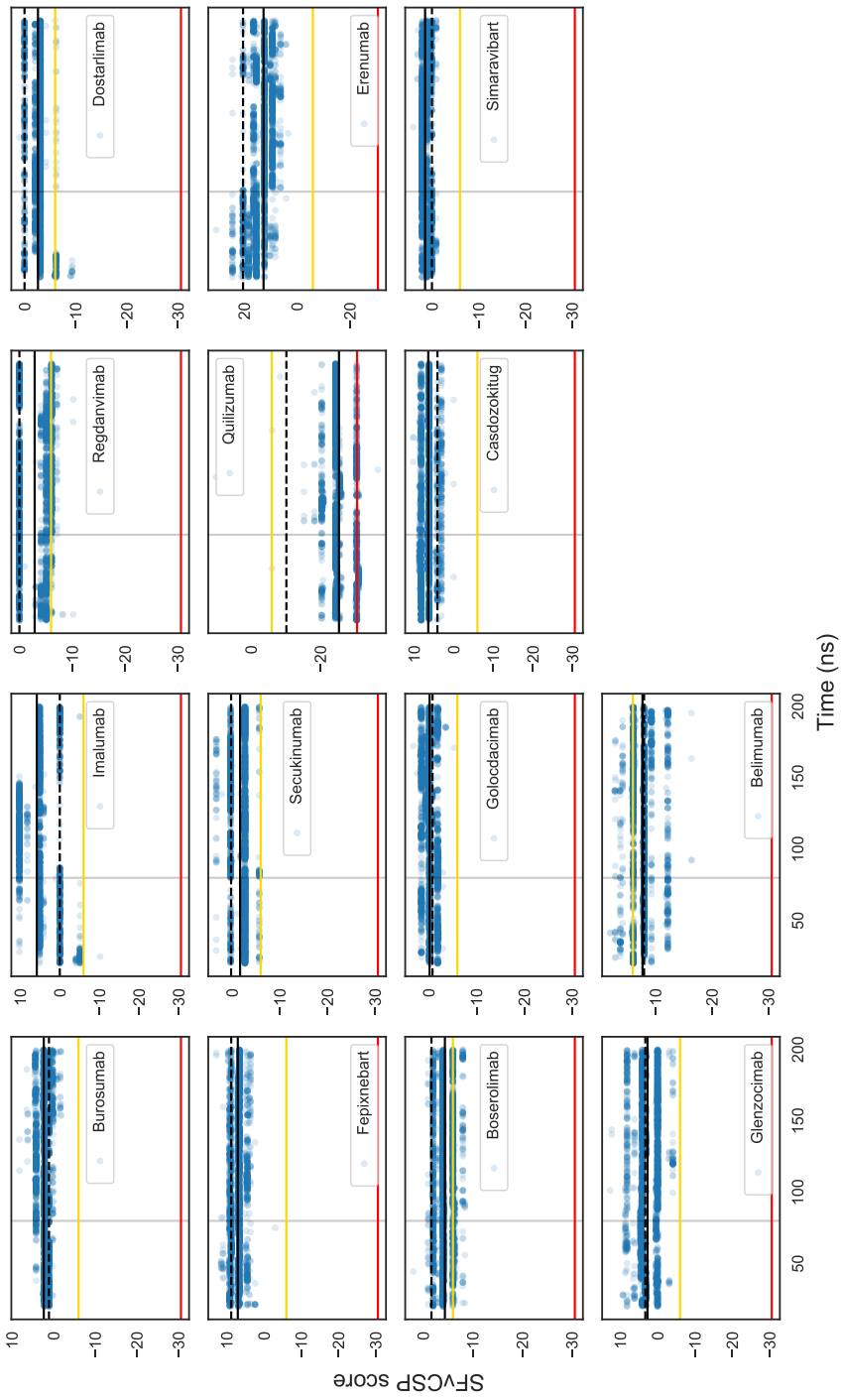


Figure A.10: Changes in asymmetry in the net heavy-and light-chain surface charges (SFvCSP) scores of antibodies during simulation. The region above the amber lines corresponds to SFvCSP scores that would be assigned a green flag by TAP and the regions below the red lines correspond to red flag values. Regions between the red and amber lines correspond to amber flag scores. Dashed lines correspond to the average SFvCSP score from three ABB2 models and the solid black line shows the average SFvCSP score over the final 120 ns of the trajectory (marked by the grey line).

B

Supplementary Material for Chapter 4

B.1 Example of dynamic time warping algorithm

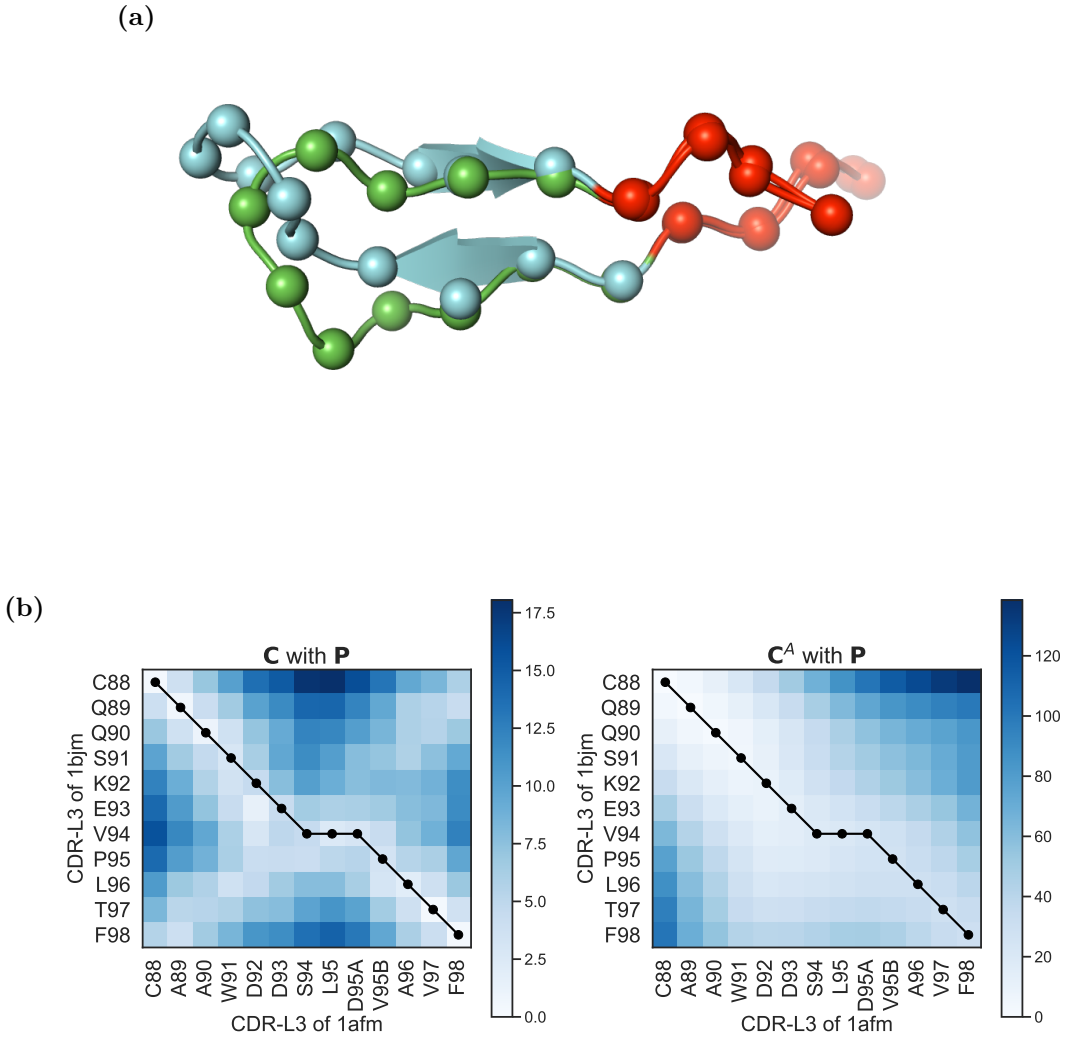


Figure B.1: Example of dynamic time warping distance calculation. **(a)** shows two CDR-L3 loops aligned on anchors (anchor Cα atoms shown in red, loop Cα atoms shown in cyan and green), to align loops for further calculations. **(b)** (left) shows the cost matrix, $\mathbf{C}$, where $\mathbf{C}_{i,j}$ is the Euclidean distance of the i^{th}Cα atom of CDR-L3 from 1AFM to the the j^{th}Cα atom of CDR-L3 from 1BJM. **(b)** (right) shows the accumulated cost matrix $\mathbf{C}^A$ with the optimal warping path tracing through each matrix. The total cost of traversing this matrix is 29.45 Å, making the distance between these two loops $29.45/13 = 2.27$ Å, since 13 steps were required to traverse through the cost matrix. ***NB:*** For ease of demonstration, DTW has only been performed using Cα atoms in this case. In our analysis of CDR-H3 loops, we used all backbone heavy atoms.

B.2 Supplementary figures

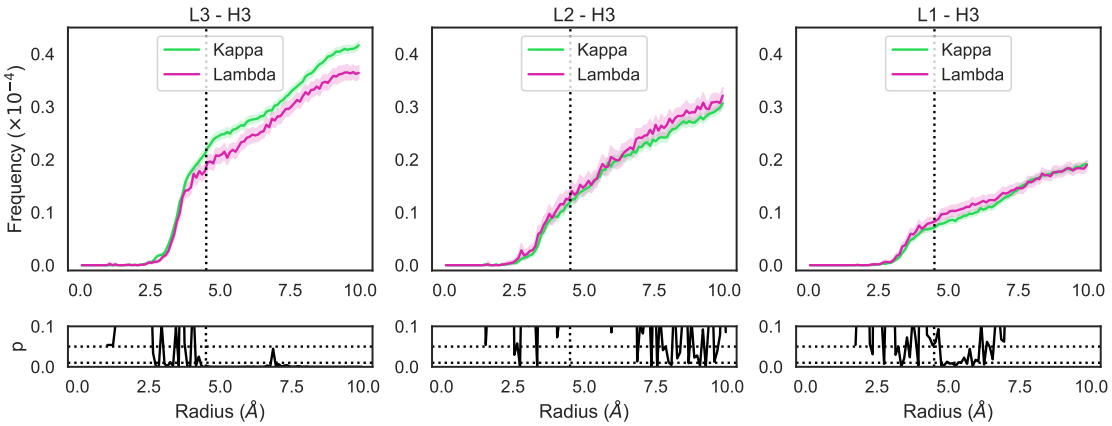


Figure B.2: Radial distribution functions of CDR loops in the crystal data set separated by light chain type. We generated histograms for all pairwise CDR loop heavy atom distances using a bin width of 0.1 Å. Each bin was then normalised by dividing by the product of the volume of a spherical shell centered at 0 and located on the edges of the bin and the number of possible pairs of heavy atoms. Shaded areas represent 95% confidence intervals. p-values are calculated for each bin using a t-test.

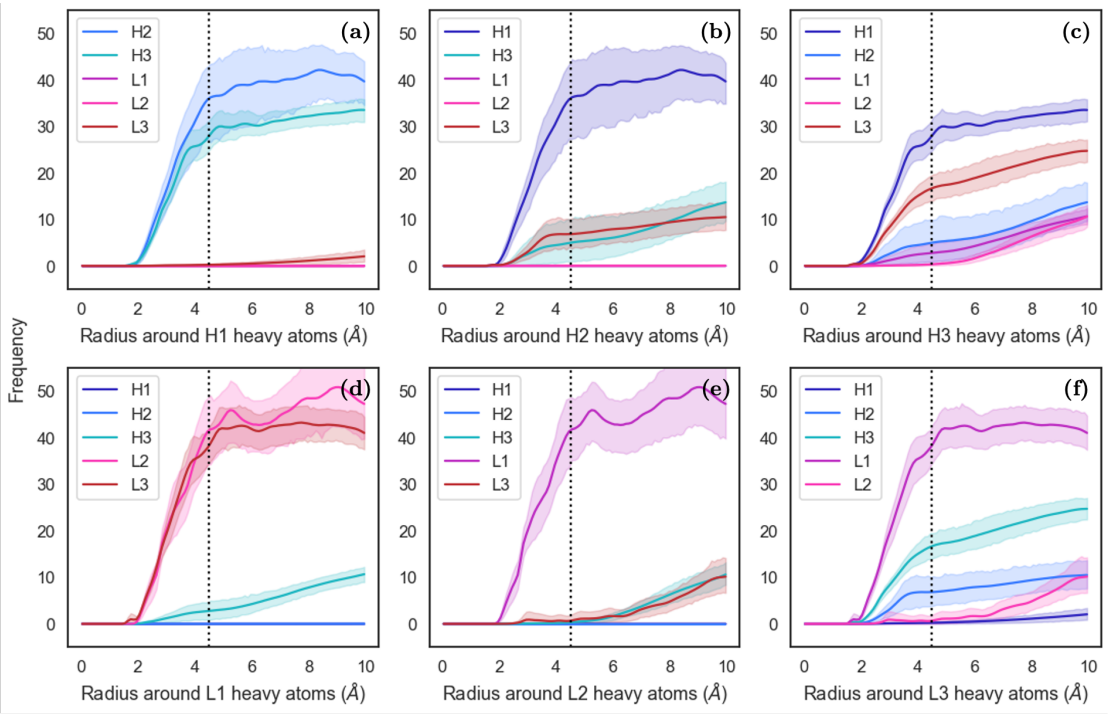


Figure B.3: Radial distribution functions of CDR loops in the MD simulations. We calculated RDFs using the mdtraj implementation with a bin width of 0.1 Å. Shaded areas represent 95% confidence intervals.

C

Supplementary Material for Chapter 5

C.1 Lists of antibodies used to train, test, and validate models

Each F_V in each set is given by its PDB accession code, followed by the heavy and light chains (xxxHL)

Train Set: 12e8HL, 12e8PM, 15c8HL, 1a2yBA, 1a3lHL, 1a3rHL, 1a4jBA, 1a4jHL, 1a4kBA, 1a4kHL, 1a6tBA, 1a6tDC, 1a6uHL, 1a6vIM, 1a7nHL, 1a7oHL, 1a7pHL, 1a7rHL, 1acyHL, 1ad0BA, 1ad0DC, 1ad9BA, 1ad9HL, 1adqHL, 1ae6HL, 1afvHL, 1afvKM, 1ahwBA, 1ahwED, 1ai1HL, 1aifBA, 1aifHL, 1aj7HL, 1aqkHL, 1ar1CD, 1axsBA, 1axsHL, 1axtHL, 1ay1HL, 1b2wHL, 1b4jHL, 1bafHL, 1bbdHL, 1bbjBA, 1bbjHL, 1beyHL, 1bfoBA, 1bfoDC, 1bfoFE, 1bfoHG, 1bfvHL, 1bj1HL, 1bj1KJ, 1blnBA, 1blnDC, 1bogBA, 1bqlHL, 1bvkBA, 1bvkED, 1bvlAB, 1bvlCD, 1bz7BA, 1c08BA, 1c12BA, 1c1eHL, 1c5bHL, 1c5cHL, 1c5dBA, 1c5dHL, 1cbvHL, 1ce1HL, 1cf8HL, 1cfnBA, 1cfqBA, 1cfsBA, 1cftBA, 1cfvHL, 1cgsHL, 1cicBA, 1cicDC, 1ck0HL, 1cl7HL, 1cloHL, 1clyHL, 1clzHL, 1cr9HL, 1ct8BA, 1ct8DC, 1cu4HL, 1cz8HL, 1cz8YX, 1d5bBA, 1d5bHL, 1d5iHL, 1d6vHL, 1dbaHL, 1dbbHL, 1dbjHL, 1dbkHL, 1dbmHL, 1deeBA, 1deeDC, 1deeFE, 1dfbHL, 1dl7HL, 1dn0BA, 1dn0DC, 1dqdHL, 1dqjBA, 1dqlHL, 1dqqBA, 1dqqDC, 1dsfHL, 1dvfBA, 1dvfDC, 1e4wHL, 1e4xHL, 1e4xIM, 1e6jHL, 1e6oHL, 1eapBA, 1egjHL, 1ehlHL, 1emtHL, 1eo8HL,

1etzBA, 1etzHL, 1ezvXY, 1f11BA, 1f11DC, 1f3dHL, 1f3dKJ, 1f4wHL, 1f4xHL, 1f4yHL, 1f58HL, 1f8tHL, 1f90HL, 1faiHL, 1fbiHL, 1fbiQP, 1fdlHL, 1fe8HL, 1fe8IM, 1fe8JN, 1fgnHL, 1figHL, 1fj1BA, 1fj1DC, 1fl3AB, 1fl3HL, 1fl5BA, 1fl5HL, 1fl6BA, 1fl6HL, 1forHL, 1fptHL, 1frgHL, 1fskCB, 1fskFE, 1fskIH, 1fskLK, 1fvcBA, 1fvcDC, 1fvdBA, 1fvdDC, 1fveBA, 1fveDC, 1g7hBA, 1g7iBA, 1g7mBA, 1g9mHL, 1g9nHL, 1gafHL, 1gc1HL, 1ggbHL, 1ggcHL, 1ggiHL, 1ggiJM, 1ghfHL, 1gpoHL, 1gpoIM, 1h0dBA, 1h3pHL, 1hezBA, 1hezDC, 1hh6BA, 1hh9BA, 1hi6BA, 1hilBA, 1hilDC, 1himLH, 1himMJ, 1hinHL, 1hklHL, 1hq4BA, 1hq4DC, 1hysDC, 1hzhHL, 1hzhKM, 1i3gHL, 1i7zBA, 1i7zDC, 1i8iBA, 1i8kBA, 1i8mBA, 1i8mHL, 1i9iHL, 1i9jHL, 1i9rHL, 1i9rKM, 1i9rXY, 1iaiHL, 1iaiIM, 1ibgHL, 1ic4HL, 1ic5HL, 1ic7HL, 1ifhHL, 1igcHL, 1igfHL, 1igfJM, 1igiHL, 1igjBA, 1igjDC, 1igmHL, 1igtBA, 1igtDC, 1ikfHL, 1il1AB, 1indHL, 1ineHL, 1iqdBA, 1iqwHL, 1it9HL, 1j1oHL, 1j1pHL, 1j1xHL, 1j5oHL, 1jguHL, 1jgvHL, 1jhlHL, 1jnlHL, 1jnnHL, 1jrhHL, 1jv5BA, 1k4cAB, 1k4dAB, 1k6qHL, 1kb5HL, 1kb9JK, 1kc5HL, 1kcrHL, 1kcsHL, 1kcuHL, 1kcvHL, 1kegHL, 1kelHL, 1kemHL, 1kenHL, 1kenTU, 1kfaHL, 1kfaIM, 1kipBA, 1kiqBA, 1kirBA, 1kn2HL, 1kn4HL, 1knoBA, 1knoDC, 1knoFE, 1kyoJK, 1kyoUV, 1l7iHL, 1l7tHL, 1lk3HL, 1lk3IM, 1lo0HL, 1lo0YX, 1lo2HL, 1lo2YX, 1lo3HL, 1lo3YX, 1lo4HL, 1m71BA, 1m7dBA, 1m7iBA, 1mamHL, 1mcoHL, 1mcpHL, 1mf2HL, 1mf2NM, 1mh5BA, 1mh5HL, 1mhpHL, 1mhpXY, 1mimHL, 1mj8HL, 1mjjBA, 1mjjHL, 1mjuHL, 1mnuHL, 1mpaHL, 1n0xHL, 1n0xKM, 1n4xHL, 1n4xIM, 1n5yHL, 1n6qHL, 1n7mLH, 1n8zBA, 1nakHL, 1nakIM, 1nbvHL, 1nbyBA, 1nbzBA, 1nc2BA, 1nc2DC, 1nc4BA, 1nc4DC, 1ncaHL, 1ncbHL, 1nccHL, 1ncdHL, 1ncwHL, 1nd0BA, 1nd0DC, 1nd0FE, 1nd0HG, 1ndmBA, 1nfdFE, 1nfdHG, 1ngpHL, 1ngqHL, 1ngwBA, 1ngwHL, 1ngxBA, 1ngxHL, 1ngyBA, 1ngzBA, 1nj9HL, 1nl0HL, 1nldHL, 1nmaHL, 1nmbHL, 1nmcBC, 1nmcHL, 1nsnHL, 1ob1BA, 1ob1ED, 1om3HL, 1om3KM, 1op3HL, 1op3MK, 1op5HL, 1op5MK, 1opgHL, 1orqBA, 1orsBA, 1ospHL, 1otsCD, 1otsEF, 1ottCD, 1ottEF, 1otuCD, 1otuEF, 1p2cBA, 1p2cED, 1p7kBA, 1p84JK, 1pg7XW, 1pg7ZY, 1pkqBA, 1pkqGF, 1plgHL, 1pskHL, 1pz5BA, 1q0xHL, 1q0yHL, 1q1jHL, 1q1jIM, 1q72HL, 1q9kBA, 1q9lBA, 1q9lDC, 1q9oBA, 1q9oDC, 1q9wBA, 1q9wDC, 1qblHL, 1qbmHL, 1qfuHL, 1qgcA4, 1qkzHL, 1qleHL, 1qlrBA, 1qlrDC,

1qygHL, 1r0aHL, 1r24BA, 1r24DC, 1r3iHL, 1r3jBA, 1r3kBA, 1r3lBA, 1rfdHL, 1rhhBA, 1rihHL, 1riuHL, 1rivHL, 1rjlBA, 1rmfHL, 1rukHL, 1rumHL, 1rupHL, 1ruqHL, 1rurHL, 1rvfHL, 1rz7HL, 1rz8BA, 1rz8DC, 1rzjHL, 1rzkHL, 1s3kHL, 1s5hBA, 1s5iHL, 1s78DC, 1s78FE, 1sbsHL, 1sm3HL, 1sy6HL, 1t03HL, 1t04BA, 1t04DC, 1t2qHL, 1t3fBA, 1t4kBA, 1t4kDC, 1t66DC, 1t66HL, 1tetHL, 1tjgHL, 1tjhHL, 1tjiHL, 1tpxBC, 1tqbBC, 1tqcBC, 1tzgHL, 1tzgIM, 1tzhBA, 1tzhHL, 1tziBA, 1u6aHL, 1u8hBA, 1u8iBA, 1u8jBA, 1u8kBA, 1u8lBA, 1u8mBA, 1u8nBA, 1u8oBA, 1u8pBA, 1u8qBA, 1u91BA, 1u92BA, 1u93BA, 1u95BA, 1ua6HL, 1uacHL, 1ub5AB, 1ub5HL, 1ub6AB, 1ub6HL, 1uj3BA, 1um4HL, 1um5HL, 1um6HL, 1uweHL, 1uweVU, 1uweYX, 1uwgHL, 1uwgYX, 1uwxHL, 1uwxMK, 1uywHL, 1uywMN, 1uz6FE, 1uz6HL, 1uz6PM, 1uz6WV, 1uz8BA, 1uz8HL, 1v7mHL, 1v7mIM, 1v7nHL, 1v7nIM, 1v7nJN, 1v7nKO, 1vfaBA, 1vfbBA, 1vpoHL, 1w72HL, 1w72IM, 1wcbBA, 1wcbHL, 1wejHL, 1xf2HL, 1xgpBA, 1xgqBA, 1xgrBA, 1xgtBA, 1xguBA, 1xgyHL, 1xgyIM, 1xiwDC, 1xiwHG, 1y0lBA, 1y0lDC, 1y0lFE, 1y0lHL, 1y18BA, 1y18DC, 1y18FE, 1y18HL, 1yecHL, 1yedBA, 1yedHL, 1yeeHL, 1yefHL, 1yegHL, 1yehHL, 1yeiHL, 1yejHL, 1yekHL, 1yjdHL, 1ynkHL, 1ynlHL, 1yntBA, 1yntDC, 1yqvHL, 1yy8BA, 1yy8DC, 1yy9DC, 1yylHL, 1yylRQ, 1yymHL, 1yymRQ, 1z3gHL, 1z3gIM, 1za3BA, 1za3HL, 1za6BA, 1za6DC, 1za6FE, 1za6HG, 1zanHL, 1zeaHL, 1zlsHL, 1zluHL, 1zluMK, 1zlvMK, 1zlwHL, 1zlwMK, 1ztxHL, 1zwiAB, 25c8HL, 2a0lDC, 2a0lFE, 2a1wHL, 2a1wIM, 2a6dBA, 2a6dHL, 2a6iBA, 2a6jBA, 2a6jHL, 2a6kBA, 2a6kHL, 2aabHL, 2adgBA, 2adiBA, 2adjBA, 2aepHL, 2agjHL, 2ai0IM, 2ai0JN, 2aj3BA, 2aj3DC, 2aj3FE, 2ajsHL, 2ajuHL, 2ajvHL, 2ajxHL, 2ajyHL, 2ajzBA, 2ajzHL, 2ak1HL, 2arjBA, 2arjHL, 2atkAB, 2b0sHL, 2b1aHL, 2b1hHL, 2b2xHL, 2b2xIM, 2b4cHL, 2bdnHL, 2bfvHL, 2bmkBA, 2bmkHL, 2bobAB, 2bocAB, 2brrHL, 2brrYX, 2c1oBA, 2c1oHL, 2c1pBA, 2c1pHL, 2cgrHL, 2cjuHL, 2ck0HL, 2cmrHL, 2d03HL, 2dblHL, 2dd8HL, 2ddqHL, 2dqcHL, 2dqdHL, 2dqeHL, 2dqfBA, 2dqfED, 2dqgHL, 2dqhHL, 2dqiHL, 2dqjHL, 2dqtHL, 2dquHL, 2dtgAB, 2dtgCD, 2dwdAB, 2dweAB, 2eh7HL, 2eh8HL, 2eizBA, 2eksBA, 2exwCD, 2exwEF, 2exyCD, 2exyEF, 2ez0CD, 2ez0EF, 2f19HL, 2f58HL, 2f5aHL, 2f5bHL, 2fatHL, 2fd6HL, 2fecIL, 2fecJO, 2fedCD, 2fedEF, 2feeIL, 2feeJO, 2fjfBA, 2fjfDC, 2fjfFE, 2fjfHL, 2fjfIG, 2fjfKJ,

2fjfNM, 2fjfPO, 2fjfRQ, 2fjfTS, 2fjfVU, 2fjfXW, 2fjgBA, 2fjgHL, 2fjhBA, 2fjhHL, 2fl5BA, 2fl5DC, 2fl5FE, 2fl5HL, 2fr4BA, 2fr4HL, 2fx7HL, 2fx8HL, 2fx8IM, 2fx8JN, 2fx8KO, 2fx9HL, 2fx9IM, 2g2rBA, 2g2rHL, 2g5bBA, 2g5bDC, 2g5bFE, 2g5bHG, 2g60HL, 2g75AB, 2g75CD, 2gcyBA, 2gfbBA, 2gfbDC, 2gfbFE, 2gfbHG, 2gfbJI, 2gfbLK, 2gfbNM, 2gfbPO, 2gjzBA, 2gjzHL, 2gk0HL, 2gsgBA, 2gsgDC, 2gsiHG, 2h1pHL, 2h2pCD, 2h2pEF, 2h2sCD, 2h2sEF, 2h8pAB, 2h9gBA, 2h9gHL, 2hfeAB, 2hfgHL, 2hg5AB, 2hjfAB, 2hkfHL, 2hlfCD, 2hlfEF, 2hmiDC, 2hrpHL, 2hrpNM, 2ht2CD, 2ht2EF, 2ht3CD, 2ht3EF, 2ht4CD, 2ht4EF, 2htkCD, 2htkEF, 2htlCD, 2htlEF, 2hvjAB, 2hvkAB, 2i5yHL, 2i5yRQ, 2i60HL, 2i60RQ, 2i9lBA, 2i9lDC, 2i9lFE, 2i9lHG, 2ibzXY, 2iffHL, 2igfHL, 2ih1AB, 2ih3AB, 2iptHL, 2ipuGK, 2ipuHL, 2iq9HL, 2itcAB, 2itdAB, 2j4wHL, 2j5lCB, 2j6eHL, 2j6eIM, 2jb5HL, 2jelHL, 2jixDG, 2jixFA, 2jixHL, 2jk5AB, 2mcpHL, 2mpaHL, 2nljBA, 2nr6DC, 2nr6FE, 2ntfBA, 2ntfHL, 2nxyDC, 2nxzDC, 2ny0DC, 2ny1DC, 2ny2DC, 2ny3DC, 2ny4DC, 2ny5HL, 2ny6DC, 2ny7HL, 2nyyDC, 2nz9DC, 2nz9FE, 2o5xHL, 2o5yHL, 2o5zHL, 2ojzHL, 2ojzIM, 2ok0HL, 2op4HL, 2oqjBA, 2oqjED, 2oqjHG, 2oqjKJ, 2or9HL, 2or9IM, 2oslAB, 2oslHL, 2otuBA, 2otuDC, 2otuFE, 2otuHG, 2otwBA, 2otwDC, 2oz4HL, 2p7tAB, 2p8lBA, 2p8pBA, 2pcpBA, 2pcpDC, 2pw1BA, 2pw2BA, 2q76BA, 2q76DC, 2q8aHL, 2q8bHL, 2qhrHL, 2qqkHL, 2qqlHL, 2qqnHL, 2qr0BA, 2qr0FE, 2qr0HG, 2qr0LK, 2qr0NM, 2qr0RQ, 2qr0TS, 2qr0XW, 2qscHL, 2r0kHL, 2r0lHL, 2r0wHL, 2r0zHL, 2r1wBA, 2r1xBA, 2r1yBA, 2r23BA, 2r29HL, 2r2bBA, 2r2eBA, 2r2hBA, 2r4rHL, 2r4sHL, 2r56HL, 2r56IM, 2r9hCD, 2r9hEF, 2rcsHL, 2uudHL, 2uudJK, 2uylBA, 2uylNM, 2uylWV, 2uylYX, 2uziHL, 2v7hBA, 2v7hHL, 2v7nBA, 2v7nDC, 2v7nFE, 2v7nHG, 2vc2HL, 2vdkHL, 2vdlHL, 2vdmHL, 2vdnHL, 2vdoHL, 2vdpHL, 2vdqHL, 2vdrHL, 2vh5HL, 2vl5AB, 2vl5CD, 2vq1FE, 2vweLJ, 2vxqHL, 2vxsHL, 2vxsIM, 2vxsJN, 2vxsKO, 2vxtHL, 2vxuHL, 2vxuIM, 2w0fAB, 2w9dHL, 2w9eHL, 2wubHL, 2wubRQ, 2wucHL, 2x7lAB, 2x7lCD, 2x7lEF, 2x7lGI, 2x7lHL, 2x7lJK, 2xa8HL, 2xknBA, 2xknDC, 2xqyGL, 2xqyJK, 2xraHL, 2xtjDB, 2xzaHL, 2xzcHL, 2xzqHL, 2y07HL, 2y36HL, 2y5tAB, 2y6sDC, 2y6sHL, 2ybrAB, 2ybrDE, 2ybrGH, 2yc1AB, 2yc1DE, 2yk1HL, 2ymxHL, 2yssBA, 2z4qBA, 2z91AB, 2z91CD, 2z92AB, 2zchHL, 2zckHL, 2zclHL, 2zjsHL, 2zkhHL, 2zpkHL, 2zpkIM, 32c2BA, 35c8HL,

3a67HL, 3a6bHL, 3a6cHL, 3aazAB, 3aazHL, 3ab0BC, 3ab0EF, 3b2uCD, 3b2uFG, 3b2uHL, 3b2uJK, 3b2uNO, 3b2uQR, 3b2uTU, 3b2uWX, 3b2vHL, 3b9kDC, 3b9kHL, 3baeHL, 3bdyHL, 3be1HL, 3bgfBC, 3bgfHL, 3bkcHL, 3bkjHL, 3bkmHL, 3bkyHL, 3bn9DC, 3bn9FE, 3bpcBA, 3bszHL, 3bszNM, 3bt2HL, 3bz4BA, 3bz4DC, 3bz4FE, 3bz4HG, 3c09CB, 3c09HL, 3c2aHL, 3c2aIM, 3c5sDC, 3c6sBA, 3c6sDC, 3c6sFE, 3c6sHG, 3cfbBA, 3cfbHL, 3cfcHL, 3cfdBA, 3cfdHL, 3cfeBA, 3cfeHL, 3cfjBA, 3cfjDC, 3cfjFE, 3cfjHL, 3cfkBA, 3cfkDC, 3cfkFE, 3cfkHL, 3cfkIG, 3cfkKJ, 3cfkNM, 3cfkPO, 3ck0HL, 3cleHL, 3clfHL, 3cmoHL, 3cmoYX, 3cvhHL, 3cvhQR, 3cviHL, 3cx5JK, 3cx5UV, 3cxdHL, 3cxhJK, 3cxhUV, 3d0vBA, 3d69BA, 3d69HL, 3d85BA, 3d9aHL, 3detCD, 3detEF, 3dggBA, 3dggDC, 3difBA, 3difDC, 3dsfHL, 3durBA, 3durDC, 3dusBA, 3dusDC, 3duuBA, 3duuDC, 3dv6BA, 3efdHL, 3effBA, 3effDC, 3ejyCD, 3ejyEF, 3ejzCD, 3ejzEF, 3eo0BA, 3eo0DC, 3eo1BA, 3eo1ED, 3eo1HG, 3eo1KJ, 3eotHL, 3eyfBA, 3eyfDC, 3eyoDC, 3eysHL, 3eyuHL, 3eyvBA, 3eyvHL, 3f58HL, 3f5wAB, 3f7vAB, 3f7yAB, 3fb5AB, 3fb6AB, 3fb7AB, 3fb8AB, 3fctBA, 3fctDC, 3ffdAB, 3fmgHL, 3fn0HL, 3fo0HL, 3fo1BA, 3fo1HL, 3fo2BA, 3fo2HL, 3fo9BA, 3fo9HL, 3fzuCD, 3fzuHL, 3g04BA, 3g5vBA, 3g5xBA, 3g5xDC, 3g5zBA, 3g6aBA, 3g6aHL, 3g6dHL, 3g6jFE, 3g6jHG, 3gb7AB, 3ggwBA, 3ggwDC, 3ghbHL, 3ghbIM, 3gheHL, 3gi8HL, 3gi9HL, 3gjeBA, 3gjeHL, 3gjfHL, 3gjfMK, 3grwHL, 3h0tBA, 3h42HL, 3haeHL, 3haeIG, 3haeON, 3haeTS, 3hb3CD, 3hc0AB, 3hc0HL, 3hfmHL, 3hi1BA, 3hi1HL, 3hi5HL, 3hi6HL, 3hi6XY, 3hmwHL, 3hmxHL, 3hplAB, 3hr5BA, 3hr5HL, 3hr5IP, 3hr5JQ, 3hzkBA, 3hzmBA, 3hzvBA, 3hzyBA, 3i02BA, 3i02DC, 3i2cHL, 3i50HL, 3i75BA, 3idgBA, 3idjBA, 3idmBA, 3idnBA, 3idxHL, 3idyBC, 3idyHL, 3ietBA, 3ietDC, 3if1BA, 3if1DC, 3ifnHL, 3ifoAB, 3ifoHL, 3ifpAB, 3ifpCD, 3ifpEF, 3ifpHL, 3igaAB, 3ijhBA, 3ijhDC, 3ijsBA, 3ijsDC, 3ijyBA, 3ijyDC, 3ikcBA, 3ikcDC, 3inuHL, 3inuMN, 3iu3AB, 3iu3CD, 3iu3HL, 3ixtAB, 3ixtHL, 3iy0HL, 3iy3BA, 3iy5BA, 3iywHL, 3iywKM, 3j2xBA, 3j2xDC, 3j2xFE, 3j2xHG, 3j2yBA, 3j2yDC, 3j2yFE, 3j2yHG, 3j2zBA, 3j2zDC, 3j2zFE, 3j2zHG, 3j30BA, 3j30DC, 3j30FE, 3j30HG, 3j42GH, 3j42IJ, 3j42KL, 3j5mDC, 3j5mHG, 3j5mLK, 3j70AB, 3j70MN, 3j70RS, 3j7eBA, 3j7eDC, 3j7eFE, 3j7eHL, 3j8vFJ, 3j8vGK, 3j8vHL, 3j8vIM, 3jabHL, 3jabQR, 3jauHL, 3jwdHL, 3jwdPO, 3jwoHL, 3k2uHL, 3kj4HL,

3kj6HL, 3klhDC, 3kr3HL, 3ks0HL, 3ks0KJ, 3kymBA, 3kymDC, 3kymFE, 3kymHG, 3kymJI, 3kymLK, 3kymNM, 3kymPO, 3l5wBA, 3l5wHL, 3l5xHL, 3l5yHL, 3l7eBA, 3l7eHL, 3l7fBA, 3l7fED, 3l7fHL, 3l95BA, 3l95HL, 3levHL, 3lexAB, 3lexHL, 3leyHL, 3lh2JN, 3lh2KO, 3lohAB, 3ls4HL, 3ls5HL, 3lzfHL, 3ma9HL, 3macHL, 3mbxHL, 3mckBA, 3mckHL, 3mclHL, 3mj8BA, 3mj8HL, 3mj9HL, 3mlrHL, 3mlsHL, 3mlsIM, 3mlsJN, 3mlsKO, 3mluHL, 3mlvHL, 3mlvNM, 3mlwHL, 3mlwIM, 3mlxHL, 3mlxIM, 3mlyHL, 3mlyIM, 3mlzHL, 3mmeAB, 3mmeHL, 3mnvBA, 3mnvDC, 3mo1BA, 3moaHL, 3mobHL, 3modHL, 3n85HL, 3n9gHL, 3na9HL, 3naaHL, 3nabHL, 3nacHL, 3ncjHL, 3ncyPS, 3ncyQW, 3nfpAB, 3nfpHL, 3nfsHL, 3ngbBC, 3ngbEF, 3ngbHL, 3ngbJK, 3nh7HL, 3nh7IM, 3nh7JN, 3nh7KO, 3nidEF, 3nidHL, 3nifEF, 3nifHL, 3nigEF, 3nigHL, 3ntcHL, 3nz8AB, 3nz8HL, 3nzhHL, 3o0rHL, 3o11BA, 3o11HL, 3o2dHL, 3o2vHL, 3o2wHL, 3o41AB, 3o41HL, 3o45AB, 3o45HL, 3o6kHL, 3o6lHL, 3o6mHL, 3oauHL, 3oayHL, 3oayMK, 3oazHL, 3oazMK, 3ob0HL, 3ob0MK, 3ogcAB, 3ojdBA, 3okdBA, 3okeBA, 3okkBA, 3oklBA, 3okmBA, 3oknBA, 3okoBA, 3opzIM, 3opzJN, 3or6AB, 3or7AB, 3p30HL, 3pgfHL, 3phoBA, 3piqCD, 3piqGI, 3piqHL, 3piqJK, 3pjsBA, 3pjsDC, 3pnwBA, 3pnwED, 3pnwHG, 3pnwKJ, 3pnwNM, 3pnwQP, 3pnwTS, 3pnwWV, 3pp3HL, 3pp3IK, 3pp4HL, 3q1sHL, 3q3gBA, 3q3gDC, 3q3gHF, 3q3gKJ, 3q6gHL, 3qa3BA, 3qa3DC, 3qa3HF, 3qa3KJ, 3qctHL, 3qcuHL, 3qcuIM, 3qcvHL, 3qcvIM, 3qehAB, 3qehCD, 3qehEF, 3qehGH, 3qg6BA, 3qg6HL, 3qg7HL, 3qhfHL, 3qosHL, 3qotHL, 3qpqDC, 3qpqFE, 3qpqHL, 3qpqJI, 3qpxHL, 3qq9DC, 3qq9HL, 3qrgHL, 3qumBA, 3qumDC, 3qumHL, 3qumKM, 3qwoAB, 3qwoHL, 3r06BA, 3r06DC, 3r06FE, 3r06HL, 3r08HL, 3r1gHL, 3ra7HL, 3ra7IM, 3rajHL, 3rhwFN, 3rhwGK, 3rhwHL, 3rhwIO, 3rhwJM, 3ri5FN, 3ri5GK, 3ri5HL, 3ri5IO, 3ri5JM, 3riaFN, 3riaGK, 3riaHL, 3riaIO, 3riaJM, 3rifFN, 3rifGK, 3rifHL, 3rifIO, 3rifJM, 3rkdDC, 3rkdHL, 3rpiAB, 3rpiHL, 3ru8HL, 3rvtDC, 3rvuDC, 3rvvDC, 3rvwDC, 3rvxDC, 3s34HL, 3s35HL, 3s36HL, 3s37HL, 3s62HL, 3s88HL, 3s96AB, 3s96CD, 3sdyHL, 3se9HL, 3sgeJI, 3skjHL, 3skjIM, 3sm5HL, 3sm5IM, 3sm5JN, 3so3CB, 3sqoHL, 3stlAB, 3stzAB, 3sy0BA, 3t2nHL, 3t2nIM, 3t3mEF, 3t3mHL, 3t3pEF, 3t3pHL, 3t4yBA, 3t65BA, 3t77BA, 3tclAB, 3tclHL, 3thmHL, 3tjeHL, 3tnmHL, 3tnnAB, 3tnnCD, 3tnnEF, 3tnnHL, 3tt1HL, 3tt1IM, 3tv3HL, 3twcHL,

3tygHL, 3u0tBA, 3u0tDC, 3u0wHL, 3u30CB, 3u30FE, 3u46AB, 3u46HL, 3u4bHL, 3u6rAB, 3u6rHL, 3u7wHL, 3u7yHL, 3u9pKM, 3u9uAB, 3u9uCD, 3uajCD, 3uajHL, 3ubxGI, 3ubxHL, 3uc0HL, 3uc0IM, 3ujiHL, 3ujjHL, 3ujtHL, 3ujtIM, 3ulsBA, 3ulsHL, 3uluDC, 3uluFE, 3uluHL, 3ulvDC, 3ulvFE, 3ulvHL, 3uo1HL, 3utzBA, 3uyrHL, 3v4pHL, 3v4pMN, 3v4uHL, 3v4vHL, 3v4vMN, 3v52HL, 3v6fAB, 3v6fCD, 3v6fEF, 3v6fHL, 3v6oCE, 3v6zAB, 3v6zCD, 3v7aEH, 3v7aFG, 3ve0AB, 3vfgHL, 3vg0HL, 3vg9CB, 3vgaCB, 3vi3FE, 3vi3HL, 3vi4FE, 3vi4HL, 3vrlEF, 3vrlHL, 3vw3HL, 3w11CD, 3w12CD, 3w13CD, 3w2dHL, 3w9dAB, 3w9dCD, 3w9eAB, 3wd5HL, 3we6AB, 3wfbHL, 3wfcHL, 3wfdHL, 3wfeHL, 3wfhAB, 3whe12, 3whe34, 3whe56, 3whe78, 3whe90, 3wheMN, 3wheOP, 3wheQR, 3wheST, 3wheUV, 3wheWX, 3wheYZ, 3whxAB, 3wifAB, 3wihHL, 3wihIM, 3wiiHL, 3wiiIM, 3wkmHL, 3wkmIM, 3wlwCD, 3wlwHL, 3wsqHL, 3wxvHL, 3wxwHL, 3x3fHL, 3x3gHL, 3zdxEF, 3zdxHL, 3zdyEF, 3zdyHL, 3zdzEF, 3zdzHL, 3ze0EF, 3ze0HL, 3ze1EF, 3ze1HL, 3ze2EF, 3ze2HL, 3zkmCD, 3zkmHL, 3zknCD, 3zknHL, 3zl4HL, 3ztjIJ, 3ztjKL, 3ztnHL, 43c9BA, 43c9DC, 43c9FE, 43c9HG, 43caBA, 43caDC, 43caFE, 43caHG, 4a6yBA, 4a6yHL, 4aehHL, 4aeiHL, 4aeiIM, 4aeiJN, 4ag4HL, 4al8HL, 4alaHL, 4am0AB, 4am0CD, 4am0EF, 4am0HL, 4amkHL, 4bh7BA, 4bh8BA, 4bklAB, 4bz1HL, 4bz2HL, 4c83AB, 4c83CD, 4cadBA, 4cadED, 4cadHG, 4cadKJ, 4ckdHL, 4ckdIM, 4ckdJN, 4ckdKO, 4cmhBC, 4cniAB, 4cniHL, 4d9lHL, 4d9lIM, 4d9lJN, 4d9lKO, 4d9qED, 4d9qHL, 4d9rED, 4d9rHL, 4dagHL, 4dcqBA, 4dgiHL, 4dgvHL, 4dgyHL, 4dkeHL, 4dkeIM, 4dkfHL, 4dkfIM, 4dn3HL, 4dn4HL, 4dtgHL, 4dvbAB, 4dvbHL, 4dvrHL, 4dw2HL, 4ebqHL, 4edwHL, 4edxBA, 4edxHL, 4eneCD, 4eneEF, 4eowHL, 4ersHL, 4etqAB, 4etqHL, 4evnAB, 4evnCD, 4evnEF, 4evnGH, 4evnIJ, 4evnKL, 4evnMN, 4evnOP, 4f2mAB, 4f2mCD, 4f33BA, 4f33DC, 4f33FE, 4f33HG, 4f37FK, 4f37HL, 4f3fBA, 4f58HL, 4f58IM, 4f58JN, 4f58KO, 4fabHL, 4ffvDC, 4ffvHL, 4ffwDC, 4ffwHL, 4ffyHL, 4ffzHL, 4ffzZY, 4fg6CD, 4fg6EF, 4fnlHL, 4fnlIM, 4fp8HL, 4fp8IM, 4fp8JN, 4fp8KO, 4fqhAB, 4fqhHL, 4fqiHL, 4fqjHL, 4fqlHL, 4fqqBA, 4fqqHL, 4fqrab, 4fqrcd, 4fqref, 4fqrgh, 4fqrij, 4fqrkl, 4fqrmn, 4fqrop, 4fqrqr, 4fqrst, 4fqruv, 4fqrwx, 4fqvHL, 4fqvIM, 4fqvJN, 4fzeHL, 4g3yHL, 4g5zHL, 4g6aCD, 4g6aHL, 4g6fBD, 4g6fHL, 4g6jHL, 4g6kHL, 4g6mHL, 4g7vHL, 4g7yHL, 4g80AB, 4g80CD,

4g80EF, 4g80GH, 4gagHL, 4gayHL, 4glrHI, 4glrJK, 4gmsHL, 4gmsIM, 4gmsJN, 4gmtHL, 4gmtIM, 4gq9HL, 4gw1BA, 4gw1DC, 4gw5BA, 4gw5DC, 4gxuMN, 4gxuOP, 4gxuQR, 4gxuST, 4gxuUV, 4gxuWX, 4gxvHL, 4gxvIM, 4h88HL, 4h8wHL, 4hbcHL, 4hc1HL, 4hc1MN, 4hcrHL, 4hcrMN, 4hdiBA, 4hdiHL, 4hf5HL, 4hfuHL, 4hfwBA, 4hfwHL, 4hg4LM, 4hg4NO, 4hg4PQ, 4hg4RS, 4hg4TU, 4hg4VW, 4hg4XY, 4hh9BA, 4hh9DC, 4hhaBA, 4hieBA, 4hihBA, 4hiiBA, 4hijDC, 4hixHL, 4hj0CD, 4hj0PQ, 4hjgBA, 4hk3JN, 4hkzBA, 4hlzGH, 4hlzIJ, 4hlzKL, 4hpoHL, 4hs6BA, 4hs6HL, 4hs8HL, 4hwbHL, 4hweHL, 4hxaHL, 4hxbHL, 4hzlAB, 4hzlHL, 4i18AB, 4i18HL, 4i2xBA, 4i2xDC, 4i3rHL, 4i3sHL, 4i77HL, 4i9wED, 4i9wGF, 4idjHL, 4ij3CB, 4imkAD, 4imkBC, 4imlAB, 4imlHL, 4ioiBA, 4irzHL, 4j1uDC, 4j4pCD, 4j4pHL, 4j6rHL, 4j8rBA, 4j8rDC, 4jamHL, 4janAB, 4janHL, 4jb9HL, 4jdvAB, 4jdvHL, 4jfxBA, 4jfxHL, 4jfyBA, 4jfyHL, 4jfzHL, 4jg0HL, 4jg1HL, 4jhaHL, 4jhwHL, 4jkpHL, 4jlrAB, 4jlrHL, 4jm2AB, 4jm2DC, 4jm4HL, 4jn1HL, 4jn2HL, 4jo1HL, 4jo1IM, 4jo2HL, 4jo2IM, 4jo3HL, 4jo3IM, 4jo4HL, 4jo4IM, 4jpiHL, 4jpkHL, 4jpvHL, 4jpwHL, 4jqiHL, 4jr9HL, 4jreBC, 4jreHL, 4jzjAB, 4jzjHL, 4jznIP, 4jzoAB, 4jzoCF, 4jzoDE, 4k23HL, 4k24HL, 4k2uIM, 4k3dHL, 4k3eIM, 4k3jHL, 4k7pHL, 4k7pYX, 4k8rDC, 4k8rHL, 4k94HL, 4k9eHL, 4kaqHL, 4khtHL, 4khxHL, 4ki5CD, 4kjpCD, 4kjpEF, 4kjqCD, 4kjqEF, 4kjwCD, 4kjwEF, 4kk5CD, 4kk5EF, 4kk6CD, 4kk6EF, 4kk8CD, 4kk8EF, 4kk9CD, 4kk9EF, 4kkaCD, 4kkaEF, 4kkbCD, 4kkbEF, 4kkcCD, 4kkcEF, 4kklCD, 4kklEF, 4kmtHL, 4kq3HL, 4kq4HL, 4kroDC, 4krpDC, 4ktdHL, 4kteHL, 4kuzHL, 4kv5EF, 4kv5GK, 4kv5HL, 4kv5JI, 4kvcHL, 4kxzHL, 4kxzJI, 4kxzNM, 4kxzQP, 4ky1HL, 4l5fHL, 4larHL, 4lasHL, 4lbeAB, 4lciHL, 4lcuAB, 4leoAB, 4lexHL, 4lf3BA, 4lf3ED, 4liqHL, 4lkcBA, 4lkxAB, 4lluAB, 4llvAB, 4llvHL, 4llwAB, 4llwCD, 4llyAB, 4llyCD, 4lmqEI, 4lmqHL, 4louCD, 4louEF, 4lqfHL, 4lssHL, 4lstHL, 4lsuHL, 4lsvHL, 4lu5HL, 4lu5IM, 4lvhBC, 4lvhEF, 4lvhHI, 4lvnCB, 4lvoCB, 4m1dHL, 4m1dIM, 4m1gHL, 4m43HL, 4m48HL, 4m5yHL, 4m5yIM, 4m5zHL, 4m61BA, 4m61DC, 4m62HL, 4m62IM, 4m6mHL, 4m6nHL, 4m6oHL, 4m7jHL, 4m7kHL, 4m7lHL, 4m93BC, 4m93HL, 4ma1BC, 4ma1HL, 4ma3BA, 4ma3HL, 4ma7HL, 4ma8HL, 4mauHL, 4mhhGI, 4mhhHL, 4mhhJK, 4mqxCD, 4mqxEF, 4mswAB, 4mwfHL, 4mxvFE, 4mxvHL, 4mxvYX, 4mxwHL, 4mxwWV, 4n0yHL,

4n8cHL, 4n8cIM, 4n90DE, 4n90FG, 4n90HI, 4n9gMN, 4ncc21, 4nccHL, 4nghHL, 4nhcHL, 4nj9HL, 4njaHL, 4nkdAB, 4nkdCD, 4nkdEF, 4nkdGH, 4nkiHL, 4nkmAB, 4nkmCD, 4nkmEF, 4nkmGH, 4nkoBA, 4nkoDC, 4nkoFE, 4nm4HL, 4nm4IM, 4nm8HL, 4nm8IM, 4nm8JN, 4nnpHL, 4nnpXY, 4np4HL, 4np4IM, 4npyHL, 4nrxAB, 4nrxHL, 4nrzAB, 4nrzHL, 4nugHL, 4nwtHL, 4nx3BA, 4nylAB, 4nylCD, 4nylEF, 4nylHL, 4nzrHL, 4nztHL, 4o02HL, 4o4yHL, 4o51BA, 4o51DC, 4o51FE, 4o51HL, 4o58HL, 4o5iMN, 4o5iOP, 4o5iQR, 4o5iST, 4o5iUV, 4o5iWX, 4o5lHL, 4o9hHL, 4oawBA, 4oawDC, 4ocxHL, 4ocyHL, 4od2BA, 4odsHL, 4odtHL, 4oduAB, 4oduHL, 4odvHL, 4odwAB, 4odwHL, 4odxAL, 4odxHB, 4ogaCD, 4ogxHL, 4ogyHL, 4oiiHL, 4oiiIM, 4ojfHL, 4okvAB, 4okvCD, 4oluHL, 4olvHL, 4olwHL, 4olxHL, 4olyHL, 4olzHL, 4om0HL, 4om1HL, 4onfHL, 4oqtHL, 4ot1HL, 4otxHL, 4otxIM, 4ouuBA, 4ouuHL, 4oz4AB, 4oz4HL, 4p3cHL, 4p3dAB, 4p3dHL, 4p59HL, 4pb0HL, 4pb9HL, 4pd4JK, 4plkDC, 4plkHL, 4plkJK, 4pozDC, 4pp1DC, 4pp1FE, 4pp2BA, 4pp2DC, 4ps4HL, 4pttBA, 4ptuBA, 4py7AB, 4py7IJ, 4py8IJ, 4q0xHL, 4q2zAB, 4q2zHL, 4q6iBA, 4q6iED, 4q6iGF, 4q6iHL, 4qciBA, 4qciHL, 4qexHL, 4qexIM, 4qf1AB, 4qf1HL, 4qhkMN, 4qhkOP, 4qhlAB, 4qhlCD, 4qhmAB, 4qhmCD, 4qhmEF, 4qhmGH, 4qhnAB, 4qhuBA, 4qhuHL, 4qnpEF, 4qnpHL, 4qt5BA, 4qt5HL, 4qthAB, 4qthHL, 4qtiHL, 4qwwDC, 4qwwFE, 4qxgHL, 4qxtAB, 4qxuHL, 4qy8AB, 4qyoAB, 4r0lAB, 4r3sAB, 4r4bBA, 4r4bHL, 4r4hHL, 4r7dAB, 4r7dCD, 4r7dEF, 4r7dGH, 4r7dIJ, 4r7dKL, 4r7dMN, 4r7dOP, 4r7nAB, 4r7nCD, 4r7nEF, 4r7nGH, 4r7nIJ, 4r7nKL, 4r7nMN, 4r7nOP, 4r7nQR, 4r7nST, 4r8wHL, 4r90HL, 4r96BA, 4r96DC, 4r96FE, 4r96HL, 4r97CB, 4r9yHL, 4r9yNM, 4rauBA, 4rauED, 4rauHG, 4rauKJ, 4rauNM, 4rauQP, 4rauTS, 4rauWV, 4ravAB, 4ravCD, 4rbpHL, 4rbpMK, 4rdqGF, 4rdqIH, 4rdqKJ, 4rdqML, 4rdqON, 4rfeAB, 4rfeCD, 4rfeEF, 4rfeHL, 4rfnBC, 4rfnHL, 4rfoHL, 4rgmCB, 4rgmHL, 4rgnBC, 4rgnDE, 4rgnFG, 4rgnHL, 4rgoHL, 4rirAB, 4rirHL, 4risHL, 4rqsDC, 4rrpGA, 4rrpHB, 4rrpIC, 4rrpJD, 4rrpKE, 4rrpLF, 4rwyHL, 4rx4AD, 4rx4HL, 4s1dCD, 4s1dEF, 4s1dHL, 4s1dOP, 4s1qHL, 4s1rHL, 4s1sHL, 4s2sAB, 4s2sHL, 4tnvFN, 4tnvGK, 4tnvHL, 4tnvIO, 4tnvJM, 4tnvUh, 4tnvVZ, 4tnvWf, 4tnvXi, 4tnvYg, 4tnwFN, 4tnwGK, 4tnwHL, 4tnwIO, 4tnwJM, 4tnwUh, 4tnwVZ, 4tnwWf, 4tnwXi, 4tnwYg, 4toyHL, 4tprHL, 4tqeHL, 4tsbHL,

4tscHL, 4ttdHL, 4tuoCD, 4tvpDE, 4u0rBC, 4u1gBC, 4u1gEF, 4u6gAB, 4u6gHL, 4u6hAB, 4u6hCD, 4u6vHL, 4u6vKM, 4uaoCB, 4ub0HL, 4ubdCD, 4ubdGH, 4ubdKL, 4ubdOP, 4ubdST, 4ubdYX, 4uikHL, 4uilHL, 4uimAB, 4uimCD, 4uimEF, 4uimHL, 4uinHL, 4uokAL, 4ut7HL, 4ut9HL, 4ut9IM, 4ut9JN, 4ut9KO, 4utaHL, 4utaIM, 4utbHL, 4utbIM, 4uu9AB, 4uu9HL, 4uujAB, 4uv4HL, 4uv7HL, 4wcyHL, 4webHL, 4wfeED, 4wfeGF, 4wffED, 4wffGF, 4wfgED, 4wfgGF, 4wfhED, 4wfhGF, 4whtAB, 4whtCD, 4whtEF, 4whtGH, 4whtIJ, 4whtKL, 4whtMN, 4whtOP, 4whtQR, 4whtST, 4whtUV, 4whtXY, 4whyGH, 4whyIJ, 4whyKL, 4whyMN, 4wukHL, 4wuuED, 4wv1BA, 4wv1ED, 4wy7HL, 4x0kAB, 4x0kHL, 4x4xAB, 4x4yAB, 4x7sHL, 4x7tAB, 4x7tHL, 4xakDE, 4xakHL, 4xbgAB, 4xbgHL, 4xbpAB, 4xbpCD, 4xbpEF, 4xbpHL, 4xceAB, 4xceCD, 4xcnAB, 4xcnCD, 4xcnHL, 4xcyAB, 4xcyHL, 4xi5DC, 4xmkHL, 4xmkIM, 4xmkJN, 4xmlHL, 4xmpHL, 4xnmBA, 4xnmHL, 4xnqBA, 4xnqHL, 4xnuHL, 4xnxHL, 4xp1HL, 4xp4HL, 4xp5HL, 4xp6HL, 4xp9HL, 4xpaHL, 4xpbHL, 4xpfHL, 4xphHL, 4xptHL, 4xrcBA, 4xrcHL, 4xtrCD, 4xtrEF, 4xvjHL, 4xvsHL, 4xvtHL, 4xvuCD, 4xvuEF, 4xvuIJ, 4xvuKL, 4xwgHL, 4xwoCD, 4xwoEF, 4xwoIJ, 4xwoKL, 4xwoOP, 4xwoQR, 4xwoUV, 4xwoWX, 4xx1EG, 4xx1HL, 4xx1MO, 4xxdBA, 4xxdED, 4xzuAB, 4xzuEF, 4y5vAB, 4y5vDE, 4y5vGH, 4y5xAB, 4y5xDE, 4y5xGH, 4y5xJK, 4ybqDC, 4ybqFE, 4yc2HL, 4ydiHL, 4ydjAB, 4ydjHL, 4ydkHL, 4ydlBC, 4ydlHL, 4ye4HL, 4ygvAB, 4ygvFL, 4yhiAB, 4yhiHL, 4yhkHL, 4yhmHL, 4yhnAB, 4yhnFL, 4yhoHL, 4yhpAB, 4yhpCD, 4yhpEF, 4yhpHL, 4yk4CB, 4yk4ZY, 4ynyAB, 4ynyCD, 4yo0AB, 4yo0CD, 4ypgBA, 4ypgHL, 4yqxHL, 4yr6AB, 4yr6DE, 4yueHL, 4ywgIM, 4yx2HL, 4yxhHL, 4yxkHL, 4yxlHL, 4yzfEF, 4yzfGH, 4yzfIJ, 4yzfKL, 4z0xBA, 4z5rBA, 4z5rKJ, 4z5rML, 4z5rQP, 4z5rSR, 4z5rUT, 4z5rWV, 4z5rZY, 4z7nEF, 4z7nHL, 4z7oEF, 4z7oHL, 4z7qEF, 4z7qHL, 4z8fHL, 4z95HL, 4zd3HL, 4zffAB, 4zffHL, 4zfgHL, 4zfoAB, 4zfoHL, 4zptAB, 4zpvAB, 4zs6CD, 4zs6HL, 4zs7HL, 4zsoBA, 4zsoDC, 4ztoHL, 4ztoIM, 4ztpHL, 4zxbAB, 4zykAB, 4zypDE, 4zypFG, 4zypHI, 4zypJL, 4zypKM, 4zypNO, 5a16AB, 5a16CD, 5a16EF, 5a16GH, 5a3iCD, 5a3iGH, 5a7xDC, 5a7xHG, 5a7xLK, 5a8hDC, 5a8hJI, 5a8hPO, 5acoGJ, 5acoHL, 5acoIK, 5albHL, 5anmBA, 5anmDC, 5anyIM, 5anyJN, 5anyKO, 5anyLP, 5aumAB, 5aumHL, 5awnHL, 5ayuHL, 5az2HL, 5b4mCD, 5b4mHL, 5b71BA,

5b71DC, 5bjzCL, 5bjzDH, 5bk0BA, 5bk2CD, 5bk2HL, 5bk3AB, 5bk3CD, 5bk3EF, 5bk3HL, 5bk5DC, 5bo1HL, 5bo1IM, 5bv7CB, 5bv7HL, 5bvjBA, 5bvjDC, 5bvjFE, 5bvjHG, 5bvpHL, 5bzdBA, 5bzdDC, 5bzwBA, 5c0nCD, 5c0nHL, 5c0rHL, 5c7kEF, 5c7xHL, 5c8jAB, 5c8jCD, 5c8jEF, 5c8jGH, 5cckHL, 5cd3AB, 5cd3CD, 5cd3EF, 5cd3GH, 5cd5CD, 5cezDE, 5cgyCD, 5cgyHL, 5chnAB, 5chnHL, 5cilHL, 5cinHL, 5cipAB, 5cipHL, 5cjqHL, 5cjsFE, 5cjsHL, 5cjxAC, 5cjxDE, 5cjxHL, 5cmaBA, 5cp3HA, 5cp7BC, 5cp7DE, 5cp7HA, 5cusHL, 5cusIM, 5cusJN, 5cusKO, 5czvHL, 5czxCD, 5czxHL, 5d1qBA, 5d1qCD, 5d1xBA, 5d1zBA, 5d1zDC, 5d1zFE, 5d1zGH, 5d72HL, 5d7sHL, 5d8jHL, 5d93CB, 5d93FE, 5d96CB, 5d96JI, 5dd0HL, 5dd5HL, 5dfvCD, 5dfvEF, 5dhvAB, 5dhvHL, 5dhyAB, 5dhyHL, 5dhzHL, 5dk3BA, 5dk3GF, 5dlmHL, 5dlmIM, 5dmgCD, 5dmgEF, 5dmgHL, 5dmiHL, 5do2CD, 5do2HL, 5dq9AB, 5dq9CD, 5dq9HL, 5dqdHL, 5dqjAB, 5dqjHL, 5dr5HL, 5drwAB, 5drxAB, 5drxHL, 5drzBA, 5drzHL, 5dumHL, 5durBD, 5durHL, 5dwuHL, 5dyoAB, 5dyoHL, 5e08HL, 5e1aAB, 5e2tHL, 5e2vHL, 5e2wHL, 5e8eBA, 5e94BA, 5e94DC, 5e99HL, 5ea0HL, 5eblAB, 5ebmAB, 5ebwAB, 5ec1AB, 5ec2AB, 5eiiAB, 5eiiHL, 5en2AB, 5eoqHL, 5eorHL, 5epmAB, 5epmEF, 5erwAB, 5esaAB, 5esqBA, 5esvAB, 5esvCD, 5esvHL, 5eszHL, 5etuBA, 5etuDC, 5eu7EC, 5eu7FD, 5eukBA, 5eukDC, 5ewiHL, 5eziHL, 5ezjAB, 5ezlAB, 5ezlHL, 5ezoHL, 5f3bAB, 5f3bEF, 5f3hCD, 5f3hEF, 5f3hGH, 5f6hIJ, 5f6hKL, 5f6hMN, 5f6hOP, 5f6iBC, 5f6jBA, 5f6jHF, 5f7eHL, 5f88BA, 5f88DC, 5f89AB, 5f89HL, 5f96HL, 5f9oHL, 5f9wBC, 5f9wHL, 5fb8BA, 5fcuHL, 5fecAB, 5fehFE, 5ff6BA, 5ff6DC, 5fgcEB, 5fhaHL, 5fhbHL, 5fhcAB, 5fhcHL, 5fuoHL, 5fuzHL, 5fylDE, 5g64HL, 5ggqHL, 5ggrAB, 5ggrHL, 5ggsAB, 5ggsCD, 5ggtHL, 5gguAB, 5gguHL, 5ggvHL, 5ghwHL, 5girAB, 5girHL, 5gisHL, 5gkrAB, 5gksAB, 5gksCD, 5gmqBC, 5grjHL, 5grvLK, 5gs0DC, 5gs0FE, 5gs0HL, 5gs0XY, 5gs1BA, 5gs1NM, 5gs1PO, 5gs1RQ, 5guxHL, 5gz0BA, 5gz0DC, 5gznCD, 5gznHL, 5gzoCD, 5gzoHL, 5h2bAB, 5h35AB, 5h35FG, 5h35HI, 5h37GH, 5h37IM, 5h37KL, 5hbtDC, 5hbvDC, 5hd8CD, 5hd8EF, 5hdbEF, 5hdbHL, 5hdqHL, 5hhvHL, 5hhxHL, 5hi3CD, 5hi3HL, 5hi4CD, 5hi4HL, 5hi5CD, 5hi5HL, 5hj3AB, 5hj3EF, 5hj3IJ, 5hj3MN, 5hpmBA, 5hpmDC, 5hyqBA, 5hyqDC, 5hysAB, 5hysCD, 5hysEF, 5hysHL, 5i15HL, 5i16BA, 5i16HL, 5i17HL, 5i18HL, 5i19HL, 5i1aBA, 5i1aHL, 5i1cHL, 5i1dHL, 5i1eHL, 5i1gHL,

5i1hHL, 5i1lBA, 5i1lHL, 5i2iBA, 5i2iDC, 5i30HL, 5i5kHL, 5i5kXY, 5i66AB, 5i6xBC, 5i6zBC, 5i71BC, 5i73BC, 5i74BC, 5i75BC, 5i76BA, 5i76DC, 5i9qBC, 5i9qHL, 5iblCD, 5iblHL, 5ibtHL, 5ibuGB, 5ibuHL, 5icxBA, 5icxDC, 5icyBA, 5icyDC, 5iczBA, 5iczDC, 5id0BA, 5id0DC, 5id1BA, 5id1DC, 5ifhHL, 5ifjAB, 5ifjDE, 5ifjGH, 5ifjJK, 5ig7AB, 5ig7DE, 5ig7GH, 5ig7JK, 5igxHL, 5ihuHL, 5iieHL, 5ijkAC, 5ijkBD, 5ijvBA, 5ijvDC, 5ijvFE, 5ijvHL, 5ik3CD, 5ikcBA, 5ikcHL, 5iltHL, 5iopBA, 5iopDC, 5iq7HL, 5iq9AB, 5iq9HL, 5ir1BA, 5ir1DC, 5it2HL, 5itfBA, 5itfDC, 5iv2BA, 5iv2DC, 5ivzBA, 5ivzDC, 5j13CB, 5j3hCD, 5j9pAB, 5jhlHL, 5jnyAB, 5jnyCD, 5jo4AB, 5jo4CD, 5jofAB, 5jofCD, 5jofEF, 5jofHL, 5jopED, 5jopHL, 5jorBC, 5jorDA, 5jorFE, 5jorHL, 5jq6HL, 5jqdDA, 5jr1HL, 5jrpHL, 5jsaEF, 5jueHL, 5jw5HL, 5jxaHL, 5jxeDC, 5jxeGF, 5jz7CD, 5jz7GI, 5jz7HL, 5jz7JK, 5k59EF, 5k59HL, 5k8aBA, 5k8aFE, 5k8aHL, 5k8aVM, 5k9jHL, 5k9kAB, 5k9kHL, 5k9qGI, 5k9qHL, 5k9qJK, 5k9qST, 5k9qUV, 5k9qXY, 5kelCD, 5kelHL, 5kelJN, 5kelMO, 5kelPT, 5kelQU, 5kemDE, 5kemIJ, 5kenCD, 5kenGH, 5kenJI, 5kenNO, 5kenQP, 5kg9HL, 5kjrHL, 5kmvEI, 5kmvHL, 5kn5DE, 5kqvCD, 5kqvPQ, 5kteHL, 5kvdHL, 5kveHL, 5kvfHL, 5kvgHL, 5kw9HL, 5kzcBD, 5kzcEG, 5kzcHL, 5kzpEI, 5kzpFJ, 5kzpGK, 5kzpHL, 5l0qCB, 5l0qFE, 5l6yHL, 5l7xHL, 5l88HL, 5l9dHL, 5lbsHL, 5lbsIM, 5lcvHL, 5ldnHL, 5lghHL, 5lqbHL, 5lspHL, 5lspST, 5lwyHL, 5lxgHL, 5m63HL, 5m63MN, 5mesHL, 5mevHL, 5mheHL, 5mhgHL, 5mhrIH, 5mhrML, 5mhrNG, 5mhrPO, 5mhrRQ, 5mhsNG, 5mi0BC, 5mo3HL, 5mo9HL, 5mp1CD, 5mp1FG, 5mp1HL, 5mp3AB, 5mp3HL, 5mp5AB, 5mp5CD, 5mp5EF, 5mp5HL, 5mp6HL, 5mthAB, 5mthHL, 5mu0AB, 5mu0CD, 5mu0EF, 5mu0GH, 5mu0IJ, 5mu0KL, 5mu0MN, 5mu0OP, 5mu2AB, 5mu2CD, 5mu2EF, 5mu2GH, 5mu2IJ, 5mu2KL, 5mu2MN, 5mu2OP, 5mubAB, 5mubCD, 5mubFG, 5mubIJ, 5mubLM, 5mubOP, 5mubRS, 5mubUV, 5muqAB, 5muqHL, 5mv3AB, 5mv3CD, 5mv3FG, 5mv3IJ, 5mv3LM, 5mv3OP, 5mv3RS, 5mv3UV, 5mv4AB, 5mv4CD, 5mv4FG, 5mv4IJ, 5mv4LM, 5mv4OP, 5mv4RS, 5mv4UV, 5mvjAB, 5mvjCD, 5mvjEF, 5mvjHL, 5mvzAB, 5mvzHL, 5mx3AB, 5mx3HL, 5my4BA, 5mykBA, 5myoBA, 5myoDC, 5myxBA, 5myxDC, 5n09HL, 5n09IM, 5n0aHL, 5n0aIM, 5n2kBA, 5n2kDC, 5n2kFE, 5n2kHG, 5n2kJI, 5n2kLK, 5n2kNM, 5n2kPO, 5n4gHL, 5n4jHL, 5n7wAB, 5n7wHL, 5nb5HL, 5nb5IM, 5nbiHL, 5nbwBA, 5nbwHL, 5ngvHL,

5nh3HL, 5nh3IM, 5nhrAB, 5nhrHL, 5nhwHL, 5nivBA, 5nj3CD, 5nj3EF, 5nj6HL, 5njdNM, 5njdPO, 5njdRQ, 5njdTS, 5njdVU, 5njdXW, 5njgCD, 5njgEF, 5nmvHL, 5nphHL, 5nuzAB, 5nuzHL, 5nyxHL, 5nyxMN, 5o14CD, 5o14HL, 5o1rHL, 5o1rIM, 5o4gBA, 5o4oBA, 5o6vHL, 5o6vIM, 5o7pBA, 5ob5HL, 5obfHL, 5ocaHL, 5occHL, 5ockHL, 5ocxHL, 5ocyHL, 5od0HL, 5od8AB, 5odbAB, 5opyHL, 5otjBA, 5sx4HL, 5sx4JI, 5sx5HL, 5sx5JK, 5sy8HL, 5szfHL, 5t0yAB, 5t0yHL, 5t0yIG, 5t1dHL, 5t1kBA, 5t1kDC, 5t1lBA, 5t1lDC, 5t1mBA, 5t1mDC, 5t33HL, 5t3sDE, 5t3xDE, 5t3zDE, 5t4zAB, 5t4zHL, 5t5bHL, 5t5fHL, 5t5nGF, 5t5nIH, 5t5nKJ, 5t5nML, 5t5nON, 5t6lAB, 5t6lHL, 5t6pBA, 5t6pDC, 5t78BA, 5t78DC, 5t80HL, 5t85HL, 5tbdAB, 5tbdDG, 5tbdEF, 5tbdJK, 5tdnDC, 5tdoDC, 5tdpDC, 5te4HL, 5te6HL, 5te7HL, 5tf1HL, 5tfsHL, 5tfwHL, 5tgbHL, 5th2BA, 5th2DC, 5th9HL, 5th9IM, 5tihHL, 5tkjAB, 5tkjDE, 5tkjGH, 5tkjJK, 5tkkHL, 5tl5HL, 5tljBA, 5tljDC, 5tljFE, 5tljHG, 5tlkBA, 5tlkDC, 5tlkFE, 5tlkHG, 5tplHL, 5tpnHL, 5tppHL, 5tqaAB, 5tqaHL, 5tqqHL, 5tqqIM, 5tr1HL, 5tr1IM, 5tr8HL, 5trpHL, 5tudCB, 5tudFE, 5ty6HL, 5tz2HL, 5tztAB, 5tztHL, 5tzuHL, 5u0rHL, 5u15AB, 5u15HL, 5u3dBA, 5u3jHL, 5u3kAB, 5u3kHL, 5u3lAB, 5u3lHL, 5u3mHL, 5u3nHL, 5u3oHL, 5u3pHL, 5u4rAB, 5u4rHL, 5u5fBA, 5u5mBA, 5u6aBA, 5u7mDE, 5u7oDE, 5u8qHL, 5u8rHL, 5ubyHL, 5ubzHL, 5ucbHL, 5udcBC, 5udcEG, 5udcHL, 5uddJP, 5uddKQ, 5uddLR, 5uddMS, 5uddNT, 5ueaAB, 5ueaHL, 5uekHL, 5uelHL, 5uemHL, 5ug0DC, 5ugyHL, 5ugyIM, 5ugyJN, 5uixHL, 5uknHL, 5ukoHL, 5ukpHL, 5ukqHL, 5ukrHL, 5um8DE, 5umiAB, 5umiHL, 5umnCD, 5umnEF, 5uoeHL, 5uoeMN, 5uoePQ, 5uoeST, 5uoeVW, 5uqyDC, 5uqyHG, 5uqyLK, 5uqyPO, 5ur8AB, 5ushDE, 5ushHL, 5usiAB, 5usiHL, 5uslHL, 5utfDE, 5utyDE, 5utzBC, 5utzFG, 5utzHL, 5utzJK, 5uxqHL, 5v2aHL, 5v6lHL, 5v6lIM, 5v6mHL, 5v7jDE, 5v7rHL, 5v7uHL, 5v8lGK, 5v8lHL, 5v8lIM, 5v8lJN, 5v8mHL, 5v8mRT, 5v8mSU, 5vagCB, 5vcnDC, 5vcnFE, 5vcoBA, 5vcoDC, 5vebAB, 5vebHL, 5vgaBA, 5vgjHL, 5vh3AB, 5vh3HL, 5vh4AB, 5vh4HL, 5vicHL, 5vigAB, 5vigHL, 5viyJI, 5viyKL, 5viyMN, 5viyOP, 5vj6MN, 5vj6OP, 5vj6QR, 5vjoAB, 5vjoCD, 5vjqAB, 5vjqCD, 5vjqEF, 5vjqGH, 5vk2DE, 5vk2FG, 5vk2HL, 5vk6AB, 5vkdHL, 5vkeAB, 5vkhAB, 5vkkHL, 5vl3AB, 5vl3CD, 5vl3HL, 5vl7HL, 5vlpHL, 5vobHL, 5vocHL, 5vodHL, 5vpgDC, 5vphDC, 5vplDC, 5vqmDC,

5vqmHL, 5vr9HL, 5vshAB, 5vshCD, 5vshEF, 5vshHL, 5vsiHL, 5vtaFE, 5vtaHL, 5vvfHL, 5vyfBA, 5vyfHL, 5vzxEI, 5vzxHL, 5vzyHL, 5w05HL, 5w06HL, 5w08GH, 5w08IJ, 5w08KL, 5w08MN, 5w08OP, 5w08QR, 5w0kEF, 5w0kHL, 5w1gHL, 5w1kBA, 5w1kDC, 5w1kGF, 5w1kIH, 5w1kLK, 5w1kON, 5w1kQM, 5w1kTS, 5w1mBA, 5w1mDC, 5w1mFE, 5w1mHG, 5w23HL, 5w23IM, 5w23JN, 5w2bHL, 5w3eEG, 5w3lEG, 5w3mEG, 5w3oDE, 5w3pHL, 5w42HL, 5w4lBC, 5w4lDE, 5w4lFI, 5w4lHL, 5w5xHL, 5w5zHL, 5w6cHL, 5w6dDE, 5w6gHL, 5wawHL, 5wcaHL, 5wcdHL, 5wdfAB, 5wdfHL, 5wduDE, 5wduMN, 5wduOP, 5wduUV, 5wduWX, 5weqBA, 5whjHL, 5whkHL, 5wi9FE, 5wi9HL, 5wk2HL, 5wk3QP, 5wk3SR, 5wk3UT, 5wk3WV, 5wkoAB, 5wkoCD, 5wkoEF, 5wkoGH, 5wkoIJ, 5wkoKL, 5wkzHL, 5wl2AB, 5wnaCD, 5wnaHL, 5wnbIM, 5wt9HL, 5wtgBA, 5wtgDC, 5wthED, 5wttAB, 5wttHL, 5wuvHL, 5wuxAB, 5wuxCD, 5wuxHL, 5x0tAB, 5x0tCD, 5x2mHL, 5x2mJK, 5x2nHL, 5x2nJK, 5x2oHL, 5x2oJK, 5x2pHL, 5x2pJK, 5x2qHL, 5x2qJK, 5x4gAB, 5x5xHL, 5x8lFK, 5x8lGL, 5x8lHM, 5x8lJO, 5x8lSN, 5x8mBC, 5xajAB, 5xajEC, 5xajFD, 5xajHL, 5xbmBA, 5xbmED, 5xcqAB, 5xcrAB, 5xcrDE, 5xcsAB, 5xctAB, 5xcuAB, 5xcvDE, 5xcxAB, 5xezCD, 5xezHL, 5xf1CD, 5xf1HL, 5xhfBA, 5xhfDC, 5xhgBA, 5xhgDC, 5xhvHL, 5xhvPQ, 5xj3AB, 5xj3DE, 5xj3GH, 5xj3JK, 5xj4HL, 5xjmHL, 5xkuCB, 5xlica, 5xlidb, 5xmhDC, 5xmhHL, 5xqwHL, 5xrqAB, 5xrqHL, 5xwdHD, 5xxyHL, 5y11AB, 5y2kAB, 5y2kEF, 5y2lIJ, 5y2mIJ, 5y2mKL, 5y9cHL, 5y9fLK, 5y9fNM, 5y9fPO, 5y9fRQ, 5y9fTS, 5y9fVU, 5y9fXW, 5y9fZY, 5y9fba, 5y9fdc, 5y9jHL, 5y9kHL, 5ye3BA, 5ye4AB, 5ye4CD, 5yhlHL, 5yoyGD, 5yoyHE, 5yoyIF, 5yoyPM, 5yoyQN, 5yoyRO, 5yslAB, 5yslCD, 5yslEF, 5yslGH, 5yupHL, 5ywyHL, 5yy5HL, 5ziaAB, 5ziaDE, 5ziaGL, 5ziaHI, 5ziaKM, 5ziaOP, 5zmjHL, 5zs0BA, 5zufED, 5zv3HL, 6a0xAB, 6a0xCD, 6a0xEF, 6a0xGH, 6a0zHL, 6a3wAB, 6a3wDE, 6a3wGH, 6a3wJK, 6a4kHL, 6a4kIM, 6a4kJN, 6a4kKO, 6a67CD, 6a67HL, 6a76HL, 6a77HL, 6a78HL, 6a78IM, 6a79HL, 6a79IM, 6a9kBA, 6a9kHL, 6ad7CD, 6ad7EF, 6ad8CD, 6ad8EF, 6adaCD, 6adaEF, 6adbCD, 6adbEF, 6adcCD, 6adcEF, 6al0HL, 6al1HL, 6al4BA, 6al4DC, 6al4FE, 6al5HL, 6anaHL, 6aniHL, 6aniIM, 6apbHL, 6apcHL, 6apdDE, 6apdFG, 6apdHI, 6apdJL, 6apdKM, 6apdNO, 6aq7HL, 6arpBA, 6arpDC, 6aruCB, 6attHL, 6au5BA, 6au5DC, 6avnHL, 6avqHL, 6avrHL, 6avuHL, 6awnBC, 6awoBC,

6awpBC, 6awqBC, 6axkAB, 6axlAB, 6axlDC, 6axpBA, 6axpDC, 6aynBA, 6aynDC, 6ayzBD, 6ayzRC, 6az2AF, 6az2CE, 6azkBA, 6azkDC, 6azlBA, 6azlDC, 6azmBA, 6azmDC, 6azxAB, 6azxCD, 6azzCB, 6azzFE, 6b08CB, 6b0aHL, 6b0eBA, 6b0gDC, 6b0hBA, 6b0hDC, 6b0nDE, 6b0sHL, 6b0wHL, 6b3dHL, 6b3sCD, 6b3sFG, 6b3sHL, 6b3sJK, 6b5lHL, 6b5mHL, 6b5nHL, 6b5oHL, 6b5pHL, 6b5rHL, 6b5sHL, 6b5tHL, 6b70CD, 6b70EF, 6b7zCD, 6b7zEF, 6b9yBA, 6b9zBA, 6baeBA, 6bahBA, 6bb4HL, 6bb4IM, 6bb4JN, 6bckHL, 6bdzHL, 6be2HL, 6be3HL, 6be4HL, 6bf4BC, 6bf4HL, 6bf7CD, 6bf7EF, 6bf9CD, 6bfqAB, 6bfqCD, 6bfqEF, 6bfqHL, 6bfsHL, 6bftAB, 6bftHL, 6bgtBA, 6bhzHL, 6bi0HL, 6bi2HL, 6bi2IM, 6bitIK, 6bitJL, 6bkcHL, 6blaHL, 6blhHL, 6bliAB, 6bliDE, 6bliGH, 6bliJK, 6bp2HL, 6bpaBC, 6bpaEF, 6bpcBC, 6bpcEF, 6bqbHL, 6bt3BA, 6bt3DC, 6bt3FE, 6bt3HG, 6btjAB, 6btjHL, 6by2AB, 6by3AB, 6bzuAB, 6bzuCD, 6bzuEF, 6bzuGH, 6bzvAB, 6bzvCD, 6bzvEF, 6bzvGH, 6bzwAB, 6bzwCD, 6bzwEF, 6bzwGH, 6bzyHL, 6c08AB, 6c08DE, 6c5hHL, 6c5iHL, 6c5jHL, 6c5kAB, 6c5kHL, 6c5vHL, 6c6xAB, 6c6xCD, 6c6xEF, 6c6xHL, 6c6yAB, 6c6yHL, 6c6zCD, 6c6zHL, 6c9uHL, 6cbjHL, 6cbvHL, 6cde56, 6cdeQR, 6cdi34, 6cdi87, 6cdihl, 6cdiqr, 6cdmAB, 6cdmDE, 6cdoAB, 6cdpAB, 6ce0DE, 6cezHL, 6cf2AB, 6ch7DE, 6ch7QR, 6ch8DE, 6ch8QR, 6chbDE, 6chbMN, 6chbOP, 6cm3PQ, 6cm3RS, 6cm3TU, 6cmgCB, 6cmiDC, 6cnjGF, 6cnjIH, 6cnjKJ, 6cnkGF, 6cnkKJ, 6cnrHL, 6cnvHL, 6co3HL, 6cr1HL, 6crqDE, 6crqIK, 6crqJL, 6cseAB, 6cseHL, 6csfAB, 6csfHL, 6ct7AB, 6ct7HL, 6cue34, 6cuehl, 6cuf34, 6cuf87, 6cufQR, 6cw2AB, 6cw3AB, 6cw3CD, 6cxcGJ, 6cxcIK, 6cxcXY, 6cxgHL, 6cxgMK, 6cxlHL, 6cxlMK, 6cxyHL, 6cyfDC, 6cyfFE, 6cyfML, 6cyfUT, 6d01AB, 6d01CD, 6d01EF, 6d01GH, 6d0uCD, 6d0uEF, 6d0uHI, 6d0uKL, 6d0xAB, 6d11AB, 6d11CD, 6d6tJI, 6d6tKL, 6d6uJI, 6d6uKL, 6d9gAB, 6d9gCD, 6d9wHL, 6db5HL, 6db5IM, 6db6HL, 6db7IM, 6dc3HL, 6dc4HL, 6dc5BC, 6dc5EF, 6dc5HI, 6dc7HL, 6dc7IM, 6dc8HL, 6dc9HL, 6dc9IM, 6dcaHL, 6dcaIM, 6dcaJN, 6dcaKO, 6dcqHL, 6dcqMN, 6dcvBA, 6dcvHL, 6dcwHL, 6ddmBA, 6ddrBA, 6ddvBA, 6de7DE, 6dezHL, 6df0HL, 6df1HL, 6df2BA, 6df2HL, 6dfgGJ, 6dfgHL, 6dfgIK, 6dfhGJ, 6dfhHL, 6dfhIK, 6dfiHL, 6dfjHL, 6dg2AB, 6dg2CD, 6djpCD, 6djpEF, 6dkjAB, 6dkjHL, 6dlaHL, 6dlbHL, 6dn0AB, 6dw2BA, 6dwiBA, 6dwiDC, 6dwiHG, 6dwiJI, 6dwiLK, 6dwiNM, 6dwiPO, 6dzlGJ, 6dzlHK,

6dzlIL, 6dzmGH, 6dzmIK, 6dzmJL, 6dznHL, 6dzrHL, 6e3hHL, 6e4xZY, 6e56GI, 6e56HJ, 6e62BC, 6e62HL, 6e63BC, 6e63HL, 6e64AB, 6e64HL, 6e65HL, 6e9iAB, 6e9iCD, 6e9iHL, 6e9uAB, 6ea5MN, 6ea5QR, 6ea7QR, 6eduJK, 6eduLM, 6eduNO, 6eduPQ, 6eduRS, 6eduTU, 6eleHL, 6eljHL, 6ellHL, 6eluBC, 6eluEF, 6eluHI, 6eluKL, 6emjHL, 6erxDE, 6erxFG, 6erxHI, 6erxMN, 6erxOP, 6erxQR, 6etiDC, 6etiFE, 6ev1AB, 6ev1CD, 6ev1EF, 6ev1GH, 6ev1KL, 6ev2AB, 6ev2CD, 6ev2EF, 6ev2GH, 6ewbEF, 6ewbGI, 6ewbHL, 6ewbJK, 6eyoHL, 6f7tBA, 6f7tHL, 6faxHL, 6fbjHL, 6feqDC, 6feqFE, 6fg1AD, 6fg1HL, 6fg2DE, 6fg2FG, 6fg2HL, 6fg2IM, 6fgbHL, 6flaAB, 6flaHL, 6flbHL, 6flcBA, 6flcHL, 6fn1CB, 6fn4CB, 6foeAB, 6foeHL, 6fxnDE, 6fxnFG, 6fxnHI, 6fxnMN, 6fxnOP, 6fxnQR, 6fy0HL, 6fy0IM, 6fy1HL, 6fy2HL, 6fy2XY, 6fy3HL, 6fy3XY, 6gc2HL, 6gfeHL, 6gfeKM, 6gffLK, 6ghgAB, 6ghgHL, 6gk7HL, 6gk8HL, 6gkuHL, 6glwCD, 6glwHL, 6gv4HL, 6gxxHL, 6h06AB, 6h06CD, 6h06EF, 6h06HL, 6h0eAB, 6h0eCD, 6h0eEF, 6h0eHL, 6h2yHL, 6h3hAB, 6h3hCD, 6h3tHL, 6h3tIM, 6h3uHL, 6h3uIM, 6h5nCB, 6h5nFE, 6hcoDC, 6hcoFE, 6hf1CB, 6hf1FE, 6hgaHL, 6hguAB, 6hguHL, 6hhcHL, 6hjpGH, 6hjpIJ, 6hjpKL, 6hjqIJ, 6hjqKL, 6hkgAB, 6hkgCD, 6hx4HL, 6hx4IM, 6hxwCD, 6hxwHL, 6i04CD, 6i04HL, 6i1oHL, 6i8sEI, 6i8sFJ, 6i8sGK, 6i8sHL, 6i9iAB, 6i9iHL, 6iapED, 6iapHL, 6iasHL, 6iccHL, 6icfHL, 6id4CD, 6id4HL, 6idgHL, 6idhHL, 6ieaHL, 6iebHL, 6iekBC, 6iekEF, 6ieqDE, 6iguHL, 6ii4EF, 6ii4HL, 6ii8GI, 6ii8HL, 6ii9HL, 6ii9MN, 6ii9PQ, 6ionHL, 6iuvCD, 6iuvHL, 6j11DE, 6j11FG, 6j11HL, 6j5dHL, 6j5fHL, 6j5gHL, 6j6yBC, 6j6yEF, 6jbtHL, 6jc2HL, 6jepHL, 6jepKI, 6jhqDE, 6jhsED, 6jhtED, 6jjpAB, 6jjpDE, 6jodHL, 6jp7HL, 6ju0HL, 6ju0IM, 6ju0JN, 6ju0KO, 6jwcHL, 6k0yAB, 6k5aCD, 6k5aEF, 6k5dCD, 6k5dEF, 6k5fCD, 6k5fEF, 6k5iCD, 6k5iEF, 6k65HL, 6k7oBC, 6k7oDE, 6k7oFG, 6k7oHL, 6ko5HL, 6krzDE, 6krzFG, 6krzHI, 6ks0HL, 6ks1HL, 6ks2BA, 6ks2DC, 6ktrAB, 6ktrFG, 6kvahl, 6kvfAC, 6kx0AB, 6kx1AB, 6kyzBC, 6kyzEF, 6kz0BC, 6kz0EF, 6kz0HI, 6kz0KL, 6l62HL, 6l8tAB, 6l8tCD, 6l8tHL, 6lbhCD, 6lbhEF, 6lcsAB, 6ldwBA, 6ldwHL, 6ldxHL, 6ldyAB, 6ldyHL, 6lhqHL, 6lhtHL, 6lktAB, 6ln2CB, 6lraHL, 6lunAB, 6lunCD, 6lxiCD, 6lxiHL, 6lxjHL, 6lxkEF, 6lxkHL, 6lxkJK, 6lynAB, 6lynHL, 6lz4AB, 6lz9HL, 6m0fHL, 6m0zHL, 6m2rHL, 6m38HL, 6m3cFE, 6m3cHL, 6m3cKJ, 6m3zHL, 6m47HL, 6m58CD, 6m58HL,

6m87BA, 6m87DC, 6m87FE, 6m87HG, 6m87JI, 6m87LK, 6mamAB, 6mamCD, 6mamEF, 6marHL, 6marMN, 6mb3AK, 6mb3BM, 6mb3CN, 6mb3DO, 6mb3FP, 6mb3GQ, 6mb3HL, 6mb3IR, 6mb3JS, 6mcoDE, 6mdtDE, 6me1AB, 6me1CD, 6medAB, 6medCD, 6medGI, 6meeCD, 6mefHL, 6megHL, 6mehHL, 6meiHL, 6mejAB, 6mejHL, 6mekBD, 6mekFE, 6mekHL, 6mekIG, 6mfjAB, 6mfjHL, 6mfpCD, 6mfpHL, 6mftAB, 6mftHL, 6mg7HL, 6mhgAN, 6mhgBO, 6mhgCP, 6mhgDQ, 6mhgFR, 6mhgGS, 6mhgHL, 6mhgIT, 6mhgJU, 6mhgKV, 6mhgMW, 6mhrAB, 6mhrDE, 6mi2AB, 6mi2DE, 6midHL, 6mjzHL, 6ml8HL, 6mlkHL, 6mlmEI, 6mlmFG, 6mlmKL, 6mnfHL, 6mnfMK, 6mnqHL, 6mnrHL, 6mnsAB, 6mnsHL, 6mpg34, 6mpg87, 6mpgWB, 6mpgXY, 6mph12, 6mph34, 6mphHL, 6mphQR, 6mphfh, 6mphgi, 6mqcAB, 6mqeAB, 6mqeHL, 6mqmAB, 6mqmDE, 6mqmGH, 6mqmJK, 6mqrHL, 6mqsAB, 6mqsCD, 6msyHL, 6mtoHL, 6mtpAB, 6mtpHL, 6mtqHL, 6mtrHL, 6mtsHL, 6mtxHL, 6mtyHL, 6mu3HL, 6mu3MK, 6mubHL, 6mubMK, 6muiCD, 6muiGH, 6muiKL, 6muiOP, 6mw9CD, 6mw9GH, 6mw9KL, 6mw9OP, 6mwcCD, 6mwcGH, 6mwcKL, 6mwcOP, 6mwnCD, 6mwnHL, 6mwvCD, 6mwvGH, 6mwvKL, 6mwvOP, 6mwxCD, 6mwxGH, 6mwxKL, 6mwxOP, 6mxrAB, 6mxrHL, 6mxsAB, 6mxsHL, 6my4AB, 6my4HL, 6my5AB, 6my5HL, 6n16AB, 6n16CD, 6n16HL, 6n16KJ, 6n1v12, 6n1v34, 6n1vHL, 6n1w34, 6n1w87, 6n2xHL, 6n2xMK, 6n32HL, 6n32KM, 6n35HL, 6n35MK, 6n4qJI, 6n4qLK, 6n4rJI, 6n4rLK, 6n5bHL, 6n5dCD, 6n5dEF, 6n5dLN, 6n5eED, 6n5eHI, 6n6bKL, 6n7jAB, 6n7jHL, 6n81CD, 6n81HL, 6n8dDB, 6n8dFE, 6nb3HL, 6nb8HL, 6nc2HL, 6nc2OT, 6nc2PU, 6nc2QV, 6nc2RW, 6nc2SX, 6nc3HL, 6nc3OT, 6nc3PU, 6nc3QV, 6nc3RW, 6nc3SX, 6ncpAB, 6ncpCD, 6nf2CD, 6nf2HL, 6nf2NM, 6nf2OP, 6nf2VU, 6nf2WX, 6nf5GI, 6nf5KN, 6nf5QR, 6nfcCD, 6nfcHL, 6nfcJM, 6nfcKN, 6nfnBA, 6nfnDC, 6nfnFE, 6nfnHL, 6nfnJI, 6nfnNM, 6nfnPO, 6nfnRQ, 6nfuAB, 6nfvAB, 6nhaHL, 6ni2HL, 6nijHL, 6nipAB, 6nipHL, 6nisHL, 6niuFG, 6niuHL, 6niuIJ, 6niuMN, 6njlKJ, 6njlNM, 6njmLK, 6njmPO, 6njnKJ, 6njnON, 6nm6UV, 6nmrAB, 6nmrFG, 6nmrHL, 6nmrJK, 6nmsBA, 6nmsHL, 6nmtBA, 6nmuBA, 6nmuHL, 6nmvHL, 6nnfUV, 6nnjUV, 6novCD, 6nqdCD, 6nqdGH, 6nqdKL, 6nuyHL, 6nv0DC, 6nv0HL, 6nyqHL, 6nz7GI, 6nz7HL, 6o1fHL, 6o23AB, 6o23CD, 6o24AB, 6o25AB, 6o25EF, 6o25IJ, 6o26AB, 6o28AB, 6o28CD, 6o29AB,

6o29HL, 6o2aAB, 6o2aCD, 6o2aHI, 6o2aOP, 6o2bAB, 6o2bCD, 6o2bHI, 6o2bOP, 6o2cAB, 6o39BA, 6o3aBA, 6o3bBA, 6o3bGE, 6o3dHL, 6o3gAB, 6o3gCD, 6o3gEF, 6o3gHL, 6o3jAB, 6o3jHL, 6o3kHL, 6o3lAB, 6o3lHL, 6o3uBA, 6o3uDC, 6o3uFE, 6o3uHL, 6o41BA, 6o41DC, 6o41HL, 6o42BA, 6o9hAB, 6o9hHL, 6o9iAB, 6obdBA, 6obdHL, 6obzAB, 6obzHL, 6oc3AB, 6oc3CD, 6ocbHL, 6oe4BC, 6oe4EF, 6oe5BC, 6oe5HL, 6oelHL, 6ofiHL, 6ogeCB, 6ogeED, 6ogxCD, 6ogxHL, 6okmHL, 6oknAB, 6oknCD, 6okpMN, 6okpOP, 6okpQR, 6okqAB, 6okqDC, 6okqFE, 6ol5AB, 6ol5HL, 6ol6AB, 6ol6HL, 6ol7AQ, 6ol7CB, 6ol7ED, 6ol7HL, 6ol7IF, 6ol7KJ, 6ol7MN, 6ol7PO, 6olpGI, 6olpHL, 6oorHL, 6opaFI, 6opaKJ, 6opaSR, 6opoHL, 6opoIE, 6opoOK, 6oppGI, 6oppHL, 6oppNO, 6opqHL, 6opqKN, 6opqMO, 6oroDE, 6oroHL, 6oroJK, 6orpDE, 6orpHL, 6orpJK, 6orqDE, 6orqHL, 6orqJK, 6osvHL, 6osy87, 6osyFE, 6osyPO, 6otcHL, 6ousMN, 6ousOP, 6ousQR, 6ousST, 6ousUV, 6ousWX, 6oy4DC, 6oz2HL, 6oz4HL, 6oz9HL, 6ozcCD, 6ozcHL, 6ozcJM, 6ozcKN, 6ozcOQ, 6ozcPR, 6p3rAB, 6p3rCD, 6p3sAB, 6p3sFG, 6p3sIJ, 6p3sMN, 6p3sST, 6p4aHL, 6p4bBA, 6p4bHL, 6p4cHL, 6p4yHL, 6p50HL, 6p60AB, 6p60CD, 6p60GH, 6p60JK, 6p65CD, 6p65FG, 6p65HL, 6p67AB, 6p67CD, 6p67EF, 6p67HL, 6p79HL, 6p8dAB, 6p8dDE, 6p91HL, 6p95DE, 6p95FG, 6p95HL, 6p9hCD, 6p9hHL, 6p9iAB, 6p9iCD, 6p9jHL, 6pa0AB, 6pbvBA, 6pbvDC, 6pciHK, 6pciIL, 6pciJM, 6pciNQ, 6pciOR, 6pciPS, 6pcuCB, 6pcuFE, 6pcuIH, 6pdsBA, 6pdsHL, 6pduHL, 6pdxHL, 6pdxMN, 6pdxOP, 6pe7HL, 6pe8AB, 6pe8HL, 6pe9AB, 6pe9CD, 6pe9EF, 6pe9HL, 6pe9KM, 6pefAB, 6pefDE, 6pehAB, 6phbBA, 6phbDC, 6phcAB, 6phcCD, 6phfCD, 6phhAB, 6phhCD, 6pi7CB, 6pi7FE, 6pisHL, 6pisIM, 6plhBA, 6plkHL, 6plkIM, 6ppgBA, 6ppgHL, 6pv7FG, 6pv7HI, 6pv8FG, 6pv8HI, 6pwcHL, 6pwuHL, 6pxgAB, 6pxgCD, 6pxgEF, 6pxgHL, 6pxhCD, 6pxrHL, 6pycAB, 6pzeHL, 6pzfFE, 6pzfHL, 6pzgHL, 6pzhBA, 6pzhHL, 6pzwFE, 6pzwHL, 6pzwJG, 6pzwKI, 6pzyCD, 6pzyGJ, 6pzyHK, 6pzyIL, 6pzzFE, 6pzzHL, 6pzzIG, 6pzzKJ, 6q0eHL, 6q0hHL, 6q0iHL, 6q0lHL, 6q0oZB, 6q18HL, 6q19HL, 6q1eHL, 6q1gHL, 6q1jHL, 6q1kHL, 6q1zDC, 6q1zHL, 6q20HL, 6q23EF, 6q23GI, 6q23HL, 6q23JK, 6qb6HL, 6qbcHL, 6qcuHL, 6qd7HL, 6qd7UP, 6qd7XZ, 6qd8HL, 6qd8QY, 6qd8WZ, 6qeeCB, 6qexCB, 6qigHL, 6qn7AB, 6qn7HL, 6qn8AB, 6qn8CD, 6qn8EF, 6qn8GI, 6qn8HL, 6qn8JK, 6qn9HL, 6qnaAB, 6qnaCD,

6qnaEF, 6qnaHL, 6qnkBA, 6qnkDC, 6qnoHL, 6qu9AB, 6qu9HL, 6r0xAB, 6r0xCD, 6r2sBA, 6r8xCB, 6rcoAB, 6rcoCD, 6rcqAB, 6rcuBC, 6rcvCB, 6rcvED, 6rcvHG, 6rcvJI, 6rlmAB, 6rlmCD, 6rlmEF, 6rlmGH, 6rlmIJ, 6rlmKL, 6rlmMN, 6rlmOP, 6rnnDC, 6rnnHL, 6rp8HL, 6rpsHL, 6rpsNM, 6s3dAB, 6s3dCD, 6s3dEF, 6s3dHL, 6s3tDC, 6s3tEI, 6s3tNM, 6s3tRQ, 6s5aHL, 6s8dHL, 6s8dPO, 6s8dYU, 6s8iHL, 6s8iPO, 6s8iYU, 6s8jHL, 6s8jPO, 6s8jYU, 6sf6AB, 6sf6HL, 6shgHL, 6sncBA, 6sncHL, 6sndBA, 6sndHL, 6sneBA, 6sneDC, 6sneFE, 6sneHL, 6snhHL, 6sniHL, 6srvHL, 6srvIM, 6srxHL, 6srxIM, 6ss0HL, 6ss0IM, 6ss2HL, 6ss4HL, 6ss5HL, 6ss6HL, 6ss6IM, 6ss6JN, 6suzHL, 6sv2HL, 6svlAB, 6svlDE, 6svlGI, 6svlHL, 6svlKM, 6svlOP, 6sxiHL, 6t3jAB, 6t3jFG, 6t9dAB, 6t9dHL, 6t9eAB, 6t9eHL, 6tcmHL, 6tcnHL, 6tcoHL, 6tcpDC, 6tkbHL, 6tkoHL, 6tulHL, 6txzHL, 6txzIM, 6txzJN, 6txzKO, 6tybHL, 6tysCD, 6tysFG, 6tysHL, 6u02DC, 6u02GF, 6u02HL, 6u02KI, 6u1nHL, 6u1tHL, 6u2fHL, 6u36HL, 6u38HL, 6u3iHL, 6u3zAB, 6u3zCD, 6u6mHL, 6u6oHL, 6u6uHL, 6u85HL, 6u8dHL, 6u8kDE, 6u8kFG, 6u8kHL, 6u9sAB, 6u9sDE, 6u9uHL, 6ubiAB, 6ubiDE, 6uceHL, 6ucfHL, 6udaCD, 6udaGH, 6udaKL, 6udjDE, 6udjHI, 6udjNO, 6udkDE, 6udkHL, 6udkMN, 6ug7HL, 6ug8HL, 6ug9AB, 6ug9HL, 6ug9JK, 6ugaBA, 6ugaHL, 6ugsAB, 6ugsHL, 6ugtAB, 6ugtHL, 6uguAB, 6uguHL, 6ugvAB, 6ugvHL, 6uigBC, 6uigHL, 6uj9HL, 6ujbEF, 6ujcEF, 6ulcHL, 6uleAB, 6uleCD, 6ulfAB, 6um5CD, 6um5GH, 6um5KL, 6um6CD, 6um6GH, 6um6KL, 6um7BC, 6um7FG, 6um7JK, 6umiHL, 6umjHL, 6uoeHL, 6ur5AB, 6ur5GH, 6urmDE, 6urmHL, 6usfKM, 6utaAB, 6utaHL, 6uteAB, 6uteCD, 6uteEF, 6uteGH, 6uteIJ, 6utkDE, 6utkHL, 6uudHL, 6uuhAB, 6uuhCD, 6uulAB, 6uvoHL, 6uydHL, 6uyeGJ, 6uyeHK, 6uyeIL, 6uymAB, 6v4nAB, 6v4nCG, 6v4nDE, 6v4nHL, 6v4oAB, 6v4oCG, 6v4oDE, 6v4oHL, 6v4pCD, 6v4qHL, 6v4rHL, 6v6wAB, 6v6wDE, 6v8xCD, 6v8xGH, 6v8xKL, 6v8zCF, 6v8zIL, 6v8zOR, 6vboHL, 6vbpAB, 6vbpCD, 6vbpHL, 6vbpJI, 6vbqAB, 6vbqCD, 6vbqEF, 6vbqHL, 6vc9HL, 6vcaEF, 6vcaHL, 6vcaKJ, 6vcaPO, 6velHL, 6vepCD, 6vepIJ, 6vepOP, 6vepUV, 6veqCD, 6veqIJ, 6vgrCD, 6vgrHL, 6vi1BA, 6vi1ED, 6vi1FC, 6vi1HG, 6vi1JI, 6vi1LK, 6vi2BA, 6vi2DC, 6vjaHL, 6vjaIM, 6vjnHL, 6vjtHL, 6vknGJ, 6vknHL, 6vknIK, 6vl8AB, 6vl8CD, 6vl8EF, 6vl8HL, 6vl9AB, 6vl9HL, 6vlnHL, 6vlrFN, 6vlwHL, 6vmjBA, 6vmjFE, 6vmjJI, 6vmjML, 6vmkBA,

6vmkED, 6vmkHG, 6vmkKJ, 6vmkML, 6vmkPO, 6vmkSR, 6vmkVU, 6vmkZY, 6vmkcb, 6vmkfe, 6vmkih, 6vmklk, 6vmkon, 6vmkrq, 6vmkut, 6vn0GJ, 6vn0HL, 6vn0IK, 6vn1HL, 6vn1IM, 6vn1JN, 6vo1GJ, 6vo1HL, 6vo1IK, 6vo3GJ, 6vo3HL, 6vo3IK, 6vorCD, 6vosHL, 6vpxHL, 6vpxKN, 6vpxPQ, 6vpyAB, 6vpyHL, 6vrhBC, 6vrkBC, 6vrlBC, 6vrqAB, 6vrqHL, 6vryHL, 6vtuHL, 6vtwHL, 6vu2HL, 6vugDC, 6vvuEF, 6vvuGI, 6vvuHL, 6vvuJK, 6vx4HK, 6vy2HL, 6vy2IM, 6vy2JN, 6vy4CD, 6vy4HL, 6vy5HL, 6vy6HL, 6vyhDC, 6w00HL, 6w05HL, 6w16HL, 6w2bBC, 6w2cBC, 6w41HL, 6w4sHL, 6w4vCD, 6w51MN, 6w51OP, 6w51QR, 6w52HL, 6w5dAB, 6w5dHL, 6w73HL, 6w7sHL, 6wasAB, 6wasHL, 6wbvHL, 6weqEF, 6weqGH, 6weqIJ, 6weqKL, 6werCD, 6werEF, 6wexHL, 6wezHL, 6wf0HL, 6wf1HL, 6wfyHL, 6wfzAB, 6wfzHL, 6wg0HL, 6wg1AB, 6wg1HL, 6wg2HL, 6wg2IM, 6wg8AB, 6wgjEF, 6wgkAB, 6wgkCD, 6wgkEF, 6wglAB, 6wh9BC, 6wh9EF, 6wh9HI, 6whkCB, 6wikDC, 6wioAB, 6wirAB, 6witHL, 6wiyHL, 6wizHL, 6wj0HL, 6wj1GI, 6wj1HL, 6wj1JK, 6wjlFJ, 6wjlHL, 6wk4AB, 6wklAB, 6wkmHL, 6wm9BC, 6wm9EF, 6wmwHL, 6wmwMN, 6wn1BC, 6wn1EF, 6wn4AB, 6wn4HL, 6wnoBC, 6wnoEF, 6wo3HL, 6wo4HL, 6wo5GI, 6wo5HL, 6wozBC, 6wozEF, 6wozHI, 6wozKL, 6wqoEF, 6wrpHL, 6ws6AB, 6ws6CD, 6ws6EF, 6wt3HL, 6wtuBC, 6wtuEF, 6wtuHI, 6wtuKL, 6wtvBC, 6wtvHI, 6wtvKL, 6wtyHI, 6wvzHL, 6ww2HL, 6wwcAB, 6wwcDH, 6wx2AB, 6wx2HL, 6wxlFE, 6wxlHL, 6wxlKJ, 6wyrIM, 6wytHL, 6wzjAB, 6wzjCD, 6wzkAB, 6wzkCD, 6wzlAB, 6wzlCD, 6wzmAB, 6wzmCD, 6wztED, 6wztGF, 6wztHL, 6x1tAB, 6x1tCD, 6x1tEF, 6x1tGI, 6x1tHL, 6x1uAB, 6x1uHL, 6x1vHL, 6x3sJI, 6x3sKL, 6x3tJI, 6x3tKL, 6x3uJI, 6x3uKL, 6x3vJI, 6x3vKL, 6x3wJI, 6x3wKL, 6x3xJI, 6x3xKL, 6x3zJI, 6x3zKL, 6x40JI, 6x40KL, 6x4tBE, 6x4tDF, 6x58AB, 6x58CD, 6x5eBA, 6x5eHL, 6x78BA, 6x78HL, 6x7wAC, 6x7wHL, 6x87AB, 6x87CD, 6x87EF, 6x87GH, 6x87IJ, 6x87KL, 6x87MN, 6x8pHL, 6x8qHL, 6x8sHL, 6x8uHL, 6x97EF, 6x97HL, 6x97JK, 6x98FE, 6x98HL, 6x98KJ, 6x9rHL, 6x9sHL, 6x9tHL, 6x9uHL, 6x9vHL, 6xc2HL, 6xc2XY, 6xc3BA, 6xc3HL, 6xc4HL, 6xc4XY, 6xc7CD, 6xc7HL, 6xcaHL, 6xcjHL, 6xcmHL, 6xcmNS, 6xcnBD, 6xcnFG, 6xcnHL, 6xdgBD, 6xdgCA, 6xe1HL, 6xeyFG, 6xeyHL, 6xeyJK, 6xgcHL, 6xgcIJ, 6xgcMN, 6xjaHL, 6xjqCD, 6xjqHL, 6xjwCD, 6xjwHL, 6xjyCD, 6xjyHL, 6xjzCD, 6xjzHL, 6xkpHL, 6xkpMN, 6xkqHL, 6xkrHL, 6xliCD,

6xliHL, 6xlqBC, 6xlqEF, 6xlqHI, 6xlqKL, 6xlzHL, 6xm2BA, 6xm2DC, 6xm2FE, 6xm2HG, 6xmiBA, 6xmiED, 6xobMN, 6xobOP, 6xobQR, 6xobST, 6xocAB, 6xocHL, 6xp6HG, 6xp6JI, 6xpqBC, 6xprBC, 6xprEF, 6xpxBC, 6xpyBC, 6xpzBC, 6xpzEF, 6xq0BC, 6xq0EF, 6xq2CB, 6xq2FE, 6xq4CB, 6xq4EF, 6xqwHL, 6xr0HL, 6xrjAB, 6xrjCD, 6xrjEF, 6xrjHL, 6xskGJ, 6xskHL, 6xskIK, 6xsnHL, 6xswAB, 6xswDE, 6xswGI, 6xswHL, 6xtgHL, 6xudHL, 6xukHL, 6xulAB, 6xulCD, 6xulEF, 6xulGI, 6xulHL, 6xulJK, 6xunAB, 6xunCD, 6xunHL, 6xxvAB, 6xxvDE, 6xy2HL, 6xzwHL, 6y1lHL, 6y54HL, 6y54MN, 6y90CD, 6y90HL, 6y92CD, 6y92HL, 6y97HL, 6y9aHL, 6y9bHL, 6y9bIM, 6y9bJN, 6yaxHL, 6yaxIM, 6ye3AB, 6ye3DE, 6ye3GH, 6yioHL, 6ylaBC, 6ylaHL, 6ym0HL, 6ywcAB, 6ywcDE, 6ywdAB, 6yx9HL, 6yxdHL, 6yxkAB, 6yxmHL, 6z06HL, 6z2lCB, 6z2mBC, 6z2mHL, 6z3kAB, 6z3kCD, 6z3kEF, 6z3kGI, 6z3kHL, 6z3kJK, 6z3kMN, 6z3kOP, 6z3kQR, 6z3pHL, 6z3qHL, 6z7wBA, 6z7wDC, 6z7wFE, 6z7wHG, 6z7xBA, 6z7yBA, 6z7yDC, 6z7zBA, 6z7zDC, 6zczHL, 6zdgBC, 6zdgFG, 6zdgHL, 6zdhDF, 6zdhGI, 6zdhHL, 6zecHL, 6zerBC, 6zerFG, 6zerHL, 6zfoBC, 6zfoHL, 6zh9HL, 6zlrBC, 6zlrFG, 6zlrHL, 6ztdHL, 6ztfAB, 6ztfHL, 6ztrAB, 6ztrHL, 6zvfHL, 7a0wCD, 7a0wEF, 7a0xCD, 7a0xEF, 7a0yCD, 7a0yEF, 7a3nHL, 7a3oHL, 7a3pHL, 7a3pIM, 7a3qHL, 7a3qIM, 7a3tHL, 7a3uHL, 7a5rHL, 7a5rIM, 7a5sHL, 7a5sIM, 7a65CB, 7a69CB, 7a6cCB, 7a6eCB, 7a6fCB, 7ahuBA, 7ahuHL, 7ahvAB, 7aj6HL, 7akdDE, 7akdHL, 7akjED, 7akjGF, 7akjHL, 7ampHL, 7amqHL, 7amrHL, 7amsHL, 7arnHL, 7b09HL, 7b0bHL, 7b3oHL, 7bbgHL, 7bbjHL, 7bbjMN, 7behHL, 7beiHL, 7bejHL, 7bekHL, 7belAB, 7belCD, 7belEF, 7belHL, 7bemHL, 7benAB, 7benDF, 7benGI, 7benHL, 7beoAB, 7beoCD, 7beoEF, 7beoHL, 7bepAB, 7bepDF, 7bepGI, 7bepHL, 7bg1HL, 7bh8EF, 7bh8GH, 7bm5BA, 7bm5HL, 7bm5KJ, 7bnvHL, 7bpkCD, 7bpkHL, 7bq5CD, 7bq5HL, 7bscHL, 7bsdAB, 7bsdHL, 7bwjHL, 7bxaBC, 7bxaHL, 7bxvHL, 7bz5HL, 7c01CD, 7c01HL, 7c2lHL, 7c2lIM, 7c2lJN, 7c4sHL, 7c4sJK, 7c61HL, 7c6aHL, 7c81HL, 7c88AB, 7c88HL, 7c94BA, 7c94ED, 7c95BA, 7c95DC, 7cacED, 7cahED, 7caiED, 7caiGF, 7cakED, 7cakGF, 7cakIH, 7cbpDE, 7cbpRS, 7cdiHL, 7cdjHL, 7ce2ZB, 7ceaAB, 7cecFG, 7cgwAB, 7cgwHL, 7ch4HL, 7ch5HL, 7chbHL, 7chcAB, 7chcHL, 7cheAB, 7cheHL, 7chfAB, 7chfHL, 7chhDE, 7chhGH, 7chhJK, 7choHL, 7chpHL, 7chsHL, 7chyHL, 7chzHL, 7cj2CD,

7cj2KL, 7cjfAB, 7cm4HL, 7cmkHL, 7cqcHL, 7cqdHL, 7cqdIM, 7cr5HL, 7cu7HL, 7cvsCD, 7cvsEF, 7cvtCD, 7cvtEF, 7cwlGJ, 7cwlHL, 7cwlIK, 7cwmGJ, 7cwmHL, 7cwmIK, 7cwnED, 7cwnGF, 7cwnHN, 7cwnKI, 7cwnLJ, 7cwnOM, 7cwoHL, 7cwsED, 7cwsGF, 7cwsIH, 7cwsLC, 7cwsNK, 7cwsPM, 7cwtED, 7cwtGF, 7cwtIJ, 7cwtLH, 7cwtNK, 7cwtPM, 7cwuHL, 7cwuJF, 7cwuMR, 7cwuNG, 7cwuOS, 7cwuPI, 7cyhED, 7cypED, 7cypGF, 7cypIH, 7czqHK, 7czqJN, 7czrHK, 7czrJN, 7czsHK, 7czsIM, 7czsJN, 7cztHL, 7cztIM, 7czuHK, 7czuJN, 7czvHK, 7czvIM, 7czvJN, 7czxHK, 7czxIM, 7czxJN, 7czyHK, 7czyJN, 7czzHK, 7czzIM, 7czzJN, 7d00HK, 7d00JN, 7d0bHL, 7d0cGF, 7d0cHL, 7d0dGF, 7d0dHL, 7d4gNG, 7d5zDC, 7d5zHG, 7d5zLK, 7d5zPO, 7d6iBC, 7d85BC, 7d85EF, 7d9zHL, 7daaHL, 7dc7AB, 7dc8BA, 7dc8ED, 7dccAF, 7dccBM, 7dccDH, 7dceHL, 7dcxAF, 7dcxBG, 7dcxEH, 7dd2AB, 7dd2EH, 7dd8AB, 7df1EI, 7df1FJ, 7df1GK, 7df1HL, 7df9HL, 7dfaHL, 7dfbHL, 7dfcHL, 7dfpCB, 7dhaCB, 7djzAB, 7dk0AB, 7dk2AB, 7dk2DE, 7dk2GH, 7dk2JK, 7dk4ab, 7dk4cd, 7dk4ef, 7dk5ab, 7dk6ab, 7dk6cd, 7dk7ab, 7dk7cd, 7dkjCD, 7dkjGH, 7dkjKL, 7dm1DC, 7dm1FE, 7dm2HL, 7dnhHL, 7dnkHL, 7dnlHL, 7dohHL, 7dpmAB, 7dpmDE, 7dpmGH, 7dpmJK, 7dq7HL, 7dr4AB, 7dr4CD, 7dr4EF, 7dr4HL, 7dunHL, 7duoHL, 7dx4HL, 7e3bBC, 7e3kDE, 7e3kFG, 7e3kHL, 7e3lDE, 7e3lFG, 7e3lHL, 7e3oHL, 7e5oHL, 7e5rIK, 7e5rLH, 7e5rMU, 7e5rND, 7e5rOE, 7e5rPQ, 7e5rTV, 7e5rWR, 7e5rXS, 7e5sEH, 7e5sIK, 7e5sJT, 7e5sLD, 7e5sPN, 7e5sRO, 7e5sUQ, 7e5sVS, 7e5yCB, 7e5yHL, 7e6pHL, 7e72AB, 7e72CD, 7e86AB, 7e88AB, 7e88DE, 7e88GH, 7e88JK, 7e8cDE, 7e8cHL, 7e8cKI, 7e8cMJ, 7e8cNP, 7e8cOQ, 7e8cSU, 7e8cTV, 7e8caZ, 7e8fCB, 7e8fDE, 7e8fHL, 7e8mHL, 7e9bHL, 7eamCD, 7eamHL, 7eanHL, 7ebrEF, 7ebzEF, 7ec5EF, 7ecyEF, 7eh5DO, 7eh5EP, 7eh5FQ, 7eh5HL, 7eh5IM, 7eh5JN, 7ej4HL, 7ej4IM, 7ej4JN, 7ej5HL, 7ej5IM, 7ej5JN, 7ejyHL, 7ejzHL, 7ek0HL, 7ekkHL, 7enfPO, 7engHL, 7eo0HL, 7epxHL, 7epxIM, 7ew5BC, 7ew5EF, 7ew5HI, 7ew5KL, 7ew5NO, 7ew5QR, 7ew5TU, 7ew5WX, 7ew5Za, 7ew5gj, 7ey4HL, 7ey5BC, 7ey5HL, 7eycAB, 7eycHL, 7ezvHL, 7ezvMN, 7f3qCB, 7f3qHL, 7f62HL, 7f63HL, 7f6yHL, 7f6zHL, 7f7eCL, 7f7hAB, 7f7hCD, 7f9wBC, 7faeHL, 7faeed, 7fafHL, 7fafdc, 7fafgf, 7fahCD, 7fahHL, 7fbiNM, 7fciHL, 7fcpBD, 7fcpHL, 7fcqHL, 7fgjHL, 7fgkHL, 7fglHL, 7fgrHL, 7fjcHL, 7fjsCA, 7fjsHL, 7jgjHL, 7jhgHL, 7jhhHL, 7jixHL,

7jkmHL, 7jksHL, 7jktHL, 7jlkAB, 7jlkHL, 7jlnAB, 7jlnHL, 7jmoHL, 7jmpHL, 7jmwHL, 7jn5HL, 7jooHL, 7jr7CD, 7jr7EF, 7jtgAB, 7jtiHL, 7jv2HL, 7jv4DE, 7jv4FG, 7jv4HL, 7jv6CD, 7jv6FG, 7jv6HL, 7jvaHL, 7jvcCD, 7jvcFG, 7jvcHL, 7jvdHL, 7jw0DC, 7jw0GF, 7jw0HL, 7jwpIJ, 7jwpKL, 7jwpMN, 7jx3AB, 7jx3CD, 7jx3HL, 7jxdAB, 7jxdCD, 7k3qHL, 7k66BC, 7k75AB, 7k75CD, 7k75IJ, 7k75MN, 7k76AB, 7k76HL, 7k7hHL, 7k7rBA, 7k7rED, 7k8mAB, 7k8nAB, 7k8oHL, 7k8pHL, 7k8qHL, 7k8rHL, 7k8sHL, 7k8sMN, 7k8sOP, 7k8tHL, 7k8tMN, 7k8tOP, 7k8vHL, 7k8vMN, 7k8wHL, 7k8wNS, 7k8xDF, 7k8xEG, 7k8yGI, 7k8zHL, 7k8zMN, 7k90HL, 7k90MN, 7k90OP, 7k9hHL, 7k9hIM, 7k9iHL, 7k9jHL, 7k9jIM, 7k9jJN, 7k9kHL, 7k9zBA, 7k9zHL, 7kbaAB, 7kbaHL, 7kbbAB, 7kbbHL, 7kcrHL, 7kd6CD, 7kd6IJ, 7kd6OP, 7kd6UV, 7kddDE, 7kddFG, 7kddHI, 7kdeOP, 7kdeQR, 7kdeST, 7ke1AB, 7ke1CD, 7kejGJ, 7kejHK, 7kejIL, 7keoAB, 7keoEF, 7kewGJ, 7kewHK, 7kewIL, 7kexGJ, 7kexHK, 7kexIL, 7kezHL, 7kf0AB, 7kf0HL, 7kf1AB, 7kf1DE, 7kf1GI, 7kf1HL, 7kf9GJ, 7kf9HK, 7kf9IL, 7kfbGJ, 7kfbHK, 7kfbIL, 7kfeGI, 7kfeHJ, 7kfgAB, 7kfhGJ, 7kfhHK, 7kfhIL, 7kfvCD, 7kfvFG, 7kfvHL, 7kfwCD, 7kfwFG, 7kfwHL, 7kfxHL, 7kfyHL, 7kguBA, 7kguDC, 7kguFE, 7kguHL, 7kgvCE, 7kgvDF, 7khfAL, 7khfHB, 7ki4DC, 7ki4GF, 7ki4HL, 7ki6DC, 7ki6GF, 7ki6HL, 7kkhBA, 7kkzHL, 7klcDC, 7klcHL, 7klgHL, 7klgIM, 7klhHL, 7klhIM, 7kmdAE, 7kmgAB, 7kmgDE, 7kmhAB, 7kmiAB, 7kmkHL, 7kmkIM, 7kmlHL, 7kmlIM, 7kmlJN, 7kn3HL, 7kn3MN, 7kn6HL, 7kn7HL, 7kpbHL, 7kpgHL, 7kpjAB, 7kpjCD, 7kq7HL, 7kqbHL, 7kqbKJ, 7kqeHL, 7kqeIN, 7kqeJM, 7kqgBC, 7kqhHL, 7kqiAC, 7kqkAB, 7kqkHL, 7kqlHL, 7ks9HL, 7kx4BC, 7kx4DA, 7kxjHL, 7kxjIM, 7kxjJN, 7kxkHL, 7kxkIM, 7kxkJN, 7kylHL, 7kylXY, 7kyoHL, 7kzbHL, 7kzxDC, 7kzxFE, 7kzzDC, 7kzzFE, 7l02HL, 7l02MK, 7l06DF, 7l06GE, 7l06HL, 7l06MK, 7l09HL, 7l09MK, 7l0lGJ, 7l0lHL, 7l0lIK, 7l0nAB, 7l0nCD, 7l0nHL, 7l0nMN, 7l2cCD, 7l2cHL, 7l2dHL, 7l2eGJ, 7l2eHL, 7l2eIK, 7l3nDE, 7l56FG, 7l56HL, 7l56JK, 7l57HL, 7l5bHL, 7l5jWL, 7l5jYX, 7l6mhg, 7l6mij, 7l77HL, 7l79AB, 7l79HL, 7l7dHL, 7l7eAB, 7l7eCD, 7l7eEF, 7l7eHL, 7l7eIJ, 7l7eMN, 7l7eOP, 7l7eTU, 7l7rBA, 7l7rDC, 7l8pEF, 7l8pHL, 7la4HL, 7laaHL, 7lbeFE, 7lbeHG, 7lbfFE, 7lbfHG, 7lbgFE, 7lbgHG, 7lcnBD, 7lcnHL, 7lcnMS, 7lcvAB, 7ld1HL, 7ld1MN, 7ld1PO, 7lddKJ, 7lddNM, 7ldeKJ, 7ldeNM, 7lezDE, 7lezGI,

7lezQO, 7lezRP, 7lf0DE, 7lf0FH, 7lf0GI, 7lf0KL, 7lf0QO, 7lf0RP, 7lf7AB, 7lfaBD, 7lfaHL, 7lfbHL, 7lfdHL, 7lg6CD, 7lg6JM, 7lg6KN, 7lj4ED, 7lj4GF, 7lj5ED, 7lj5GF, 7ljaED, 7ljaGF, 7ljbED, 7ljbGF, 7lk4BA, 7lk4DC, 7lk4FE, 7lk4HG, 7lk9AB, 7lk9CD, 7lk9HL, 7lkaAB, 7lkaCD, 7lkaEF, 7lkaHL, 7lkbBC, 7lkbHL, 7lkfHL, 7lkgAB, 7lkgDE, 7lkhHL, 7lkiAB, 7lkiDE, 7ll1EF, 7ll1HL, 7ll1JK, 7ll2HL, 7ll2IJ, 7ll2KM, 7lm8HL, 7lm8MN, 7lm9HL, 7lo6HG, 7lo6JI, 7lo6LK, 7lo7HL, 7lo8HL, 7lokKJ, 7lokML, 7lopBC, 7lopHL, 7lopWV, 7lopXY, 7lq7CD, 7lq7FG, 7lq7HL, 7lq7PQ, 7lq7TU, 7lq7yz, 7lqvHL, 7lqvYZ, 7lr3AB, 7lr3HL, 7lr4AB, 7lr4HL, 7lrsDE, 7lrtDE, 7lrtFG, 7lrtHL, 7ls9DF, 7ls9EG, 7ls9HL, 7lsbBD, 7lseHL, 7lsfHL, 7lsgHL, 7lsiAC, 7lsiBD, 7lssHL, 7luahg, 7luaij, 7lucDE, 7lucFG, 7lucHI, 7lucJK, 7lucLM, 7lucNO, 7ludAB, 7ludHL, 7ly0HL, 7ly2GI, 7ly2HL, 7ly2NO, 7ly3DC, 7ly3FE, 7ly9BC, 7ly9EI, 7lywAB, 7lywCD, 7lywHL, 7m2hAB, 7m2hDE, 7m2iAB, 7m2jAB, 7m30FG, 7m3iAB, 7m3iHL, 7m42BA, 7m42DC, 7m51HL, 7m52HL, 7m53HL, 7m55HL, 7m6dAB, 7m6dHL, 7m6eDC, 7m6eFG, 7m6eHL, 7m6gHL, 7m6gRS, 7m6hDF, 7m6hEG, 7m71HL, 7m74HL, 7m7bHL, 7m7eCD, 7m7fCD, 7m7fEF, 7m7gCD, 7m7gEF, 7m7hCD, 7m7hEF, 7m7iCD, 7m7iEF, 7m7jCD, 7m7jEF, 7m7wAB, 7m7wCD, 7m7wEF, 7m7wHL, 7m8jHL, 7m8lGJ, 7m8lHK, 7m8lIL, 7m8lMP, 7m8lNQ, 7m8lOR, 7mdjAB, 7mdpHI, 7mdtHL, 7mduEF, 7mduHL, 7memGJ, 7memHL, 7memIK, 7mepHL, 7mepIG, 7mepMJ, 7mepNK, 7mf1HL, 7mf7AB, 7mf7CD, 7mf7EF, 7mf7HL, 7mf8HL, 7mf9CD, 7mf9GI, 7mf9HL, 7mf9MN, 7mf9QR, 7mfaCD, 7mfaGI, 7mfaHL, 7mfaMN, 7mfaQR, 7mfaUV, 7mfbHL, 7mfgEF, 7mfgHL, 7mfgJK, 7mhxAB, 7mjjDE, 7mjjFG, 7mjkDE, 7mjkFG, 7mjkHI, 7mjlHL, 7mjsHL, 7mlhCA, 7mlhED, 7mlzHL, 7mm0HL, 7mmnDE, 7mmnHL, 7mmnMG, 7mmoAB, 7mmoDE, 7mn8DC, 7mnlHL, 7mnlKM, 7mnmHL, 7mnmKM, 7mnnHL, 7mnnKM, 7mnoHL, 7mnoKM, 7mrzXY, 7msqDE, 7msqHL, 7mtaHL, 7mtaHL, 7mtaHL, 7mtaHL, 7mtaHL, 7mtaHL, 7mtbHL, 7mtbHL, 7mtbHL, 7mtbHL, 7mtbHL, 7mtbHL, 7mu4HL, 7mw2DE, 7mw2FG, 7mw2HL, 7mw3DE, 7mw3FG, 7mw3HL, 7mw4DE, 7mw4FG, 7mw4HL, 7mw5DE, 7mw5FG, 7mw5HL, 7mw6DE, 7mw6FG, 7mw6HL, 7mwwHL, 7mwxAB, 7mwxEF, 7mxdCD, 7mxdJK, 7mxdSU, 7mxdXY, 7mxeAB, 7mxeHL, 7mxeMO, 7mxeRW, 7mxlAB, 7mxlHL, 7mxlJK, 7mxpGJ, 7mxpHL, 7mxpIK, 7mytAB, 7mytHL,

7myzHL, 7myzIM, 7mzfHL, 7mzgHL, 7mzhBC, 7mzhHL, 7mziHL, 7mzjCD, 7mzjFE, 7mzjHL, 7mzjNM, 7mzkCD, 7mzkFE, 7mzkHL, 7mzkNM, 7mzlHL, 7mznHL, 7n01HL, 7n04HL, 7n05HL, 7n07HL, 7n08AB, 7n08CD, 7n08HL, 7n0aBA, 7n0uHL, 7n0vHL, 7n0xHL, 7n28CD, 7n28JK, 7n28SU, 7n28XY, 7n3cHL, 7n3dHL, 7n3eHL, 7n3fHL, 7n3gHL, 7n3hHL, 7n3iHL, 7n4iHL, 7n4jHL, 7n4mHL, 7n5hEF, 7n5hHL, 7n5hJK, 7n62HL, 7n64HL, 7n6pAB, 7n6pHL, 7n8hCB, 7n8hJI, 7n8hON, 7n8iHL, 7n8qCD, 7n8qHL, 7nabAB, 7nabHL, 7nc0HL, 7ncsHL, 7nd0HL, 7nd4DE, 7nd4FG, 7nd4HL, 7nd5HL, 7nd6HL, 7nd7FG, 7nd7HL, 7nd7JK, 7nd8HL, 7nd9HL, 7ndaHL, 7ndbHL, 7negHL, 7nehHL, 7neqDC, 7neqFE, 7nezDC, 7nezFE, 7nfaCA, 7nfaDB, 7niuCB, 7nivCB, 7niwCB, 7njzHL, 7np1HL, 7np1IM, 7np1JN, 7np1KO, 7ntcHL, 7nw3HL, 7nwlDE, 7nx2BA, 7nx2DC, 7nx3BC, 7nx3ED, 7nx6AB, 7nx6HL, 7nx7AB, 7nx7HL, 7nx8AB, 7nx8HL, 7nx9AB, 7nx9HL, 7nxaAB, 7nxaHL, 7nxbAB, 7nxbHL, 7ny5CD, 7ny5FB, 7o2zHL, 7o30BA, 7o30HL, 7o31HL, 7o33HL, 7o4yHL, 7o52HL, 7o7fHF, 7o9wCB, 7oh0CB, 7oh1ED, 7olyHL, 7oo2AB, 7oo2CD, 7or9AB, 7or9HL, 7oraAB, 7oraDE, 7oraFG, 7oraHL, 7orbAB, 7orbCD, 7orbEF, 7orbHL, 7ow1HK, 7ox1AB, 7ox1CD, 7ox1EF, 7ox1HL, 7ox2AL, 7ox2BC, 7ox2DE, 7ox2JK, 7oxnHK, 7p40CD, 7p40FB, 7p40IH, 7pa8JI, 7pa8LK, 7pa8NM, 7pa8YF, 7pa9KL, 7pa9MN, 7pa9OP, 7pa9QR, 7pa9TS, 7pa9VW, 7pa9XU, 7pa9YZ, 7pa9ba, 7pa9cd, 7pc2MN, 7pc2OP, 7pc2QR, 7pgbHL, 7pgbRS, 7pgbUV, 7pgbXY, 7pgbab, 7pgbfg, 7pgbik, 7pgbmn, 7pgbpq, 7phgHL, 7phsHL, 7phuBC, 7phuDE, 7phvBC, 7phvEF, 7phwBC, 7pi2BC, 7pi2EF, 7pi2HI, 7pi2KL, 7pi3BC, 7pi3FG, 7pi3IJ, 7pi3MN, 7pi3PQ, 7pi3ab, 7pi7BC, 7pi7EF, 7pnmGD, 7pnmHL, 7pnmIE, 7pnqDF, 7pnqHL, 7pnqXY, 7po5HL, 7po5XF, 7po5ZD, 7pqyBC, 7pqyHL, 7pqzAB, 7pqzHL, 7pr0CB, 7pr0GF, 7pr0HL, 7pryAB, 7pryCD, 7pryFG, 7pryHL, 7przHL, 7ps0BC, 7ps0HL, 7ps1AB, 7ps2AB, 7ps2HL, 7ps3HL, 7ps4BC, 7ps4HL, 7ps5HL, 7ps6CD, 7ps6HL, 7ps7AB, 7ps7HL, 7q0aYZ, 7q0gAB, 7q0gHL, 7q0hAB, 7q0hHL, 7q0iAB, 7q0iHL, 7q4qBA, 7q4qDC, 7q6cHL, 7q6eEF, 7q6eHL, 7q6eJK, 7q9fEF, 7q9fHL, 7q9fJK, 7q9gEF, 7q9gHL, 7q9iEF, 7q9iHL, 7q9iJK, 7q9jEF, 7q9jHL, 7q9jJK, 7q9kDE, 7q9kHL, 7q9mEF, 7q9mHL, 7q9mJK, 7q9pEF, 7q9pHL, 7q9pJK, 7qezEF, 7qezHL, 7qf0HL, 7qf1CB, 7qf1ED, 7qnwAB, 7qnwHL, 7qnxAB, 7qnxHL, 7qnyAB, 7qnyHL, 7qtiBC,

7qtiFE, 7qtiGH, 7qtiIJ, 7qtjBC, 7qtjFE, 7qtkBC, 7qu1AB, 7qu2AB, 7r40DF, 7r40EG, 7r40HL, 7r58HL, 7r6wAB, 7r6wHL, 7r6xAB, 7r6xHL, 7r8lCD, 7r8lHL, 7r8mDE, 7r8mHL, 7r8nHL, 7r8nMN, 7r8nOP, 7r8uHL, 7r9dHL, 7rahBA, 7rahDC, 7raiHL, 7raiIM, 7raiJN, 7rajHL, 7raqHL, 7rbuHL, 7rcoDC, 7rcoFE, 7rcsAB, 7rcsHL, 7rd3AB, 7rd3HL, 7rd4AB, 7rd4HL, 7rd5BA, 7rd5DC, 7rd9HL, 7rdaHL, 7rdjBA, 7rdkBA, 7rdlBA, 7rdmBA, 7rdwGI, 7rdwHL, 7rdwQR, 7rdwUV, 7re7HL, 7re7IJ, 7re9AB, 7re9HL, 7rewAB, 7rewHL, 7rfbAB, 7rfbHL, 7rfcAB, 7rfcHL, 7rfpHL, 7rfpIM, 7rksHL, 7rksIM, 7rkuEF, 7rkuGI, 7rkuHL, 7rkuJK, 7rkvDH, 7rkvEF, 7rkvGL, 7rlvAB, 7rlvCD, 7rlvEF, 7rlwAB, 7rlwCD, 7rlxAB, 7rlyAB, 7rlyCD, 7rlyEF, 7rlzAB, 7rlzDC, 7rm0AB, 7rm0CD, 7rm0EF, 7rm0GH, 7rm1AB, 7rm1CD, 7rm3AB, 7rm3CD, 7rnjHL, 7rp0AB, 7rp3HI, 7rptAB, 7rptCD, 7rptEF, 7rptGH, 7rpuAB, 7rq6DE, 7rq6HL, 7rq6JK, 7rqpAB, 7rqqHL, 7rqrAB, 7rqrHL, 7rr0BC, 7rsnEJ, 7rsnFK, 7rsnHL, 7rsoGJ, 7rsoHL, 7rsoIK, 7rt9AB, 7rt9CD, 7rtaHL, 7ru3EF, 7ru3HL, 7ru4HL, 7ru5DE, 7ru5HL, 7ru8HL, 7rw2GJ, 7rw2HL, 7rw2IK, 7rxcHL, 7rxdHL, 7rxiHL, 7rxjAB, 7rxjHL, 7rxlCD, 7rxlHL, 7rxpHL, 7ryuHL, 7ryvEF, 7ryvHL, 7ryvIJ, 7ryvPQ, 7s07HL, 7s07XY, 7s08AB, 7s08HL, 7s0bAB, 7s0bCD, 7s0cHL, 7s0cMN, 7s0dHL, 7s0dMN, 7s0dOP, 7s0eHL, 7s0jHL, 7s0xCD, 7s0xHL, 7s11HL, 7s11IM, 7s11JN, 7s11KO, 7s13HL, 7s13IM, 7s1bHL, 7s3mHL, 7s4gAB, 7s4gCD, 7s4gEF, 7s4gHL, 7s4sHL, 7s5pHL, 7s5qBC, 7s5qHL, 7s5rBC, 7s5rHL, 7s6cEG, 7s6cHF, 7s6dDF, 7s6dEG, 7s6jCD, 7s6jFG, 7s6jHL, 7s6kCD, 7s6kFG, 7s6kHL, 7s6lHL, 7s7iHL, 7s7iIM, 7s7iJN, 7s7iUV, 7s7iWX, 7s7iYZ, 7s8ghl, 7s8hBC, 7s8hHL, 7sa6HL, 7sbuHL, 7sbzAB, 7sbzHL, 7sc1HL, 7sc1MN, 7sc1PQ, 7sd2AB, 7sd2CD, 7sd2HL, 7sd5HL, 7semBC, 7sfxAB, 7sfxCD, 7sfxGH, 7sg4HL, 7sg5HL, 7sg6HL, 7sgfHL, 7sgfJK, 7sgfMN, 7shtHL, 7shtJK, 7shuCD, 7shuEF, 7shyCD, 7shyEF, 7shyIJ, 7shyKL, 7shzCD, 7shzEF, 7shzIJ, 7shzKL, 7si0CD, 7si0EF, 7si0IJ, 7si0KL, 7si2ED, 7si2GF, 7si2HL, 7sixHL, 7sj0BL, 7sjmHL, 7sjnDE, 7sjnFG, 7sjnHL, 7sjsHL, 7sk3DC, 7sk3FE, 7sk4DC, 7sk5FE, 7sk7DC, 7sk8DC, 7sk8FE, 7sk9FE, 7skzHL, 7sl5AB, 7sl5DE, 7sl5GH, 7sn1HL, 7sn2HL, 7sn3FG, 7sn3HL, 7so5HL, 7so7XY, 7so9CB, 7so9JI, 7so9ON, 7soaHL, 7sobCB, 7sobFE, 7sobML, 7socHL, 7sodHL, 7soeBL, 7soeED, 7soeHG, 7sofHL, 7sq1AI, 7sq1HL, 7sqwAB, 7srsHL,

7srsQP, 7sscHL, 7st8HL, 7strHL, 7stsAB, 7stsHL, 7stzHL, 7stzIM, 7su0HL, 7su0IM, 7su1HL, 7swdJK, 7swnHL, 7swoHL, 7swpHL, 7swwHL, 7swxHL, 7t0gHL, 7t0hHL, 7t0iDC, 7t0jDC, 7t0jFE, 7t0jHL, 7t0kBA, 7t0kDC, 7t0kFE, 7t0lGK, 7t0lHL, 7t0rCB, 7t0rHL, 7t0zHL, 7t1wHL, 7t1xHL, 7t3dCD, 7t3dJM, 7t3dKN, 7t3mDL, 7t3mHJ, 7t3mIK, 7t4gGI, 7t4gHL, 7t4gJK, 7t4qFG, 7t4qIH, 7t4qJK, 7t4sHI, 7t4sJK, 7t5oHL, 7t72HL, 7t74IJ, 7t74KG, 7t74MN, 7t75HL, 7t7bHL, 7t82DC, 7t82HL, 7t87HL, 7t8wHL, 7t97AB, 7t97CD, 7t97EF, 7t97HL, 7t98AB, 7t98HL, 7t99AB, 7t99CD, 7t99EF, 7t99HL, 7t9aJK, 7t9aMN, 7t9aOP, 7t9bJK, 7t9bMN, 7t9bOP, 7t9mHL, 7tasHL, 7tatDE, 7tatFG, 7tatHL, 7tb8DE, 7tb8FG, 7tb8HL, 7tb8JK, 7tbfDE, 7tbfHL, 7tc9HL, 7tcaDE, 7tcaHL, 7tccDE, 7tccFG, 7tccHL, 7tccJK, 7tccMN, 7tccOP, 7tcqAB, 7tcqHL, 7tctEF, 7tctHL, 7td8EF, 7td8HL, 7tdmHL, 7tdnHL, 7te1HL, 7te6CD, 7te6GH, 7teqHL, 7teqMN, 7terHL, 7terMN, 7tesHL, 7tesMN, 7tfnHL, 7tfnIJ, 7tfnPQ, 7tfoHL, 7tfoIJ, 7theBC, 7thoEF, 7thoHL, 7thtHL, 7thtac, 7thtdb, 7tjqAB, 7tjqDE, 7tjqHL, 7tjqIM, 7tjqJK, 7tjqNO, 7tl0DE, 7tl0FG, 7tl0HI, 7tl0JK, 7tl0LM, 7tl0NO, 7tm0NO, 7tmzEF, 7tmzHL, 7tn0BA, 7tn0CD, 7tn0HG, 7tn0MN, 7tn9AB, 7tn9CD, 7tn9EF, 7tn9QR, 7towAB, 7towHL, 7tp3HL, 7tp4HL, 7tpdEF, 7tpdHL, 7tqaAB, 7tqaCD, 7tqaHL, 7tqbHL, 7tr4LH, 7ttmHL, 7ttxHL, 7ttyHL, 7tufAB, 7tufHL, 7tugHL, 7tuhHL, 7tusAB, 7tusCD, 7tusEF, 7tusGH, 7tuyHL, 7txtCD, 7txzEF, 7txzHL, 7tyvDE, 7tyvGK, 7tyvHL, 7u04HL, 7u09HL, 7u0aHL, 7u0bHL, 7u0dMN, 7u0eAB, 7u0eHL, 7u0kHL, 7u0pDI, 7u0pEF, 7u0qGF, 7u0qJI, 7u0xFG, 7u0xHL, 7u2dHL, 7u2eHL, 7u5bAB, 7u5bHL, 7u5gHL, 7u60EF, 7u60HL, 7u63BA, 7u63DC, 7u63HL, 7u8cHL, 7u8gDE, 7u8mNO, 7u8mQR, 7u8mTU, 7u9fEF, 7u9fHL, 7u9gGD, 7u9gHE, 7u9gIF, 7u9pHL, 7u9vEF, 7u9vHL, 7uapHL, 7uapMN, 7uaqHL, 7uarDE, 7uarFG, 7uarHL, 7ubrEF, 7ubrHL, 7uceHL, 7ucfDE, 7ucgCF, 7ucgDE, 7ucgNO, 7ucyEF, 7ucyHL, 7udgEF, 7udgHL, 7udhEF, 7udhHL, 7udsDE, 7udsFG, 7udsHL, 7ue0EF, 7ue0HL, 7ue9HL, 7uedHL, 7uelHL, 7uemDC, 7uemHL, 7uenHL, 7ufhEF, 7ufhHL, 7ugnGJ, 7ugnHK, 7ugnIL, 7ugoGJ, 7ugoHK, 7ugoIL, 7ugpGJ, 7ugpHK, 7ugpIL, 7ugqGJ, 7ugqHK, 7ugqIL, 7uh8EF, 7uh8HL, 7uhzHL, 7uhzIM, 7uhzJN, 7ui0HL, 7ui0IM, 7ui0JN, 7ujeEF, 7ujeHL, 7ujkEF, 7ujkHL, 7uk9EF, 7uk9HL, 7ukoEF, 7ukoHL, 7ukpEF, 7ukpHL,

7uktEF, 7uktHL, 7ul0HL, 7ul1HL, 7ul7ED, 7ul7GF, 7ul7HL, 7ummHL, 7ummIM, 7ummJN, 7ums45, 7ums67, 7unbFE, 7unbHL, 7uotOP, 7uotQR, 7uotST, 7uovOP, 7uovQR, 7uovST, 7uplED, 7uplFI, 7upwDE, 7upwFG, 7upwHL, 7upxHL, 7upyDE, 7upyFG, 7upyHL, 7uraHL, 7urcHL, 7urdHL, 7ureHL, 7urqHL, 7ursHL, 7urxHL, 7uslHL, 7ut3HL, 7uvfHL, 7uvfXY, 7uvhAB, 7uvhDE, 7uviAB, 7uviDE, 7uvlHL, 7uvqCB, 7uvqDE, 7uxlAB, 7uxlFE, 7uz4HL, 7uz4MN, 7uz4PQ, 7uz5HL, 7uz6HL, 7uz6MN, 7uz6PQ, 7uz7HL, 7uz7MN, 7uz7PQ, 7uz8HL, 7uz8MN, 7uz8PQ, 7uz9HL, 7uz9MN, 7uz9PQ, 7uzaHL, 7uzbHL, 7uzdHL, 7v05Aa, 7v05Bb, 7v05Cc, 7v05Dd, 7v05Ee, 7v05Ff, 7v05Gg, 7v05Hh, 7v05Ii, 7v05Jj, 7v05Kk, 7v05Ll, 7v05Mm, 7v05Nn, 7v23OP, 7v23QR, 7v23ST, 7v24HL, 7v26ab, 7v26cd, 7v26ef, 7v27BA, 7v3hGH, 7v3hIM, 7v3hKL, 7v3lDE, 7v3lFG, 7v3lHI, 7v3qBA, 7v3qDC, 7v4wBA, 7v5jED, 7v5jGF, 7v5jIH, 7v5kED, 7v5kGF, 7v5kIH, 7v5nBA, 7v5nFE, 7v61HK, 7v61IM, 7v64BA, 7v6nED, 7v6nGF, 7v6nIH, 7v6oIH, 7v7kBA, 7v8qBA, 7v8qDC, 7v8qFE, 7vacBA, 7vacDC, 7vacFE, 7vadHL, 7vaeHL, 7vafHL, 7vagHL, 7vazBA, 7vazDC, 7vazHF, 7venBA, 7vgrDC, 7vgrFE, 7vgsDC, 7vgsFE, 7vmzHL, 7vn9AB, 7vn9HL, 7vsuBA, 7vsuDC, 7vswBA, 7vswFE, 7vswJI, 7vswNM, 7vtjHL, 7vuxHL, 7vyrAB, 7vyrHL, 7vytBC, 7vytHL, 7w71HL, 7w71IM, 7w9eDE, 7w9fAB, 7wb5HL, 7wbhOQ, 7wbhPS, 7wbhUV, 7wbzHL, 7wc0AB, 7wc0HL, 7wcdAB, 7wcdEF, 7wcdGI, 7wchDE, 7wchHL, 7wchMN, 7wckAB, 7wcpDE, 7wcpHL, 7wcpMN, 7wcrab, 7wcuAB, 7wczab, 7wd0ab, 7wd0cd, 7wd7ab, 7wd7cd, 7wd7ef, 7wd8EF, 7wd9EF, 7wd9GH, 7wd9IJ, 7wdfDE, 7wdfFG, 7we7HL, 7we7JM, 7we7KN, 7we8HL, 7we9CL, 7we9HJ, 7we9IK, 7weaGI, 7weaHL, 7webGI, 7webHL, 7wecGJ, 7wecHL, 7wecIK, 7wedHL, 7weeHL, 7wefCL, 7wg3EA, 7wg3FB, 7wg3GC, 7wg3HD, 7wgdHL, 7wgtHL, 7wh8BA, 7whbDE, 7whbGH, 7whbJK, 7whdDE, 7whdGH, 7whdJK, 7whzBD, 7whzCE, 7whzHL, 7wi0BD, 7wi0CE, 7wi0HL, 7wjyBC, 7wjyEF, 7wjzAC, 7wjzDN, 7wjzEO, 7wjzFP, 7wjzGQ, 7wjzHR, 7wk0AB, 7wk0DE, 7wk8CD, 7wk9ab, 7wk9cd, 7wkaab, 7wkacd, 7wkxAC, 7wkxBD, 7wlcHL, 7wlwHL, 7wn8HL, 7wn8PQ, 7wo4DE, 7wo4FG, 7wo4HI, 7wo4MN, 7wo4OP, 7wo4QR, 7wo5DE, 7wo5FG, 7wo5HI, 7wo7AB, 7woaDB, 7woaFG, 7wobDB, 7wobFG, 7wobHI, 7wocAB, 7wogAB, 7wonHL, 7wp0HL, 7wp1HL, 7wp2HL, 7wp5HL, 7wp6BI, 7wp6DC, 7wp6HL, 7wp8DC, 7wp8EG, 7wp8HL, 7wpdXY,

7wpeUV, 7wpeXY, 7wpfRS, 7wpfUV, 7wpfXY, 7wphED, 7wphHL, 7wpvHL, 7wr8AB, 7wr9AB, 7wroHL, 7wrvUV, 7wryHL, 7wrzHL, 7ws0ED, 7ws0GF, 7ws0IH, 7ws1ED, 7ws1GF, 7ws1IH, 7ws1KJ, 7ws2ED, 7ws3ED, 7ws3GF, 7ws3IH, 7ws4IH, 7ws5ED, 7ws5GF, 7ws5IH, 7ws6IH, 7ws7GF, 7ws7KJ, 7wsiHL, 7wslHL, 7wtiGJ, 7wtiHL, 7wtiIK, 7wtjHL, 7wtkGJ, 7wtkHL, 7wtkIK, 7wueCD, 7wueEF, 7wuhDF, 7wuhHI, 7wuhKL, 7wurEF, 7wurGH, 7wvdGH, 7wvdIJ, 7wvgCD, 7wviAB, 7wviCD, 7wvlHL, 7wvmAB, 7wvmCD, 7wwiED, 7wwiGF, 7wwiIH, 7wwjGF, 7wwjIH, 7wwkGF, 7wwkIH, 7wwlHL, 7wwlIM, 7wwlJN, 7wwmHL, 7wwmIM, 7wwmJN, 7x1mAB, 7x1mCD, 7x1mHL, 7x25AC, 7x25HK, 7x26HK, 7x28CD, 7x28HK, 7x29HK, 7x2aHK, 7x2gHL, 7x2iHL, 7x2kHL, 7x2oHL, 7x2wHL, 7x37HL, 7x38HL, 7x3cFE, 7x3cHL, 7x46HL, 7x47HL, 7x6aGJ, 7x6aHL, 7x6aIK, 7x6lGJ, 7x6lHL, 7x6lIK, 7x6oCD, 7x6oIK, 7x6oJL, 7x7oEF, 7x7oGI, 7x7oHL, 7x7oJK, 7x7tAB, 7x7tHL, 7x7uAB, 7x7uHL, 7x7vAF, 7x7vHL, 7x8tCB, 7x8tFE, 7x8tIH, 7x8tLK, 7x8wHL, 7x8yHL, 7x8zHL, 7x90EF, 7x91HL, 7x92HL, 7x93EF, 7x94HL, 7x95HL, 7x96HL, 7xckAB, 7xcoHL, 7xcoMN, 7xcoOP, 7xcpHL, 7xczBC, 7xczDE, 7xdaBC, 7xdaDE, 7xdbDE, 7xdbGF, 7xdbIH, 7xdkDE, 7xdkFG, 7xdkHI, 7xdkKJ, 7xdkLM, 7xdkNO, 7xdlED, 7xdlGF, 7xdlIH, 7xh8BC, 7xicEF, 7xicHI, 7xicKL, 7xltHL, 7xmxEF, 7xmxHI, 7xmxKL, 7xmzDE, 7xmzGI, 7xmzHL, 7xnfHL, 7xodXY, 7xq8CL, 7xq8vR, 7xstDE, 7xstFG, 7xstHI, 7xstJK, 7xstLM, 7xstNO, 7xswAB, 7xw6YL, 7xw7AB, 7xxlAC, 7xy8HL, 7xy8IM, 7y0cAB, 7y0cMN, 7y0vHL, 7y0wAB, 7y0wHL, 7y3jHL, 7y7jED, 7y7kBC, 7y7mFG, 7yadCD, 7yckEF, 7ycnCD, 7ycnEF, 7ycnHI, 7ycnKL, 7ydiHL, 7ydyHL, 7ye5HL, 7ye9GI, 7ye9HL, 7ye9NO, 7yegGI, 7yegHL, 7yegNO, 7yhkFH, 7ykjAD, 7ymsFE, 7yr1DF, 7yr1EG, 7yr1HL, 7yrfFE, 7yrhFE, 7yv7FE, 7yxuHL, 7yzjHL, 7z0xHL, 7z0yHL, 7z12BC, 7z3aGI, 7z3aHL, 7z3aJN, 7z3aKP, 7z3aMO, 7zbuHL, 7zf3AB, 7zf3HL, 7zf5DF, 7zf5GI, 7zf6HL, 7zf8HL, 7zf9AB, 7zf9HL, 7zfaQR, 7zfaST, 7zfaXY, 7zfbAB, 7zfbCD, 7zfbHL, 7zfcHL, 7zfdAB, 7zfdHL, 7zfeFG, 7zfeHL, 7zfeJK, 7zfeMN, 7zffHL, 7zlgHL, 7zlhHL, 7zliHL, 7zljHL, 7zlkHL, 7zlkMN, 7zlkOP, 7zlkQR, 7zmkBC, 7zmkEF, 7zmkGH, 7zmkJK, 7zmkMN, 7zmkPQ, 7zmkST, 7zmkVW, 7zr7FG, 7zr7HL, 7zr8HL, 7zr9DE, 7zr9FG, 7zr9HL, 7zrcFG, 7zrcHL, 7zrcJK, 7zwiBC, 7zwiEF, 7zwmBC, 7zwmDE, 7zwmGH, 7zwmIJ, 7zxfBC, 7zxfDE,

7zxgBC, 7zxgEF, 7zxkHL, 7zxkJI, 7zxuHL, 7zyiHL, 8a1eDE, 8adyHL, 8adyQS, 8adzHL, 8adzUV, 8ae0HL, 8ae0QS, 8ae2HL, 8ae2QS, 8ae3HL, 8ae3QS, 8as0BC, 8as0ED, 8as0HG, 8as0KJ, 8as0NM, 8as0QP, 8as0TS, 8as0WV, 8athEF, 8athHL, 8bbnAB, 8bbnCD, 8bbnEF, 8bbnHL, 8bbnJK, 8bbnMN, 8bboHL, 8bczAB, 8bczCD, 8bczEF, 8bczHL, 8bh5HL, 8c3vAB, 8c3vCD, 8c3vHL, 8c7hLK, 8cw9CD, 8cw9GM, 8cw9HL, 8cw9IK, 8cw9JB, 8cw9NO, 8cwvHL, 8cz5HL, 8d0aHL, 8d0zHL, 8d1tHL, 8d36HL, 8d3aHL, 8d47BA, 8d47HL, 8d48HL, 8d6zAB, 8d6zCD, 8d6zEF, 8d6zGI, 8d8qCB, 8d8qHL, 8d8rHL, 8dadHL, 8daoAB, 8daoCD, 8db4CD, 8db4IJ, 8dccHL, 8dd2JI, 8dd2KL, 8dd3JI, 8dd3KL, 8de6BI, 8de6EF, 8df1HL, 8df1MN, 8df1OP, 8df1QR, 8df1ST, 8df1UV, 8df5AB, 8df5CD, 8df5HL, 8df5MN, 8dfgHL, 8dfgIM, 8dfhHL, 8dfiHL, 8dgwAB, 8dgwCD, 8dgwEF, 8dgwHL, 8dimBC, 8dimEF, 8dimHI, 8diuEF, 8diuGC, 8diuHI, 8djkHL, 8djmHL, 8dl6DC, 8dl7DC, 8dl8DC, 8dlrHL, 8dlsHL, 8dm3HL, 8dm3MN, 8dm3PQ, 8dm4HL, 8ds5CB, 8dt3HL, 8duaDE, 8duaHL, 8duaJK, 8dv6CL, 8dv6HD, 8dw2DC, 8dw2HL, 8dw3DC, 8dw3HL, 8dxuAB, 8dxuHL, 8dzhED, 8dzhHL, 8dzhIJ, 8dziBD, 8dziGF, 8dziHL, 8e7mHL, 8eb2JL, 8eb2KM, 8eb2NO, 8ek1AB, 8ek1CD, 8ek1HL, 8ekaAB, 8ekaCD, 8ekaHL, 8ekfHL, 8emaAR, 8emaBL, 8epaAB, 8epaHL, 8erqHL, 8errEF, 8errHL, 8errIJ, 8f6eDC, 8f6eFE, 8f6fDC, 8f6fFE, 8f6hDC, 8f6hFE, 8f6iDC, 8f6iFE, 8f6jDC, 8f6jFE, 8fbwAB, 8fbwCD, 8gngHL, 8gngZY, 8gptDE, 8gptFG, 8gptHI, 8gpuCD, 8gpuFG, 8gpuHL, 8gpuJK, 8gpuNO, 8gpuQR, 8gpxHL, 8gscHL, 8gsdHL, 8gseHL, 8gsfEF, 8gsfHL, 8gv4AB, 8gv4CD, 8gv5CD, 8gv6GH, 8gv6IJ, 8gv7CD, 8gx9HF, 8gx9IL, 8gx9ON, 8gx9PM, 8hc2HL, 8hc3DE, 8hc3HL, 8hc4DE, 8hc4FJ, 8hc4GI, 8hc4HL, 8hc4KM, 8hc4NO, 8hc5DE, 8hc5HL, 8hc6HL, 8hc7HL, 8hc9DE, 8hc9FG, 8hc9HL, 8hcaDE, 8hcaFG, 8hcaHL, 8hcbDE, 8hcbFG, 8hcbHL, 8hhzEF, 8hhzGH, 8hhzIJ, 8hiiHL, 8hijHL, 8hikHL, 8hitAB, 8hpjBC, 8hpjEF, 8hpkHL, 8hs2BC, 8hscFE

Validation Set: 1dlfHL, 1fnsHL, 1gigHL, 1jfqHL, 1jptHL, 1mlbBA, 1nlbHL, 1oaqHL, 1seqHL, 2adfHL, 2d7tHL, 2e27HL, 2fb4HL, 2fbjHL, 2r8sHL, 2v17HL, 2vxvHL, 2w60AB, 2xwtAB, 2ypvHL, 3e8uHL, 3eo9HL, 3g5yBA, 3gizHL, 3gnmHL, 3go1HL, 3hc4HL, 3hntHL, 3i9gHL, 3lizHL, 3lmjHL, 3m8oHL, 3mxwHL, 3npsBC, 3oz9HL, 3p0yHL, 3v0wHL, 4f57HL, 4h20HL, 4hpyHL, 4nzuHL

Test Set: 7ox3AB, 7t25HL, 7t86HL, 7tqgHL, 7u61HL, 7u62HL, 7u64AB, 7u8eHL, 7ucxHL, 7uvoAB, 7uvsAB, 7uylHL, 7uymHL, 7uzcEF, 7x2hAB, 7x3nBA, 7x63HL, 7x8pCB, 7yclAB, 7z2mBA, 7z3wHL, 8bbhHL, 8be1AB, 8d29AB, 8d54AB, 8dg9HL, 8dguHL, 8dtrEF, 8dttAB, 8dxtHL, 8ek6HL, 8ffeHL

C.2 Algorithmic details for message passing layers

The MPL consists of the edge message, the message aggregation, and the node update blocks. $\mathbf{h}$ and $\mathbf{e}$ refer to node and edge features, respectively. $\cdot^l$ refers to some quantity or vector in layer l. We begin with the edge message block, where $\mathbf{h}_{src}$ and $\mathbf{h}_{dst}$ refer to source and destination nodes of a message:

$$\mathbf{x}_e^{(l-1)} = \mathbf{h}_{src}^{(l-1)} \,||\, \mathbf{e}^{(l-1)} \,||\, \mathbf{h}_{dst}^{(l-1)} \tag{C.1}$$

$$\mathbf{x}' = \text{InstanceNorm}(\mathbf{x}_e^{(l-1)}) \tag{C.2}$$

$$\mathbf{x}'' = \text{LeakyReLu}(\text{Linear}(\mathbf{x}')) \tag{C.3}$$

$$\mathbf{e}^{(l)}, \mathbf{c}_e^{(l)} = \text{LSTM}(\mathbf{x}'', \mathbf{e}^{(l-1)}, \mathbf{c}_e^{(l-1)}) \tag{C.4}$$

where $\mathbf{c}_e$ refers to the cell state of the LSTM unit. Lastly, we calculate the attention per edge, noting that the softmax calculation is sparse:

$$\boldsymbol{\alpha}^{(l)} = \text{softmax}(\text{Linear}(\mathbf{e}^{(l)})) \tag{C.5}$$

In the aggregation block, a weighted average of incoming edges m is calculated for each node i:

$$\mathbf{m}_i^{(l)} = \sum_{j \in \mathcal{N}(i)} (\alpha_i^{(l)} \odot \mathbf{e}_i^{(l)}). \tag{C.6}$$

Lastly, the node reduce block updates the nodes using aggregated messages:

$$\mathbf{x}_n = \mathbf{m}_i^{(l)} \,||\, \mathbf{h}^{(l-1)} \tag{C.7}$$

$$\mathbf{x}' = \text{InstanceNorm}(\mathbf{x}_n) \tag{C.8}$$

$$\mathbf{x}'' = \text{ReLu}(\text{Linear}(\mathbf{x}')) \tag{C.9}$$

$$\mathbf{h}^{(l)}, \mathbf{c}_h^{(l)} = \text{LSTM}(\mathbf{x}'', \mathbf{h}^{(l-1)}, \mathbf{c}_h^{(l-1)}) \tag{C.10}$$

where $\mathbf{c}_h$ refers to the cell state of the LSTM unit.

C.3 Supplementary figures

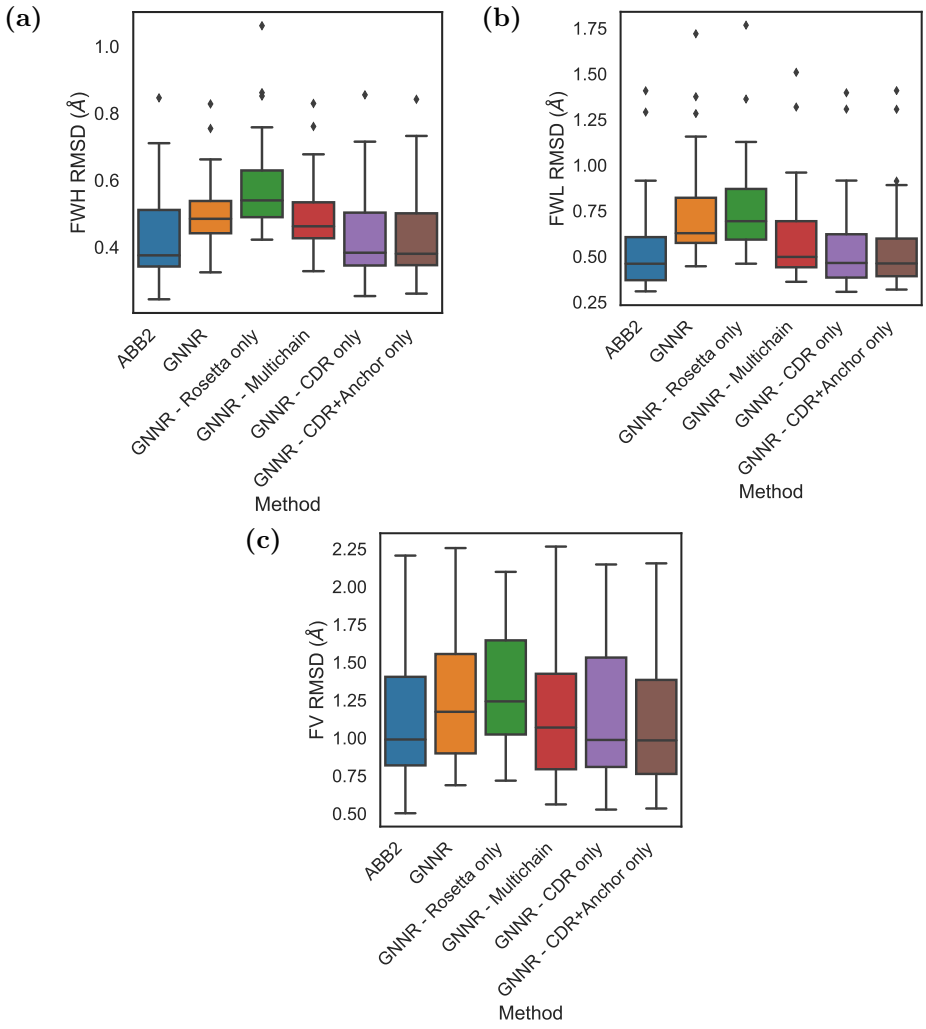


Figure C.1: Further performance of modified GNNRefine versions on ABB2 models. Antibody structures unseen by GNNRefine were modelled using ABB2 followed by refinement using GNNRefine and variants. Shown are the distributions of RMSD values compared to crystal structures for the different versions of GNNR, and for ABB2. The boxplots are drawn using the interquartile ranges, with the middle line denoting the median. **(a)** shows V_H FW regions, **(b)** shows V_L FW regions, and **(c)** shows the entire F_V .

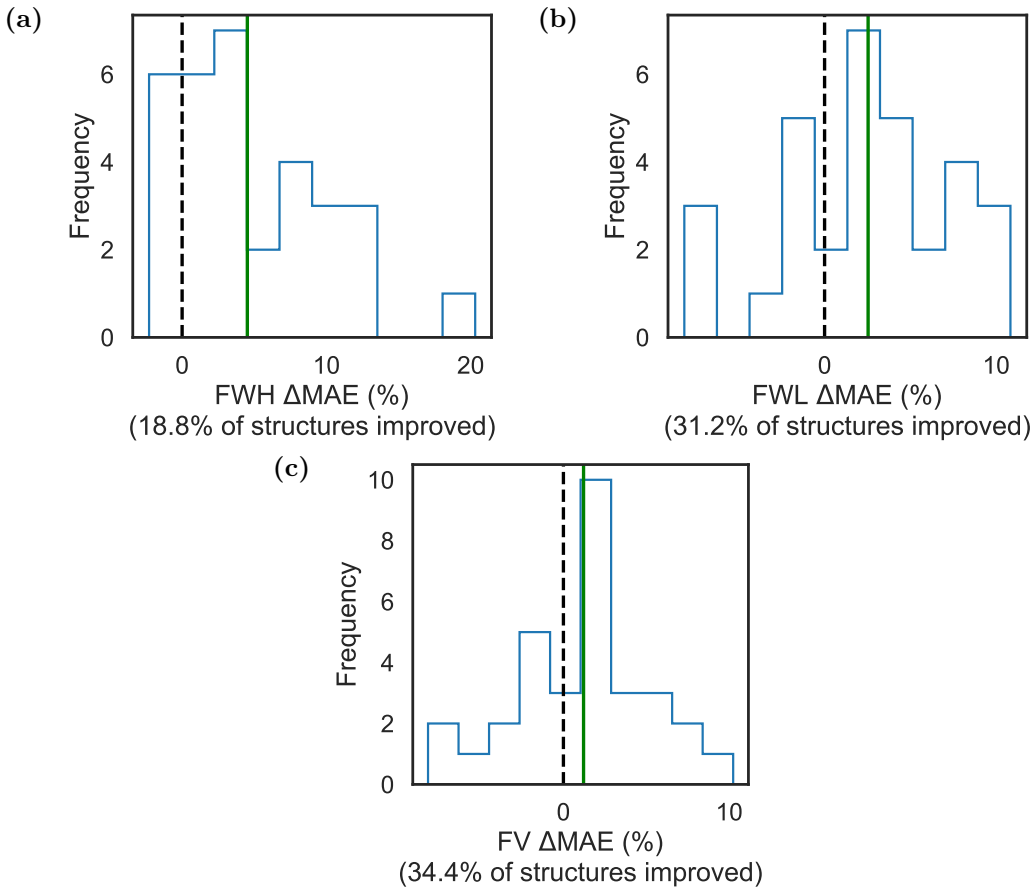


Figure C.2: Changes in MAE of inter-Cβ distances compared to ABB2. Shown are the percentage changes in MAE relative to ABB2 when refined inter-Cβ distances were inferred using our model. The green lines represent averages across all test set structures, and the percentage value shows the number of structures where our models improves the average MAE compared to ABB2. **(a)** shows V_H FW residues, while **(b)** shows V_L FW residues. **(c)** shows the entire F_V .

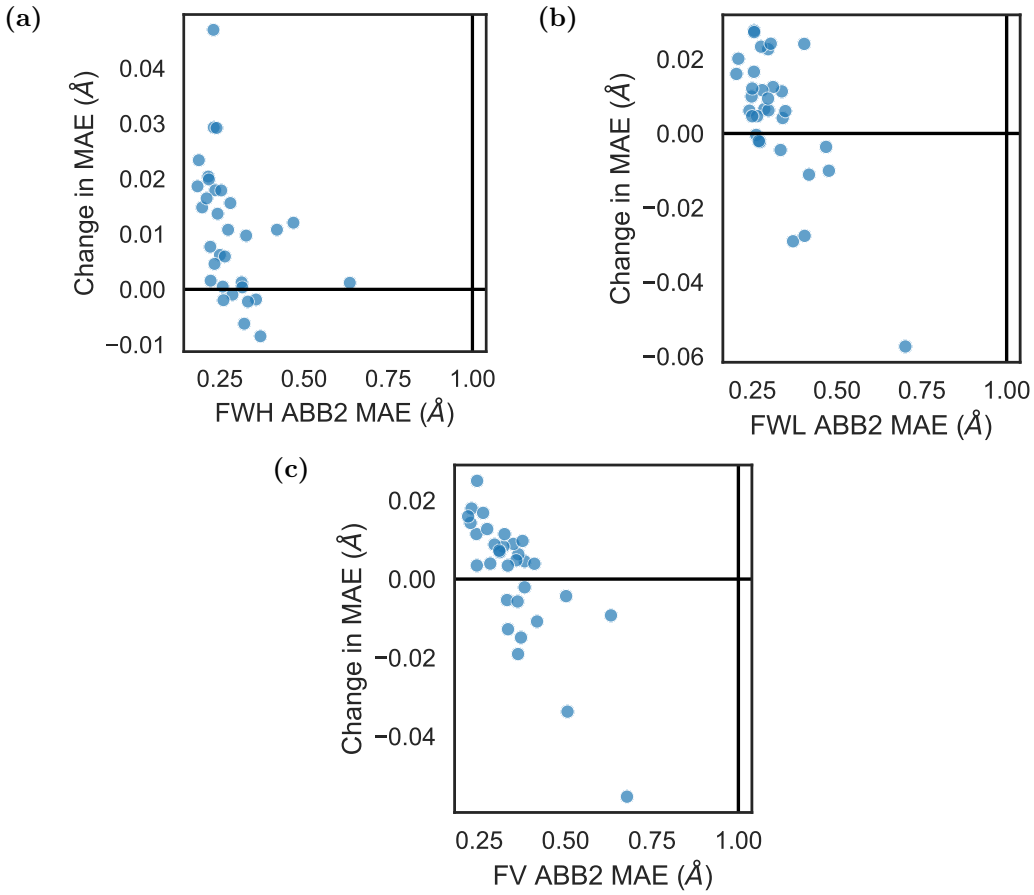


Figure C.3: Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 error. Shown are the absolute changes in MAE relative to ABB2 when refined inter-Cβ distances were inferred using our model, plotted against the original mean MAE for each ABB2 model. **(a)** shows V_H FW residues, while **(b)** shows V_L FW residues. **(c)** shows the entire F_V .

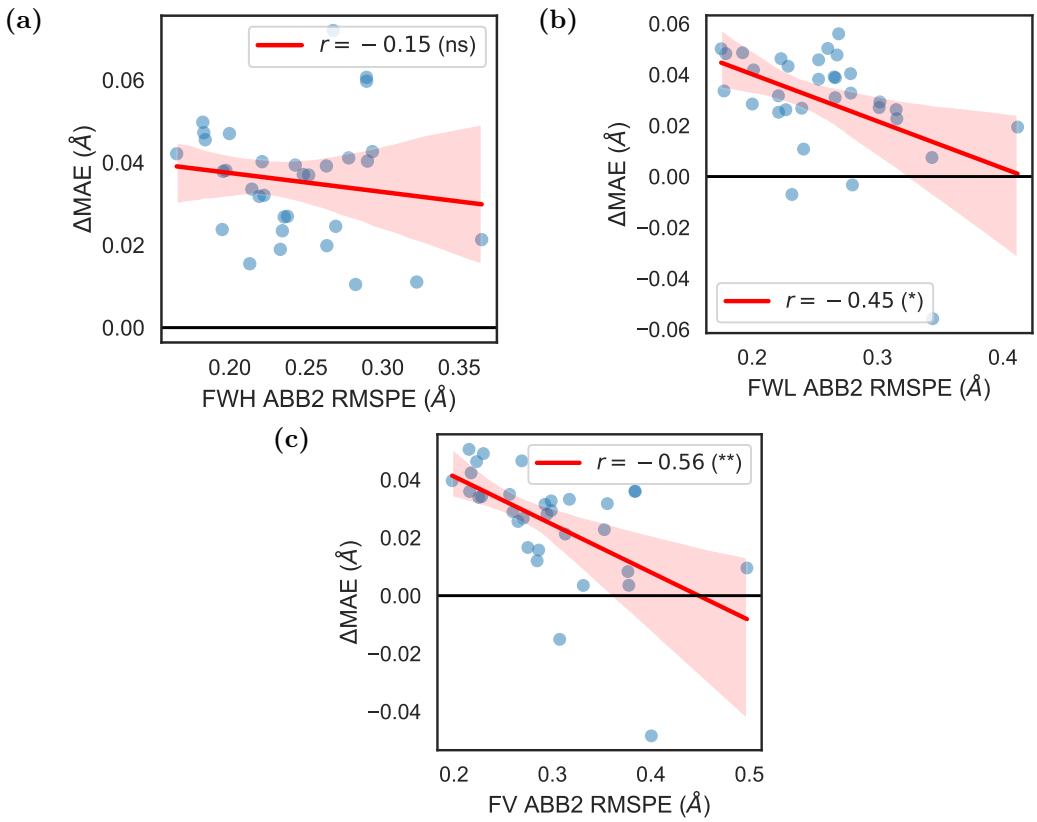


Figure C.4: Changes in MAE of inter-Cβ distances compared to ABB2 as a function of ABB2 predicted error. Shown are the absolute changes in MAE relative to ABB2 when refined inter-Cβ distances were inferred using our model, plotted against the ABB2 predicted root mean squared error of residues. The red lines show lines of best fit, with the shaded regions representing the 95% confidence interval for the regressions. Pearson correlation coefficients are reported for the linear regressions. (ns: not significant, *: $0.005 \leq p < 0.05$, **: $p < 0.005$) **(a)** shows V_H FW residues, while **(b)** shows V_L FW residues. **(c)** shows the entire F_V .

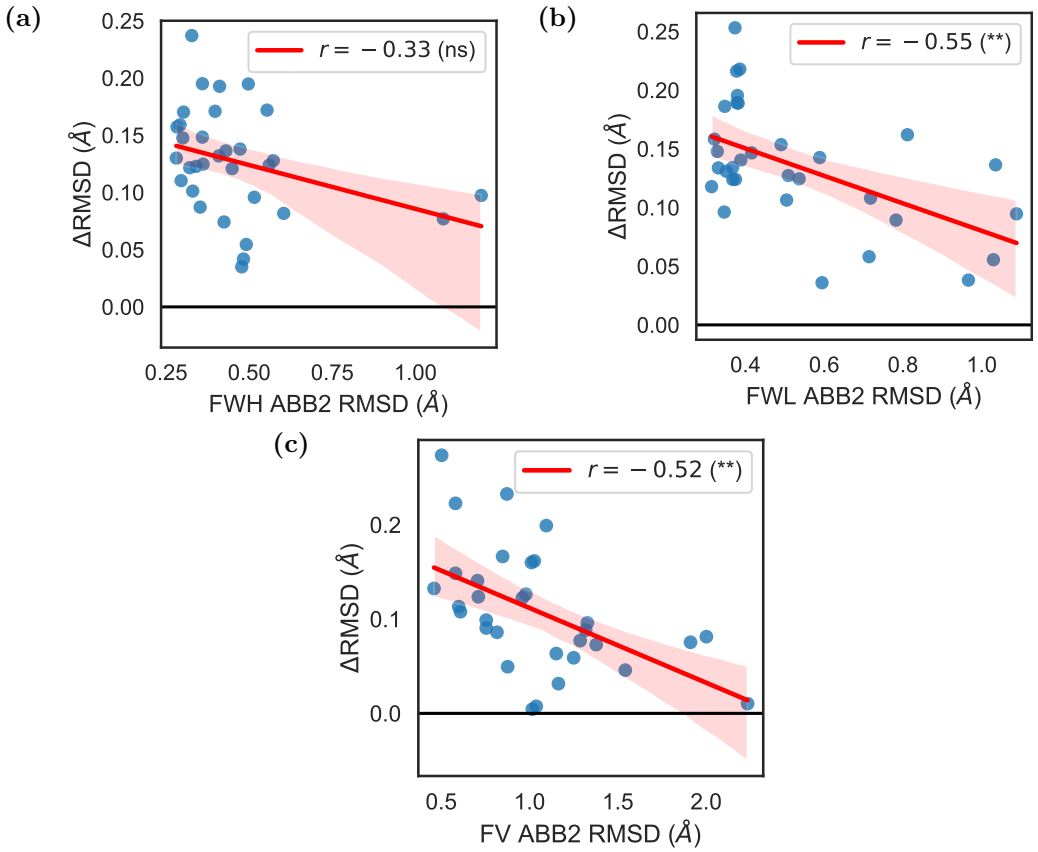


Figure C.5: Changes in RMSD of refined structures compared to ABB2. We refined the ABB2 structural models using OpenMM by placing harmonic positional restraints on framework heavy atoms and adding harmonic distance restraints on CDR loop residues centered at the predicted distance for each edge. The red lines show lines of best fit, with the shaded regions representing the 95% confidence interval for the regressions. Pearson correlation coefficients are reported for the linear regressions. (ns: not significant, *: $0.005 \leq p < 0.05$, **: $p < 0.005$) **(a)** shows V_H FW residues, while **(b)** shows V_L FW residues. **(c)** shows the entire F_V .

D
Additional Publications

While the majority of this DPhil was concerned with projects surrounding antibody structure and function, the author was also involved in projects within the field of structural parasitology, investigating the dynamics of malarial antigens. As this work is beyond the scope of this thesis, publications arising from it are included in this appendix.

nature communications

Article

https://doi.org/10.1038/s41467-023-39357-w

Structural basis for DARC binding in reticulocyte invasion by *Plasmodium vivax*

Received: 28 February 2023

Accepted: 6 June 2023

Published online: 19 June 2023

Check for updates

Re'em Moskovitz[1,3], Tossapol Pholcharee[1,3], Sophia M. DonVito[2], Bora Guloglu[1], Edward Lowe[1], Franziska Mohring[2], Robert W. Moon[2] & Matthew K. Higgins[1]

The symptoms of malaria occur during the blood stage of infection, when the parasite replicates within human red blood cells. The human malaria parasite, *Plasmodium vivax*, selectively invades reticulocytes in a process which requires an interaction between the ectodomain of the human DARC receptor and the *Plasmodium vivax* Duffy-binding protein, PvDBP. Previous studies have revealed that a small helical peptide from DARC binds to region II of PvDBP (PvDBP-RII). However, it is also known that sulphation of tyrosine residues on DARC affects its binding to PvDBP and these residues were not observed in previous structures. We therefore present the structure of PvDBP-RII bound to sulphated DARC peptide, showing that a sulphate on tyrosine 41 binds to a charged pocket on PvDBP-RII. We use molecular dynamics simulations, affinity measurements and growth-inhibition experiments in parasites to confirm the importance of this interaction. We also reveal the epitope for vaccine-elicited growth-inhibitory antibody DB1. This provides a complete understanding of the binding of PvDBP-RII to DARC and will guide the design of vaccines and therapeutics to target this essential interaction.

In many regions of the world, *Plasmodium vivax* is the predominant parasite that causes human malaria. Leading to around 14.5 million diagnosed cases each year, it accounted for an estimated 61% of malaria in the Americas and 48% in South East Asia in 2017[1]. With studies suggesting that the impact of this parasite is underestimated[2,3], morbidity caused by *P. vivax* is a major global health problem and an effective vaccine is urgently required to contribute to the malaria eradication effort.

The symptoms of malaria occur when parasites replicate within the blood. In the case of *P. vivax*, invasion of red blood cells is highly dependent on the reticulocyte Duffy antigen/receptor for chemokines (DARC). The importance of DARC as an invasion receptor[4] was first shown over 45 years ago, with the demonstration that the related parasite, *Plasmodium knowlesi*, is unable to invade blood cells taken from individuals with the Duffy-negative phenotype[5]. These people have a mutation in their DARC gene which makes the receptor undetectable on circulating reticulocytes. Duffy negativity was also found to prevent African-Americans from being infected with *P. vivax* through mosquito bite[6]. The widespread prevalence of *P. vivax* across the world, together with its much-reduced occurrence in regions of Africa where Duffy-negativity is widespread, highlight the importance of DARC as a determinant of *P. vivax* susceptibility[7]. Indeed, while *P. vivax* is occasionally capable of infecting Duffy-negative individuals, this is linked to lower parasite load and more than 15-fold reduced likelihood of causing clinical disease[8–10]. Preventing DARC-mediated invasion remains a primary approach to prevent clinical vivax malaria.

The parasite binding partner for DARC is the Duffy binding protein, PvDBP[11] and multiple lines of evidence highlight the important nature of their interaction. In the closely related, genetically tractable parasite, *P. knowlesi*, knockout of PkDBPα prevents invasion of Duffy-positive erythrocytes in vitro[12–14]. While PvDBP knockout in *P. vivax* has not been technically possible, due to challenges with maintaining this parasite in culture conditions, immunisation of mice, rabbits and non-human primates with the RII region of PvDBP (PvDBP-RII) induces

[1]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [2]London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [3]These authors contributed equally: Re'em Moskovitz, Tossapol Pholcharee. e-mail: matthew.higgins@bioch.ox.ac.uk

inhibitory antibodies that block binding of PvDBP to DARC[15,16]. In humans, high titres of naturally acquired antibodies that target PvDBP-RII and prevent DARC binding in vitro, are associated with reduced risk of *P. vivax* infection[17], lower parasite densities following invasion and decreased risk of clinical malaria[18,19]. Immunisation of human volunteers with recombinant viral vectors expressing PvDBP-RII induces strain-transcending antibodies which prevent recombinant PvDBP-RII from binding to DARC while human antibodies, derived from either vaccination or natural infection, are inhibitory of invasion[13,20]. PvDBP is therefore the primary blood stage vaccine candidate to prevent vivax malaria.

With both intact PvDBP and DARC proving challenging to produce, molecular studies have used smaller fragments to narrow down functionally important regions[21], with a ~ 350 amino acid residue Duffy-binding-like (DBL) domain known as PvDBP-RII, shown to bind DARC[22]. PvDBP-RII binds to the sixty-residue N-terminal ectodomain of DARC[23]. The first structural study related to PvDBP was a crystal structure of PkDBP-RII, from *P. knowlesi*[24]. This was followed by structures of *Plasmodium vivax* DBP-RII bound to a 30 residue-long peptide ($DARC_{14-43}$), of which 11 residues ($DARC_{19-30}$) were ordered and were observed in the structure as a helical peptide[25].

However, DARC is post-translationally modified by sulfation of two tyrosine residues, Y30 and Y41 and mutation of Y41 to phenylalanine reduced PvDBP binding[26]. These residues are not observed in existing structures, most likely because the DARC ectodomain used was bacterially expressed and therefore lacked tyrosine sulfation[26,27]. Studies of *P. knowlesi* DBP-RII also proposed the locations for binding sites of the sulfated tyrosines, with positively charged patches on PkDBP-RII suggested to form the binding sites for Y41[24] and Y30[28]. Indeed, a later refinement of the PkDBP-RII structure showed a sulfate ion bound in the proposed binding site for Y41[28]. Mutations have also been described in regions of PvDBP which do not contact $DARC_{19-30}$ but which are known to contribute to full-length DARC binding[29,30] and a polymorphism in residue 42 of DARC is the sole change in the Fy^a blood group, which reduces the likelihood and severity of *P. vivax* infection[31]. Indeed, NMR experiments indicate that residues from 14-43 of DARC show chemical shifts on binding to PvDBP-RII[25]. The full interaction between PvDBP and DARC therefore remains to be defined.

Structural approaches have also started to reveal how the PvDBP-DARC interaction can be blocked by mouse and human monoclonal antibodies. Two human monoclonals bind the region of PvDBP involved in $DARC_{19-30}$ binding and dimerisation[20]. In contrast, a broadly inhibitory human monoclonal, from a vaccinated human volunteer, binds away from this site[13], as do inhibitory mouse antibodies[32]. The molecular mechanisms by which these antibodies work are uncertain.

In this study, we determine the structural basis for the interaction of PvDBP with the DARC ectodomain and the human invasion blocking antibody, DB1. By using DARC proteins expressed in human cells, we define the binding site for sulfated tyrosine 41. This reveals the molecular basis for the role of tyrosine sulfation in the essential interaction between PvDBP and DARC in reticulocyte invasion by *P. vivax*.

Results

Structural characterisation of the PvDBP-DARC interaction

To determine the structure of PvDBP-RII bound to sulfated DARC ectodomain we started by expressing the N-terminal 60 residues of DARC ($DARC_{ecto}$) in HEK293 cells. This was purified and assessed by mass spectrometry, with the predominant peak having the mass expected for the peptide with the addition of two sulfates (Supplementary Fig. 1). We also expressed PvDBP-RII in HEK293 cells and this was mixed with $DARC_{ecto}$ for crystallisation. As no crystals formed, we attempted to use Fab fragments from monoclonal antibodies which bind to PvDBP-RII, but do not prevent it from binding to $DARC_{ecto}$, as crystallisation chaperones[13]. A complex of PvDBP-RII, $DARC_{ecto}$ and the Fab fragment of the DB1 monoclonal antibody[13] formed crystals which diffracted to 2.49 Å resolution. Molecular replacement using the PvDBP-RII[13] structure as a search model provided phase information and the resultant electron density map revealed density for residues 19-47 of $DARC_{ecto}$ (Fig. 1 and Supplementary Table 1).

A comparison with the previously determined structure[25] of PvDBP-RII bound to $DARC_{ecto}$ expressed in bacteria shows DARC residues 19-30 to adopt the same binding mode in both cases, forming an α-helix which packs against subdomain 2 of PvDBP-RII, with a root mean square deviation of 0.22 Å over the helical residues 22-29 of DARC (Fig. 1b). While this is the only region of DARC observed in the previous study, we now also observe clear density for residues 31-47. These form an elongated peptide which wraps, in a horse-shoe-shaped trajectory, around a protrusion on the surface of PvDBP-RII subdomain 2. Clear density was observed for each residue, including a binding pocket for Y41 and the sulfate on Y41 (Fig. 1c), which closely matched a binding site previously proposed[28]. Also visible is the side chain of residue D42, which is polymorphic (D42G) in the Fy^a/Fy^b blood group variant (Fig. 1c), but which does not make clear interactions with PvDBP-RII. Residues 48-60 are not visible in the crystal structure, suggesting a disordered linker between the PvDBP-RII-binding region of the ectodomain and the DARC transmembrane region, which starts at residue 61.

Sulfated Y41 binds to a stable binding pocket on PvDBP while Y30 is dynamic

The structure allows us to assess the role of sulfation of Y30 and Y41 of DARC on the interaction with PvDBP. A previous study has shown that mutation of Y30 to F, which removes the hydroxyl group, as well as the potential for sulfation, did not affect PvDBP-RII binding[26]. Indeed, in our structure, the Y30 side chain was clearly visible in the electron density, but no evidence was seen for sulfation, suggesting either that the sulfate was not present, or that it was disordered in the electron density (Fig. 1d). The side chain did not contact PvDBP-RII, but instead interacts with a neighbouring DB1 Fab fragment through crystal contacts.

In contrast, the side chain of Y41, together with clear density for a sulfate group, was observed. This docks into a positively charged pocket on the PvDBP-RII surface, where the sulfate makes direct salt-bridges with K301 and R304 (Fig. 1c). This is consistent with sulfated Y41 (Y41-S) forming an important part of the binding interface. Indeed, the more substantial Y41F mutation, which removes both sulfate and hydroxyl group, reduces PvDBP binding[26] and this interaction is likely to be important for residues 31-47 to adopt their correct binding conformation.

We next assessed the stability of the binding modes of Y30 and Y41 using metadynamics simulations allowing us to generate the relative free energy landscapes of the side chains of sulfated and non-sulfated versions of Y30 and Y41[33,34]. These simulations were run for three different molecular models, generated from our crystal structure. These were sulfated on both Y30 and Y41, sulfated on just Y30 and sulfated on just Y41. In each case, we analysed the ensemble of structures generated during the simulation, the free energy surfaces and the number and duration of contacts formed between each tyrosine residue and PvDBP-RII (Fig. 2).

In the case of Y41, sulfation leads to an average free energy decrease of 7.63 kJ/mol with the decrease in free energy concentrated in the existing global minimum, at approximately $\chi_1 = -161°$, where we observed a local decrease of 17.87 kJ/mol, compared with $\chi_1 = -144.3°$ in the crystal structure (Fig. 2a). This corresponds with Y41 adopting a more preferred location when sulfated (Fig. 2b). In contrast, while sulfation of Y30 leads to an average free energy decrease of 8.75 kJ/mol, this changes the shape of the free energy surface, with the free energy barrier at approximately $\chi_1 = -120°$ disappearing (Fig. 2c) and Y30 being able to explore a

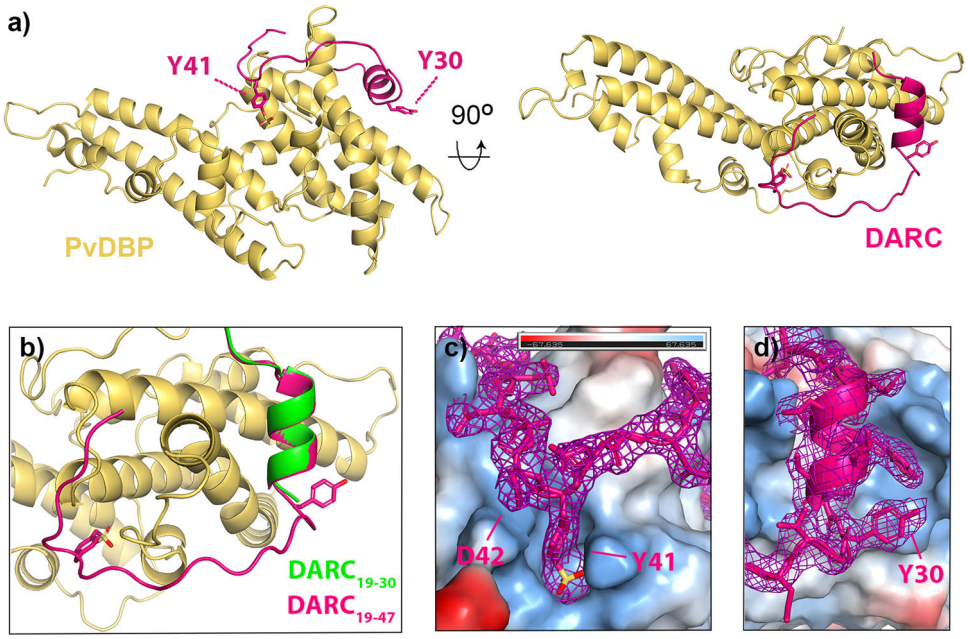


Fig. 1 | The structure of PvDBP-RII bound to sulfated DARC ectodomain. a The structure of PvDBP-RII (yellow) bound to the DARC ectodomain (pink). Residues Y30 and Y41 from DARC are highlighted as sticks, with sulfur in yellow and oxygen in red. **b** A close up of the DARC ectodomain, showing residues 19-47 of the sulfated ectodomain coloured as **a**), overlayed with residues 19-30 of the non-sulfated ectodomain in green (from PDB: 4NUV)[25]. **c** A close up of residue 41 of DARC, with DARC and the electron density surrounding DARC in pink and PvDBP-RII as a surface coloured by electrostatics (blue as positive charge and red as negative, estimated in pymol). **d** A close up of residue 30 of DARC, with DARC and the electron density surrounding DARC in pink and PvDBP-RII as a surface coloured by electrostatics (blue as positive charge and red as negative, estimated in pymol). In both **c**, **d**, the region of the $2F_O$-F_C map within 2 Å of DARC is shown at a contour level of 1.1.

wider set of χ_1 configurations when sulfated (Fig. 2d). The location of Y30 in the crystal structure has $\chi_1 = 71.6°$, which lies close to the local energy minima at $\chi_1 = \sim 65°$ in the free energy landscape. Therefore, while the sulfate on Y30 is not observed in the crystal structure, and Y30 does not contact PvDBP-RII, simulations suggest that a flexible Y30S can interact with PvDBP-RII through a dynamic electrostatic interaction.

We next assessed the number of contacts which each residue from DARC forms with PvDBP and determined the fraction of the simulation during which each contact occurs. Using a heavy atom distance cut-off of 4.5 Å to define contacts, we observe that both Y30 and Y41 make more contacts with PvDBP when sulfated. This is much more pronounced for Y41 where the average number of contacts increases from 3.64 to 7.46, whereas in Y30 we observe a shift from 1.30 to 1.87 (Fig. 2e). We observe a similar pattern when considering the durability of specific interactions, with Y41 making more interactions with PvDBP than Y30 and with sulfation increasing both the number and durability of interactions (Fig. 2f). In particular, Y41 interacts continuously with R304 during the simulation, which it can only reach when sulfated. We therefore find that sulfation of Y41 stabilises a specific location for the side chain in which it makes strong electrostatic interactions, while sulfation of Y30 does not lead to the formation of such a favourable contact, but instead increases the strength and durability of a smaller number of transient interactions.

Y41-S contributes more to binding affinity and erythrocyte invasion than Y30-S

The nature of the compact binding pocket for the sulfate on Y41, and the lack of a similar binding pocket for that on Y30, lead to the prediction that the sulfate on Y41 contributes more to the binding affinity than that on Y30. To test this, we generated a set of peptides containing the ordered region of DARC ectodomain (residues 19-47) and analysed their binding to PvDBP-RII by isothermal titration calorimetry (ITC) (Fig. 3a and Supplementary Fig. 2). A peptide in which both Y30 and Y41 are sulfated bound to PvDBP-RII with an affinity of 58 nM, while a peptide in which neither sulfate is present, bound with much lower affinity of 23 µM, quantifying the importance of sulfation for this interaction. We next tested peptides in which either Y30 or Y41 is sulfated. A peptide in which Y30 is sulfated, but Y41 is not, bound with an affinity of 552 nM, showing that the loss of Y41 sulfation leads to a ~9-fold reduction in affinity. In contrast, a peptide in which Y41 is sulfated but Y30 is not, bound with an affinity of 168 nM, with a ~3.5-fold reduction in binding affinity from the fully sulfated peptide. Therefore, both sulfates impact binding affinity, with Y41 providing a greater contribution to affinity.

We next assessed whether the differential effects of Y30 and Y41 sulfation on the affinity of DARC for PvDBP are mirrored during red blood cell invasion in vitro. To test this, we used a transgenic *P. knowlesi* line which has been modified to express PvDBP instead of the native, orthologous DARC binding protein, PkDBPα[13,14]. This transgenic parasite invades Duffy-positive erythrocytes in culture and is similarly affected by PvDBP-RII-targeting antibodies as *P. vivax* clinical isolates[13]. As it is currently technically impossible to generate transgenic erythrocyte lines in which DARC sulfation is specifically modulated, we instead assessed the ability of our four DARC ectodomain peptides to inhibit the growth of this transgenic PvDBP-expressing line, by measuring the effect of different concentrations of each peptide in a growth inhibition assay (Fig. 3b). We found that the Y30-S/Y41-S double sulfated peptide was most effective at blocking parasite growth ($IC_{50} = 0.72$ µM), followed by the Y41-S peptide ($IC_{50} = 2.99$ µM), with no growth inhibition observed for the Y30-S and non-sulfated peptides. Therefore, the effect of the peptides on parasite growth inhibition and the affinity that the peptides have for PvDBP-RII show the same pattern, with sulfation of Y41 having the greatest effect on peptide efficacy and sulfation of Y30 making a significant, but smaller difference.

We also conducted the equivalent experiment using unmodified (wild-type) *P. knowlesi* and observed a similar outcome, with

https://doi.org/10.1038/s41467-023-39357-w

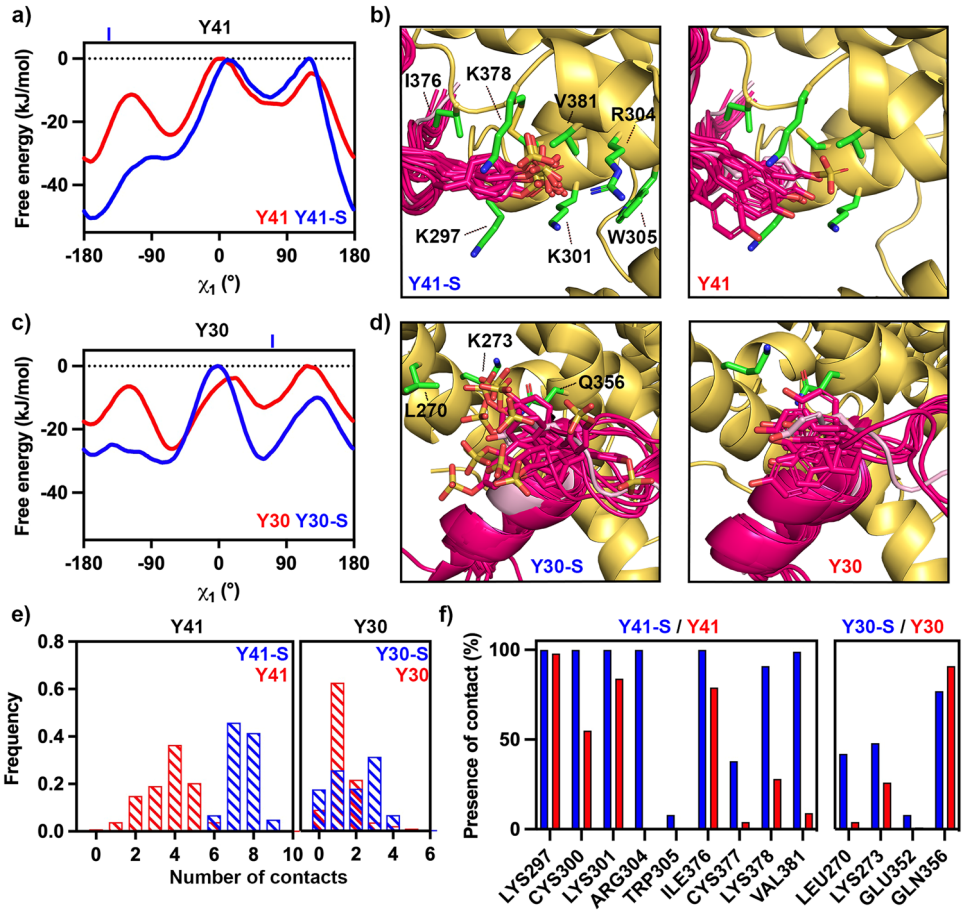


Fig. 2 | Molecular dynamics simulations indicate ordered binding of Y41 but not Y30. a Free energy landscapes for residue Y41 sulfated (blue) and non-sulfated (red) relative to the χ_1 angle of the tyrosine, showing that sulfation favours a single binding position. In the crystal structure, $\chi_1 = -144.3°$, indicated by a blue line above the graph. **b** Representative images from across the simulation, showing the degree of motion of Y41 in its sulfated (left) and non-sulfated (right) forms. In each case, PvDBP is yellow and DARC is bright pink. The crystal structure is shown for comparison in light pink. In each case, sulfate atoms are yellow, oxygen atoms red and nitrogen atoms blue. **c** Free energy landscapes for residue Y30 sulfated (blue) and non-sulfated (red) relative to the χ_1 angle of the tyrosine, showing that sulfation disfavours a single binding position. In the crystal structure, $\chi_1 = 71.6°$, indicated by a blue line above the graph. **d** Representative images from across the simulation, showing the degree of motion of Y30 in its sulfated (left) and non-sulfated (right) forms, coloured as in **b**). **e** The frequency of the number of contacts formed between PvDBP-RII and Y41 (left) and Y30 (right) in the sulfated (blue) and non-sulfated (red) forms during the simulations. **f** The percentage of frames from across the simulation in which each residue from PvDBP forms interactions with Y41 (left) and Y30 (right) in the sulfated (blue) and non-sulfated (red) forms. Source data are provided as a Source data file.

IC_{50} = 0.79 μM for the double sulfated peptide, IC_{50} = 3.92 μM for Y41-S and no inhibition observed for non-sulfated and Y30-S peptides, highlighting the parallel in DARC binding between PvDBP and PkDBPα. Therefore, while a panel of PvDBP RII-targeting antibodies which blocked invasion of *P. vivax* clinical isolates and PvDBP-expressing *P. knowlesi* (including DB1 and DB9) were not cross-reactive against wild-type *P. knowlesi*[13], the similar effect of DARC-based peptides on these two parasite lines indicates that PvDBP and PkDBPα engage DARC in a similar way.

Mapping the epitope for growth-neutralising antibody DB1

We next analysed the location of the epitope for DB1 (Fig. 4). This antibody inhibits growth of transgenic *P. knowlesi* parasites expressing PvDBP and of a subset of patient isolates of *P. vivax* which express different PvDBP variants[13]. Notably, DB1 binds to a site on PvDBP-RII which contains the polymorphic ^{339}DEK^{341} motif[35], making direct contact with these residues (Supplementary Fig. 3), explaining in part why it is not a strain-transcendent neutralising antibody[13]. DB1 also binds to a site which does not overlap the DARC binding surface, explaining why it does not directly block $DARC_{ecto}$ binding[13] (Fig. 4a).

Previous structural studies of PvDBP-RII have observed dimers to form within crystals and have also shown DARC-dependent dimerisation at high concentrations in solution in the presence of a non-sulfated DARC peptide[25,27]. DB1 would block this proposed dimerisation interface, with the Fab fragment overlapping with the second PvDBP copy and associated DARC peptide from the dimer (Fig. 4b). To determine whether the mechanism of action of DB1 is inhibition of PvDBP-RII dimerisation, we used small-angle X-ray scattering to assess the solution mass of a complex consisting of PvDBP-RII bound to Fab fragment of DB1. As predicted, the PvDBP-RII:DB1 complex showed the mass and maximum dimensions expected for a monomeric complex (Supplementary Fig. 4). We also conducted SAXS experiment for PvDBP-RII bound to Y30-S Y41-S $DARC_{19-47}$, but found that aggregation confounded collection of interpretable data. We therefore collected SAXS data for PvDBP-RII bound to Y30-S Y41-S $DARC_{19-47}$ and to the Fab fragment of antibody DB9. As DB9 binds distant from the dimerisation interface it would not prevent dimer formation. Both in SEC-SAXS experiments, in which the complex was passed through a size-exclusion column immediately before SAXS data collection (Supplementary Fig. 4), or when studied at high concentrations of up to 6 mg/ml in solution

https://doi.org/10.1038/s41467-023-39357-w

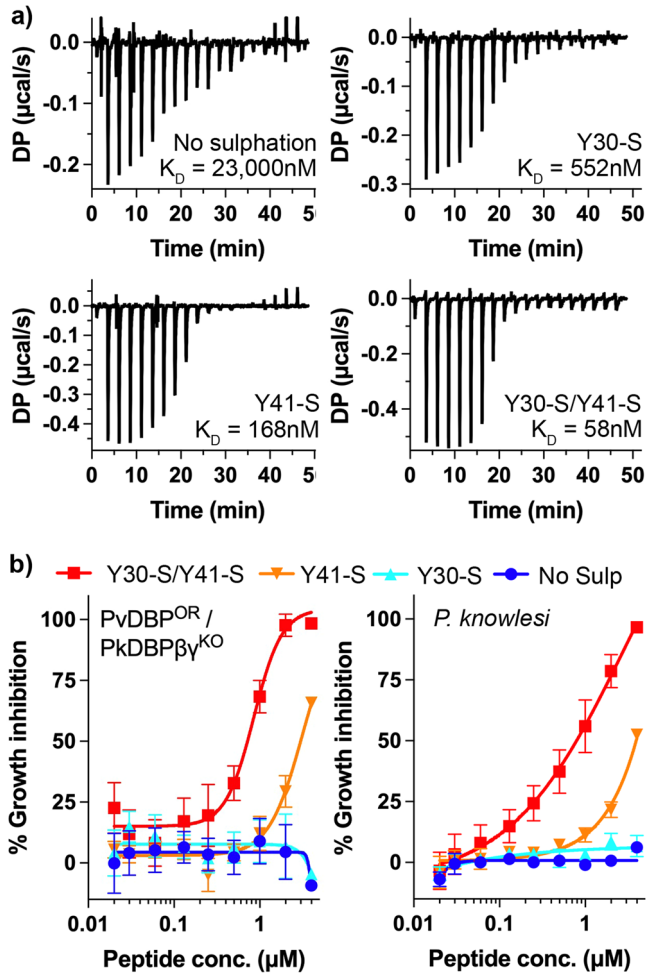


Fig. 3 | The effect of tyrosine sulfation of DARC on PvDBP-RII affinity and parasite invasion. a Isothermal titration calorimetry measurements of the binding of synthetic DARC peptides to PvDBP-RII. Shown are single representative traces for non-sulfated peptide and peptides sulfated on Y30, on Y41 or on both 30 and 41. Each stated K_D value is the mean from n = 3 technical replicates. **b** Growth-inhibitory activity for the same four peptides in an assay which assess the growth of a *Plasmodium knowlesi* line in which PkDBPs have been replaced by PvDBP ($PvDBP^{OR}$/$PkDBP\beta\gamma^{KO}$, left) or wild-type *Plasmodium knowlesi* (right). The Y30-S/Y41-S peptide inhibited with an IC_{50} of 0.72 µM for $PvDBP^{OR}$/$PkDBP\beta\gamma^{KO}$ and 0.79 µM for *P. knowlesi*. The Y41-S peptide inhibited with an IC_{50} of 2.99 µM for $PvDBP^{OR}$/$PkDBP\beta\gamma^{KO}$ and 3.92 µM for *P. knowlesi*. Technical replicates ($n = 2$) from each assay were averaged, and data presented represents the mean ± standard error of the mean of four separate biological replicates ($n = 2$ in Fy^a donor blood, and $n = 2$ in Fy^b, to account for variation between DARC alleles). IC_{50} values were identified using a variable slope four-parameter logistic curve. Source data are provided as a Source data file.

(Supplementary Fig. 5) we did not observe dimerisation of the PvDBP-RII:DB9:$DARC_{19\text{-}47}$ complex. As PvDBP-RII bound to sulfated $DARC_{19\text{-}47}$ does not dimerise in our hands, we cannot therefore conclude that DB1 acts by blocking dimerisation.

Discussion

Sulfation of DARC has been known for decades to play an important role in its binding to both cytokines and to PvDBP. However, previous structural studies used bacterially expressed DARC ectodomains, which lack tyrosine sulfation. To test whether these sulfates impact PvDBP binding, we therefore determined a structure of PvDBP-RII in complex with a sulfated DARC ectodomain. In the presence of tyrosine sulfation, we see an additional 17 residues of DARC, binding in an extended conformation around a protrusion on the PvDBP surface. We found that the sulfate on Y41 docks into a compact pocket on PvDBP, making direct interactions which are likely to be required for residues 31-47 of DARC to become ordered. Indeed, specific removal of the sulfate on Y41 leads to a ~9-fold reduction in affinity and negates growth-inhibitory activity at the concentrations tested. In contrast, Y30 plays a less significant role in the interaction, not binding to PvDBP in the structure, dynamic in molecular dynamics simulations and with specific removal of the sulfate leading to ~3.5- and ~4-fold reductions in affinity and growth inhibitory activity.

Our study also reveals the epitope for the growth-neutralising antibody DB1 and shows that it binds to a polymorphic site distant from the DARC binding site, as predicted based on previous studies which show that DB1 does not block DARC binding[13]. Instead, DB1 binds in a location which would overlap with a putative dimerisation interface on PvDBP-RII[27]. The growth-neutralising antibodies which target PvDBP-RII, and have known structures, bind to a variety of locations, suggesting a range of possible molecular mechanisms for neutralisation (Fig. 4c). DB9[13] and 2D10[32] lie on subdomain 3, distant from both DARC binding site and the putative dimerisation interface, 092096[20] blocks both DARC binding and occludes the putative dimerisation interface and DB1 blocks only the putative dimerisation interface. However, our SAXS data revealed monomers of PvDBP-RII complexes in solution, even at high concentrations, leading us to question whether PvDBP-RII dimerisation is physiological and whether blocking of dimerisation is a mechanism of inhibition (Supplementary Figs. 4 and 5). Studies in which full-length PvDBP is assessed in the context of parasite invasion will be required to resolve whether dimerisation occurs in vivo.

Our study, therefore presents a complete view of the interaction between DARC and PvDBP-RII and reveals the role of DARC sulfation on PvDBP binding. However, our view of the complete PvDBP protein remains incomplete. Studies of PvDBP have long focused on the RII domain, through ease of production and due to the demonstration that this domain contains the DARC binding site. However, questions remain, which will only be answered when we understand the structure of full-length PvDBP and how it binds to full-length DARC. For example, it is also possible that the putative dimerisation interface observed in PvDBP-RII has other roles in full-length PvDBP, which are affected by antibody binding. Further study is also required to understand the role of the DARC Fy^a/Fy^b polymorphism on PvDBP binding as the role of the polymorphic residue 42 is not certain from this study. Nevertheless, we now reveal, for the first time, the full interaction between the ectodomain of DARC and PvDBP-RII, showing the details of this interaction critical for red blood cell invasion in *P. vivax*. This identifies new surfaces of PvDBP which could be targeted by small molecules or antibodies that block the interaction with DARC, or that can be included in vaccine immunogens. These findings will guide future design of therapeutic agents to target the scourge of malaria.

Methods

Protein expression and purification

For crystallisation, the coding sequence for PvDBP-RII (D194-S508) was cloned into the pHLsec vector, which included a C-terminal 6xHis tag. DARC ectodomain (M1-S60, numbered as for the mature polypeptide) was cloned into the pENTR4 vector containing a C-terminal TEV protease cleavage site and 6xHis tag. Residues C4, C51 and C54 were changed to A to avoid disulfide formation and were not part of the PvDBP-RII binding site. DB1 heavy and light chains were supplied in the AbVec-hIgG1 and AbVec-hIgKappa vectors respectively. The constructs were transfected into Expi293 cells (Thermo Fisher Scientific) using the ExpiFectamin 293 Transfection Kit (Gibco). The transfected cells were collected after 4 days and the supernatant clarified by centrifugation at 10,000 × *g* and supplemented with 150 mM NaCl and 20 mM imidazole (final concentration).

https://doi.org/10.1038/s41467-023-39357-w

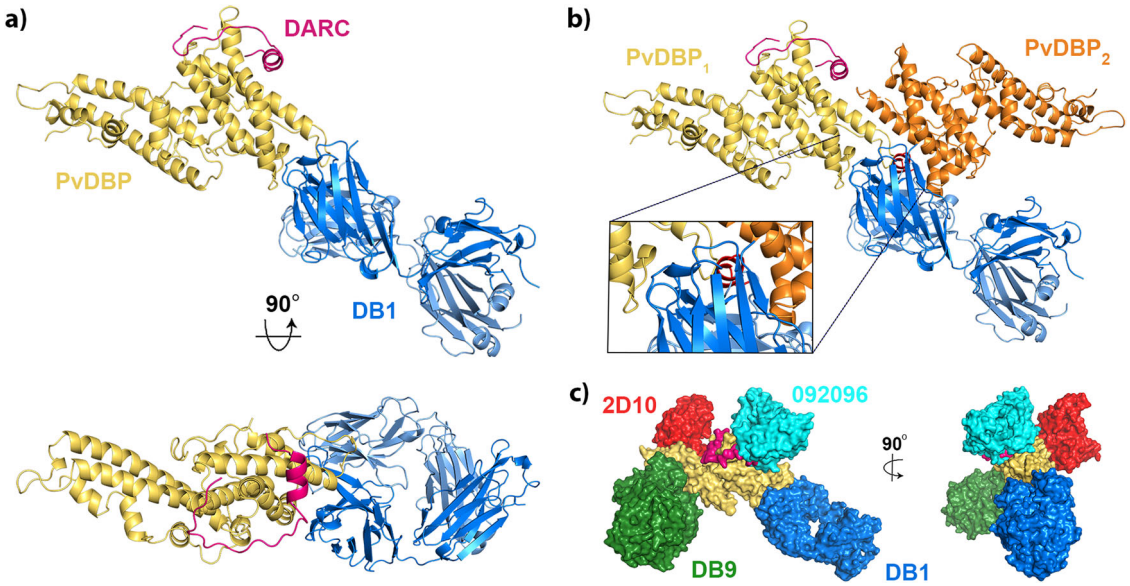


Fig. 4 | Structural basis for neutralising antibody binding to PvDBP. a Structure of PvDBP-RII (yellow) bound to DARC (pink) and antibody DB1. **b** A model of the putative PvDBP dimer (yellow and orange) bound to DARC and DB1, derived from PDB: 4NUV, showing that DB1 clashes with the putative dimerisation interface. **c** A composite structure in which four different neutralising antibodies, DB1 (blue), DB9 (green)[13], 2D10 (red)[32] and 092096 (cyan)[20] are docked on to the structure of PvDBP-RII (yellow) and DARC (pink).

For isothermal titration calorimetry, PvDBP-RII (D194-T521) was expressed with an N-terminal monomeric Fc domain (mono-Fc). Three alanine substitution mutations, T257A, S353A and T422A, were introduced to the PvDBP-RII sequence to remove potential N-linked glycosylation sites. The construct was transfected into Expi293 cells as above and harvested after 5 days. The mono-Fc tagged PvDBP-RII was purified using HiTrap Protein A HP column (Cytiva).

DB1 antibody was purified by binding clarified cell supernatant to Protein A Agarose beads (Thermo Fisher Scientific), with elution into 100 mM glycine pH 3.0 and was quickly neutralised with 1 M Tris-HCl pH 8.0. To generate DB1 Fab fragments, purified DB1 was buffer exchanged into 150 mM NaCl, 20 mM HEPES pH 7.5, 20 mM L-cysteine for overnight cleavage at 37 °C with immobilised papain, followed by protein-A agarose purification to remove Fc and uncleaved DB1 antibody.

PvDBP-RII and DARC ectodomain used for crystallisation was purified by immobilised-metal-affinity chromatography by batch binding clarified cell supernatant to Ni-NTA resin, with elution into 300 mM imidazole, 150 mM NaCl, 20 mM HEPES pH 8.0 and was buffer exchanged into 150 mM NaCl, 20 mM HEPES pH 7.5 using a Vivaspin 3 kDa column. TEV protease and PNGase F were added to DARC ectodomain overnight at 4 °C to remove the TEV-6xHis tag and glycans, respectively.

Crystallisation, data collection, and structure determination

Crystallisation trials were carried out using vapour diffusion in sitting drops, mixing 100 nl protein complex solution and 100 nl well solution. Crystals were visible from day 12 with a well solution of 15% w/v jeffamine D-2003 and 10% v/v ethanol. The crystals were cryoprotected by transfer into drops of well solution supplemented with 25% glycerol and were cryo-cooled in liquid nitrogen for data collection.

Data was collected at the I24 beamline at Diamond at a wavelength of 0.9999 Å. After processing by autoPROC, the dataset was scaled using AIMLESS (v.0.73)[36], producing a complete dataset at a resolution of 2.5 Å. The structure was solved by molecular replacement with Phaser MR (v.2.8.3)[37] using a search model for PvDBP-RII (PDB: 6R2S). The model was built and refined using cycles of COOT (v.0.8.9.2)[38] and BUSTER (v.2.10)[39].

Molecular dynamics simulations

All simulations performed using OpenMM v7.7[40]. Models of PvDBP (residues 217-509) and DARC (residues 19-47) were capped at C- and N-termini using N-methyl and acetyl groups, respectively, protonated at pH of 7.5 using H++[41] and were soaked in truncated octahedral water boxes with a padding distance of 1 nm. NaCl was added to 150 mM while neutralising charges as described in[42]. Systems were parameterised using the Amber14-SB force field[43] and water modelled using TIP3P-FB[44] and tleap[45]. For sulfotyrosine residues, we used parameters from[46]. Three systems were prepared from our crystal structure, with both Y30 and Y41 sulfated, with only Y30 sulfated and with only Y41 sulfated.

Non-bonded interactions were calculated using the particle mesh Ewald method[47] with a real-space cut-off of 0.9 nm and error tolerance of 0.0005. Water molecules and heavy atom-hydrogen bonds were rigidified using SETTLE[48] and SHAKE[49] algorithms, respectively. Hydrogen mass repartitioning[50] was used to allow 4 fs time steps. Simulations were run using the mixed-precision CUDA platform in OpenMM using the Middle Langevin Integrator with a friction coefficient of 1 ps^{-1} and the Monte-Carlo Barostat at a pressure of 1 atm. We equilibrated systems using a multi-step protocol: (i) energy minimisation over 10,000 steps, (ii) heating of the NVT ensembles from 100 K to 300 K over 200 ps, (iii) 200 ps simulation of the NPT ensembles at 300 K, (iii) cooling of the NVT ensembles from 300 K to 100 K over 200 ps, (iv) energy minimisation over 10,000 steps, (v) heating of the NVT ensembles from 100 K to 300 K over 200 ps, and (vi) 5 ns simulation of the NPT ensembles at 300 K.

We initialised well-tempered metadynamics using Plumed[51]. The first pair of simulations was biased on the χ_1 and χ_2 angles of DARC position 30. These were initialised with structures with both Y30 and Y41 sulfated and with just Y41 sulfated, using an initial hill height of 1.2 kJ/mol, a bias factor of 10, and bias widths of 0.07 rad, with biases being deposited every 500 steps. The second pair of simulations was biased on the χ_1 and χ_2 angles of DARC position 41. These were initialised with structures with both Y30 and Y41 sulfated and with just Y30 sulfated, using an initial hill height of 3.0 kJ/mol, a bias factor of 10, and bias widths of 0.07 rad, with biases being deposited every 500 steps. We found that a lower hill height of 1.2 kJ/mol was

insufficient to overcome free energy barriers in the system. For the second pair of simulations, we applied harmonic restraints on the backbone positions of the DARC peptide using the crystal structure as the reference and a force constant of 1500 kJ/(mol nm^2). Production simulations were performed for 500 ns and convergence was assessed using block error analysis on the computed free energy surfaces with respect to χ_1 and χ_2 angles of the biased positions. To recover unbiased ensembles from our simulations, we calculated weights (w) for each configuration (s) using the time-independent reweighting scheme (Eq. (1))[52]:

$$w(s) \propto \exp\left(\frac{V(s)}{k_B T}\right) \quad (1)$$

where $V(s)$ is the configuration-dependent bias at the end of the simulation and $k_B T$ is the product of the Boltzmann constant and temperature. We used MDTraj (v. 1.9.6)[53] to analyse trajectories.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed on a MicroCal PEAQ-ITC (Malvern). PvDBP-RII used in the ITC experiments was tagged N-terminally with a monomeric Fc domain (mono-Fc) and contained T257A, S353A and T422A substitutions to remove N-linked glycosylation sites; modifications are not in close proximity to the DARC binding site and hence should not interfere with binding. All synthetic peptides used in the ITC experiments were purchased from Cambridge Research Biochemicals (Billingham, UK) with the purity level of 95%. Prior to the measurement, mono-Fc tagged PvDBP-RII was extensively dialysed against 1x phosphate-buffered saline (PBS), and the peptides were dissolved in the same dialysis buffer to prevent buffer mismatch. The peptides were placed in the syringe at a concentration of ~130 μM for Ac-QLDFEDVWNSS-[Y(SO3H)]-GVNDSFPDG D-[Y(SO3H)]-DANLEL-NH_2, ~300 μM for Ac-QLDFEDVWNSSYGVNDSFP DGDYDANLEL-NH2, ~200 μM for Ac-QLDFEDVWNSS-[Y(SO3H)]-GVND SFPDGDYDANLEL-NH_2, and ~260 μM for Ac-QLDFEDVWNSSYGVNDS FPDGD-[Y(SO3H)]-DANLEL-NH_2, whereas the concentration of mono-Fc tagged PvDBP-RII in the cell was ~20 μM for all experiments. The peptide concentrations were determined by BCA assay using the Pierce Rapid Gold BCA Protein Assay Kit (ThermoFisher Scientific), and the protein concentration was determined by UV absorbance at 280 nm. The titrations were all performed with peptides in the syringe and mono-Fc tagged PvDBP-RII in the cell and consisted of a single 0.4 μl injection followed by 18 injections of 2 μl with injection duration of 4 s, injection spacing of 150 s, stir speed of 750 rpm, and reference power of 10 μcal/s. Experiments were conducted at 25°C in triplicate ($n = 3$), and the data are reported as the arithmetic mean ± SD. Fitting of the integrated titration peaks was performed with MicroCal PEAQ-ITC Analysis Software (Malvern) provided with the instrument, then all data were exported to and plotted in Prism 9 (Dotmatics).

Small angle X-ray scattering

Mono-Fc tagged PvDBP-RII used for SAXS experiments was expressed in Expi293 cells (Thermo Scientific) as described above and purified using the CaptureSelect C-tagXL Affinity Matrix (Thermo Scientific). The elution was concentrated and buffer-exchanged into 20 mM HEPES, 150 mM NaCl pH 7.5 using a PD-10 Desalting Column (Cytiva) before TEV cleavage (1:100 TEV to protein ratio) overnight at 4 °C. After cleavage, mono-Fc was removed using Pierce Protein A agarose beads (Thermo Scientific). The flow-through containing PvDBP-RII was then mixed with DB1 or DB9 Fab (1.5:1 Fab to protein ratio) for 1 h before concentration and size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva). The fractions containing PvDBP-RII with Fabs were concentrated or were mixed with $DARC_{19\text{-}47}$ Y30-S/Y41-S double sulfated peptides in case of the PvDBP-RII-DB9 Fab complex for 1 h prior to concentration.

Small-angle X-ray scattering (SAXS) data were collected at beamline B21 at the Diamond Light Source (Didcot, UK). Scattering was detected using X-rays at a wavelength of 0.99 Å and an Eiger 4 M detector with a detector-sample distance of 4.014 m. For the PvDBP-RII:DB9 Fab complexes with or without Y30-S/Y41-S double sulfated peptides, experiments were performed via two methods: (1) batch mode where 35 μl of samples at 6, 3, 1 mg/ml concentrations with corresponding matching buffers were injected directly into the capillary, using the EMBL Arinax sample handling robot, for data collection, and (2) SEC-SAXS where 50 μl of 6 mg/ml samples were gel filtered on the Shodex PROTEIN KW403-4F column (Shodex Denko Europe) equilibrated with 20 mM HEPES, 150 mM NaCl, pH 7.5 prior to data collection. The data for PvDBP-RII:DB1 complex were only collected using the SEC-SAXS method at the starting concentration of 6 mg/ml.

The data were processed using ScÅtter[54] with the ATSAS software suites[55]. For SEC-SAXS, buffer frames were averaged and subtracted from averaged frames of the complex peak fractions. The radius of gyration (Rg), distance distribution function P(r) and the maximum particle diameter (Dmax) were determined using ScÅtter. Twenty-nine ab initio initial bead models were generated with DAMMIF[55] from the ATSAS software suites in ScÅtter and averaged with DAMAVER[56] followed by refinement against the original data using DAMMIN[57]. The 20 Å envelope of the refined bead model was generated in Chimera[58]. Crystal structures of PvDBP-RII in complex with DB9 Fab[13] (PDB: 6R2S) overlayed with the Y41-S DARC peptide structure obtained in this study and the crystal structure of PvDBP-RII with DB1 Fab from the current study were fitted into the envelopes using Chimera. CRYSOL[55] was used to derive theoretical scattering data from monomeric and dimeric version of PvDBP-RII:DB9 or PvDBP-RII:DB1 complexes and to compare these data with the experimental results.

Growth inhibition assays with transgenic *P. knowlesi*

To investigate DBP-DARC interactions in vitro, the variably sulfated DARC ectodomain peptides described above were used to competitively inhibit the host-parasite interaction between erythrocytic DARC and parasite-expressed DBP, which is essential for invasion and subsequent replication. Growth inhibition assays were run as previously described in refs. 13,14 with parasite growth measured via lactate dehydrogenase (LDH) assay after one intraerythrocytic cycle (approx. 27 h) in the presence of various peptides, prepared in twofold dilution curves. Each 96-well plate included internal controls (infected, untreated erythrocytes and uninfected erythrocytes) which were used to calculate the percentage growth inhibition in each treatment well using the following formula (Eq. (2))[13,14]:

$$\%\ \mathrm{GIA} = 100 - \left(\left(\frac{\mathrm{OD}_{\mathrm{sample}} - \mathrm{OD}_{\mathrm{uninfected\ RBCs}}}{\mathrm{OD}_{\mathrm{infected\ RBCs}} - \mathrm{OD}_{\mathrm{uninfected\ RBCs}}}\right) \times 100\right) \quad (2)$$

Assays were conducted with a minimum of 2 technical replicates, and data presented represents averages from four separate experiments ($n = 2$ each using Fy^a and Fy^b donor red blood cells). Assays were performed using both *P. knowlesi* wild-type A1-H.1 (expressing PkDBPα), as well as transgenic *P. knowlesi* $PvDBP^{OR}$/$PkDBP\beta\gamma^{KO}$, which expresses the *P. vivax* DBP as an orthologue replacement of the homologous PkDBPα. IC_{50} values were identified using a variable slope four-parameter logistic curve, calculated using GraphPad Prism v9.4.1.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Coordinates and structure factors generated in this study have been deposited in the Protein Data Bank with accession code 8A44. Source data are provided with this paper.

References

1. Battle, K. E. et al. Mapping the global endemicity and clinical burden of *Plasmodium vivax*, 2000-17: a spatial and temporal modelling study. *Lancet* **394**, 332–343 (2019).
2. Howes, R. E. et al. *Plasmodium vivax* transmission in Africa. *PLoS Negl. Trop. Dis.* **9**, e0004222 (2015).
3. Price, R. N., Douglas, N. M. & Anstey, N. M. New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. *Curr. Opin. Infect. Dis.* **22**, 430–435 (2009).
4. Horuk, R. et al. A receptor for the malarial parasite *Plasmodium vivax*: the erythrocyte chemokine receptor. *Science* **261**, 1182–1184 (1993).
5. Miller, L. H., Mason, S. J., Dvorak, J. A., McGinniss, M. H. & Rothman, I. K. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* **189**, 561–563 (1975).
6. Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N. Engl. J. Med.* **295**, 302–304 (1976).
7. World Health Organisation. *World Malaria Report 2019* (WHO, 2019).
8. Popovici, J., Roesch, C. & Rougeron, V. The enigmatic mechanisms by which *Plasmodium vivax* infects Duffy-negative individuals. *PLoS Pathog.* **16**, e1008258 (2020).
9. Menard, D. et al. *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl Acad. Sci. USA* **107**, 5967–5971 (2010).
10. Twohig, K. A. et al. Growing evidence of *Plasmodium vivax* across malaria-endemic Africa. *PLoS Negl. Trop. Dis.* **13**, e0007140 (2019).
11. Adams, J. H. et al. The Duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. *Cell* **63**, 141–153 (1990).
12. Singh, A. P. et al. Targeted deletion of *Plasmodium knowlesi* Duffy binding protein confirms its role in junction formation during invasion. *Mol. Microbiol.* **55**, 1925–1934 (2005).
13. Rawlinson, T. A. et al. Structural basis for inhibition of *Plasmodium vivax* invasion by a broadly neutralizing vaccine-induced human antibody. *Nat. Microbiol.* **4**, 1497–1507 (2019).
14. Mohring, F. et al. Rapid and iterative genome editing in the malaria parasite *Plasmodium knowlesi* provides new tools for *P. vivax* research. *Elife* **8**, e45829 (2019).
15. de Cassan, S. C. et al. Preclinical assessment of viral vectored and protein vaccines targeting the Duffy-binding protein region II of *Plasmodium vivax*. *Front. Immunol.* **6**, 348 (2015).
16. Moreno, A. et al. Preclinical assessment of the receptor-binding domain of *Plasmodium vivax* Duffy-binding protein as a vaccine candidate in rhesus macaques. *Vaccine* **26**, 4338–4344 (2008).
17. Cole-Tobian, J. L. et al. Strain-specific duffy binding protein antibodies correlate with protection against infection with homologous compared to heterologous *Plasmodium vivax* strains in Papua New Guinean children. *Infect. Immun.* **77**, 4009–4017 (2009).
18. King, C. L. et al. Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage *Plasmodium vivax* infection. *Proc. Natl Acad. Sci. USA* **105**, 8363–8368 (2008).
19. Nicolete, V. C., Frischmann, S., Barbosa, S., King, C. L. & Ferreira, M. U. Naturally acquired binding-inhibitory antibodies to *Plasmodium vivax* Duffy binding protein and clinical immunity to malaria in rural Amazonians. *J. Infect. Dis.* **214**, 1539–1546 (2016).
20. Urusova, D. et al. Structural basis for neutralization of *Plasmodium vivax* by naturally acquired human antibodies that target DBP. *Nat. Microbiol.* **4**, 1486–1496 (2019).
21. Chitnis, C. E. & Miller, L. H. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. *J. Exp. Med.* **180**, 497–506 (1994).
22. Chitnis, C. E. & Sharma, A. Targeting the *Plasmodium vivax* Duffy-binding protein. *Trends Parasitol.* **24**, 29–34 (2008).
23. Chitnis, C. E., Chaudhuri, A., Horuk, R., Pogo, A. O. & Miller, L. H. The domain on the Duffy blood group antigen for binding *Plasmodium vivax* and *P. knowlesi* malarial parasites to erythrocytes. *J. Exp. Med.* **184**, 1531–1536 (1996).
24. Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E. & Sharma, A. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* **439**, 741–744 (2006).
25. Batchelor, J. D. et al. Red blood cell invasion by *Plasmodium vivax*: structural basis for DBP engagement of DARC. *PLoS Pathog.* **10**, e1003869 (2014).
26. Choe, H. et al. Sulphated tyrosines mediate association of chemokines and *Plasmodium vivax* Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). *Mol. Microbiol.* **55**, 1413–1422 (2005).
27. Batchelor, J. D., Zahm, J. A. & Tolia, N. H. Dimerization of *Plasmodium vivax* DBP is induced upon receptor binding and drives recognition of DARC. *Nat. Struct. Mol. Biol.* **18**, 908–914 (2011).
28. Yogavel, M., Chhibber-Goel, J., Jamwal, A., Gupta, S. & Sharma, A. Engagement rules that underpin DBL-DARC interactions for ingress of *Plasmodium knowlesi* and *Plasmodium vivax* into human erythrocytes. *Front. Mol. Biosci.* **5**, 78 (2018).
29. Hans, D. et al. Mapping binding residues in the *Plasmodium vivax* domain that binds Duffy antigen during red cell invasion. *Mol. Microbiol.* **55**, 1423–1434 (2005).
30. VanBuskirk, K. M., Sevova, E. & Adams, J. H. Conserved residues in the *Plasmodium vivax* Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. *Proc. Natl Acad. Sci. USA* **101**, 15754–15759 (2004).
31. King, C. L. et al. Fy(a)/Fy(b) antigen polymorphism in human erythrocyte Duffy antigen affects susceptibility to *Plasmodium vivax* malaria. *Proc. Natl Acad. Sci. USA* **108**, 20113–20118 (2011).
32. Chen, E. et al. Broadly neutralizing epitopes in the *Plasmodium vivax* vaccine candidate Duffy Binding Protein. *Proc. Natl Acad. Sci. USA* **113**, 6277–6282 (2016).
33. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
34. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
35. Chen, E., Salinas, N. D., Ntumngia, F. B., Adams, J. H. & Tolia, N. H. Structural analysis of the synthetic Duffy Binding Protein (DBP) antigen DEKnull relevant for *Plasmodium vivax* malaria vaccine design. *PLoS Negl. Trop. Dis.* **9**, e0003644 (2015).
36. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
37. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
39. Bricogne, G. et al. 'BUSTER version 2.10.4' (Global Phasing Ltd., 2017).
40. Eastman, P. et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
41. 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.* **40**, W537–W541 (2012).

42. Machado, M. R. & Pantano, S. Split the charge difference in two! a rule of thumb for adding proper amounts of ions in MD simulations. *J. Chem. Theory Comput.* **16**, 1367–1372 (2020).
43. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
44. Eyal, N. F. S. & Karpus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921 (1996).
45. Case, D. A. et al. Amber 2022 (University of California, 2022).
46. Takefumi, Y., Okajima, R., Miyanabe, K. & Tsumoto, K. Modified AMBER force-field (FUJI) parameters for sulfated and phosphorylated tyrosine residues: development and application to CCR5-derived peptide systems. *AIP Conf. Proc.* **2186**, 030013 (2019).
47. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
48. Miyamoto, S. K. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
49. Ryckaert, J., 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.* **23**, 327–341 (1977).
50. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
51. consortium, P. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).
52. Branduardi, D., Bussi, G. & Parrinello, M. Metadynamics with adaptive Gaussians. *J. Chem. Theory Comput.* **8**, 2247–2254 (2012).
53. McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
54. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).
55. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
56. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
57. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
58. Pettersen, E. F. et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

M.K.H. is a Wellcome Trust Investigator (220797/Z/20/Z). R.M. and T.P. are funded by the Skaggs-Oxford graduate scholarship. B.G. is funded by the Wellcome PhD programme in Cellular Structural Biology. S.M.D. is supported by a UK Medical Research Council LID PhD studentship. R.W.M. was supported by the UK Medical Research Council (MRC Career Development Award MR/M021157/1). The authors thank Simon Draper for the coding sequence for antibody DB1 and the plasmid construct for monoFc-tagged PvDBP-RII. We thank the staff at Diamond beamline I04 for help with crystallographic data collection and staff at Diamond beamline B21 for help with SAXS data collection.

Author contributions

R.M. produced proteins and R.M., E.L. and M.K.H. determined the crystal structure. T.P. conducted ITC and SAXS experiments. B.G. performed molecular dynamics analysis. S.M.D., F.M. and R.W.M. conducted parasite growth-inhibition experiments. R.W.M. and M.K.H. devised the study. M.K.H. drafted the manuscript and all authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-39357-w.

Correspondence and requests for materials should be addressed to Matthew K. Higgins.

Peer review information *Nature Communications* thanks Ram Kothandan, Wai-Hong Tham and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

nature communications

Article

https://doi.org/10.1038/s41467-022-33379-6

Structure of the malaria vaccine candidate Pfs48/45 and its recognition by transmission blocking antibodies

Received: 24 May 2022

Accepted: 14 September 2022

Published online: 24 September 2022

Check for updates

Kuang-Ting Ko[1,2,6], Frank Lennartz[1,6], David Mekhaiel[3,6], Bora Guloglu[1,2], Arianna Marini[3], Danielle J. Deuker[3], Carole A. Long[4], Matthijs M. Jore[5], Kazutoyo Miura[4], Sumi Biswas[3] ✉ & Matthew K. Higgins[1,2] ✉

An effective malaria vaccine remains a global health priority and vaccine immunogens which prevent transmission of the parasite will have important roles in multi-component vaccines. One of the most promising candidates for inclusion in a transmission-blocking malaria vaccine is the gamete surface protein Pfs48/45, which is essential for development of the parasite in the mosquito midgut. Indeed, antibodies which bind Pfs48/45 can prevent transmission if ingested with the parasite as part of the mosquito bloodmeal. Here we present the structure of full-length Pfs48/45, showing its three domains to form a dynamic, planar, triangular arrangement. We reveal where transmission-blocking and non-blocking antibodies bind on Pfs48/45. Finally, we demonstrate that antibodies which bind across this molecule can be transmission-blocking. These studies will guide the development of future Pfs48/45-based vaccine immunogens.

The recent approval of RTS,S as the first vaccine to prevent malaria caused by *Plasmodium falciparum* is an important step in efforts to end the devastating effects of this ancient disease[1], with promising results also recently released for the related R21 vaccine[2]. However, with large-scale trials of RTS,S indicating that it is around 30% effective at preventing severe disease[3], the quest for a more effective vaccine continues. The *Plasmodium* life cycle is complex and there are multiple stages at which such a vaccine could act[4], including stopping fusion of the male and female gametes of the parasite within the mosquito midgut. If the bloodmeal of a mosquito contains antibodies which prevent gamete fusion, then these can stop completion of the parasite life cycle and can block transmission. Vaccine immunogens which induce the production of such transmission-blocking antibodies will therefore be important components of multi-stage-targeting malaria vaccines[5].

One of the most promising transmission-blocking vaccine candidates is the gamete surface protein, Pfs48/45. This is essential for gamete fusion, as male *Plasmodium berghei* gametes that lack the homologous protein, Pbs48/45, are unable to penetrate female gametes to form zygotes[6]. In addition, antibodies induced by immunising animals with Pfs48/45 block the sexual development of the parasite within infected mosquitos[7–12]. Pfs48/45 is expressed on gametocytes found in human blood and is therefore exposed to the human immune system. Indeed, antibodies that target Pfs48/45 are also found in sera from individuals from malaria-endemic regions and the presence of such antibodies correlates with the transmission-blocking activity of these sera[13–17]. This means that individuals immunized with Pfs48/45 could also experience immune boosting through natural low-level infection. Also encouraging is the low sequence diversity of Pfs48/45 across strains of *Plasmodium falciparum*[12,16,18–20].

[1]Department of Biochemistry, South Parks Road, University of Oxford, Oxford OX1 3QU, UK. [2]Kavli Institute for Nanoscience Discovery, Dorothy Crowfoot Hodgkin Building, University of Oxford, South Parks Rd, Oxford OX1 3QU, UK. [3]Jenner Institute, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK. [4]Malaria Immunology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, NIH, Rockville, MD, USA. [5]Department of Medical Microbiology, Radboud University Medical Centre, Nijmegen, The Netherlands. [6]These authors contributed equally: Kuang-Ting Ko, Frank Lennartz, David Mekhaiel. ✉e-mail: sumi.biswas@ndm.ox.ac.uk; matthew.higgins@bioch.ox.ac.uk

These factors combine to suggest that a vaccine immunogen based on Pfs48/45 will target a conserved and essential feature of the parasite life cycle and prevent further transmission from a malaria-infected individual.

As full-length Pfs48/45 has proved challenging to produce, vaccine immunogen design has so far been guided by an understanding of the molecular architecture of Pfs48/45 and by the identification of epitopes for potent monoclonal antibodies[5,21]. Pfs48/45 is formed from three domains, with N- and C-terminal 6-cys domains joined by a central 4-cys domain, all linked to the gamete membrane through a C-terminal GPI anchor[22–24]. While these domain boundaries are well understood, how they come together to form full-length Pfs48/45 is still not known. Monoclonal antibodies have been generated against Pfs48/45, have been classified into different competition groups, and have been assessed for their transmission-blocking activity[8,12,25–28]. The most potent of these antibodies, 32F3[8] and 85RF45.1[25] bind to the C-terminal domain and the structure of the C-terminal domain bound to the Fab fragment of 85RF45.1 has been determined[12,29]. This domain has therefore been the focus of most vaccine immunogen design approaches targeting Pfs48/45, primarily administered as part of fusion proteins, such as that with GLURP[21]. However, other transmission-blocking antibodies have been identified which target the N-terminal and central domains of Pfs48/45[12,25,28]. Here, we therefore present the structure of the full Pfs48/45 ectodomain, show how antibodies recognise different regions of Pfs48/45 and demonstrated that a significant fraction of the transmission-blocking activity of sera targeting Pfs48/45 is due to antibodies which target its N-terminal and central domains.

Results

A structural comparison of two antibodies targeting domain 3 of Pfs48/45

The two antibodies which have the highest transmission-blocking activity, 32F3 and 85RF45.1, both bind to the C-terminal domain of Pfs48/45[8,12]. However, these two antibodies show substantial differences in transmission-blocking activity, with 85RF45.1 completely preventing transmission at 14 µg/ml, while 32F3 is only ~40% effective at 70 µg/ml[12]. To rationalise these differences, and to guide future vaccine design, we determined the structure of the C-terminal 6-cys domain of Pfs48/45 bound to 32F3 (Fig. 1A, Supplementary Table 1). We prepared Fab fragments from 32F3, combined these with the Pfs48/45 C-terminal domain and conducted crystallisation trials. Crystals formed and a complete dataset was collected to 1.9 Å resolution, allowing structure determination by molecular replacement.

The structure of the C-terminal domain of Pfs48/45 was largely unchanged in conformation from that observed in complex with 85RF45.1, with a root mean square deviation of 0.66 Å (Fig. 1A, B)[12,29]. The most substantial change was in the loop comprising residues 357–369. This loop is disordered when bound to antibody 85RF45.1 but becomes ordered through interactions with 32F3 (Fig. 1A). While antibodies 32F3 and 85RF45.1 bind to overlapping epitopes, many of the contacts are not shared and the Fabs approach Pfs48/45 at different angles.

We also determined the binding kinetics for both 85RF45.1 and 32F3 to Pfs48/45, using surface plasmon resonance measurements (Fig. 1C, Supplementary Fig. 1). We first captured 85RF45.1 and 32F3 on different flow paths of a chip coated with protein A/G and flowed two-fold dilution series of Pfs48/45 over these antibody-coupled surfaces. Both antibodies bound to Pfs48/45 with similar affinities in the low nanomolar range (2.5 nM for 85RF45.1 and 5.2 nM for 32F3). However, binding kinetics differed, with 85RF45.1 showing faster association- and dissociation-rates, while 32F3 binds more slowly but forms a more stable complex. It seems likely that the slower binding kinetics of 32F3 may be due to the need for loop 357–369 to become ordered on binding, while antibody 85RF45.1 has an epitope which is unchanged in structure on antibody binding, allowing faster binding kinetics.

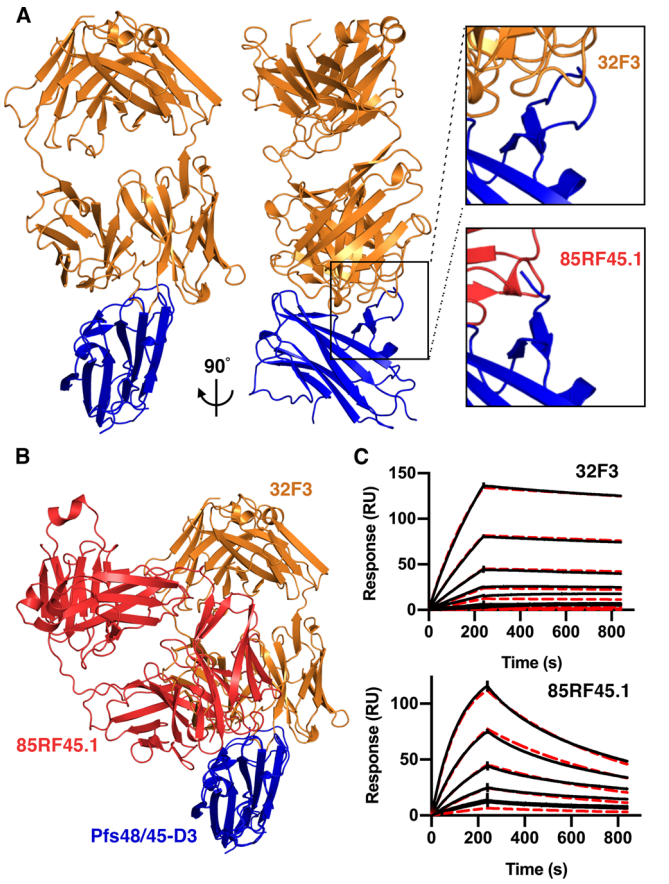


Fig. 1 | Comparison of the epitopes of transmission-blocking antibodies 85RF45.1 and 32F3. A The structure of the C-terminal domain of Pfs48/45 (Pfs48/45-D3; blue) bound to antibody 32F3 (orange). The upper right inset shows a close-up of a Pfs48/45 loop which becomes ordered on 32F3 binding. The lower inset shows the equivalent view of the complex of Pfs48/45-D3 (blue) bound to 85RF45.1 (red). **B** An alignment of the structures of 32F3 (orange) and 85RF45.1 (red) bound to Pfs48/45-D3 (blue). **C** Surface plasmon resonance analysis of the binding of Pfs48/45-D3 to immobilised 32F3 and 85RF45.1. In both cases, the black lines show the responses due to a two-fold dilution series with a top concentration of 7.8 nM for 85RF45.1 and 125 nM for 32F3. The red dashed lines show fitting to a 1-to-1 binding model.

Determining the structure of full-length Pfs48/45 by combining crystallography with an AlphaFold2 model

We next aimed to determine the structure of the complete Pfs48/45 molecule, allowing us to reveal the architecture of this three-domain protein and the locations of antibody epitopes outside the C-terminal domain. Insect cell expression systems were available to produce full-length Pfs48/45 ectodomain, or to generate a protein containing the central and C-terminal domains, Pfs48/45-D2 + 3 [11]. Both were expressed without the GPI-anchor modification site and were purified and mixed in different combinations with one or more Fab fragments, selected from a set of antibodies which bind to different domains of Pfs48/45 (85RF45.1, 85RF45.3, 85RF45.5[25], 32F3[8], 10D8, 9D1, 7A6[12]). Three of these complexes generated crystals: full-length Pfs48/45 bound to 10D8; full-length Pfs48/45 bound to both 10D8 and 85RF45.1; and Pfs48/45-D2 + 3 bound to both 10D8 and 32F3. Datasets were collected to 4.2 Å, 3.72 Å and 3.69 Å, respectively.

We attempted to determine the structures of these complexes by molecular replacement, using the structures of the C-terminal domain bound to either 32F3 or 85RF45.1 Fab as search models. In each case,

https://doi.org/10.1038/s41467-022-33379-6

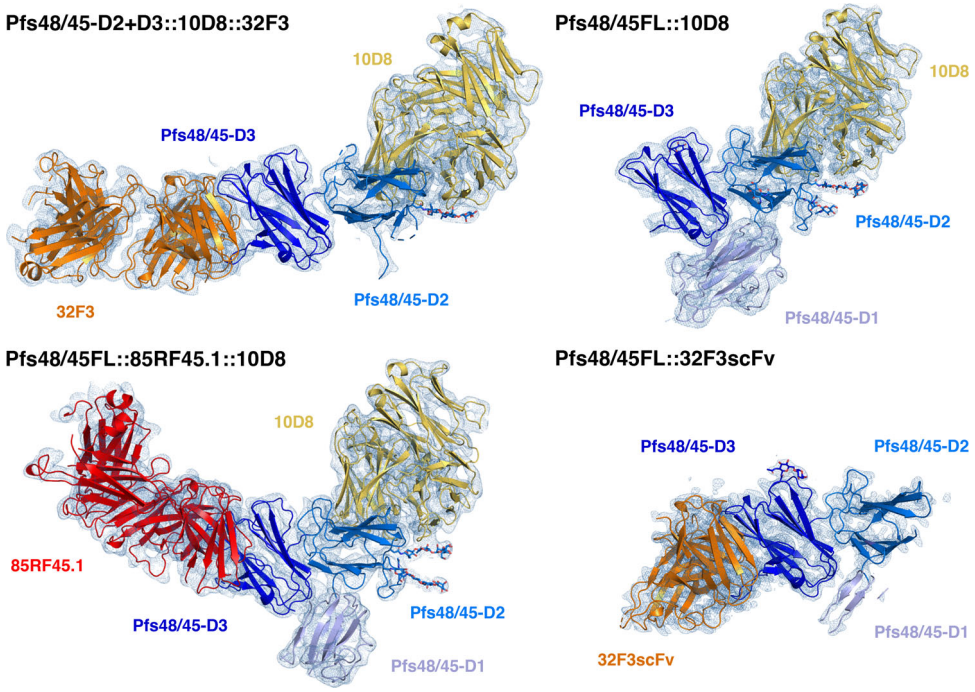


Fig. 2 | Structures of Pfs48/45 bound to antibodies 85RF45.1, 32F3 and 10D8. Structures of Pfs48/45 bound to different combinations of Fab fragments or scFvs. Pfs48/45 is shown with the three domains in different shades of blue. The N-terminal domain is light blue, the central domain is mid-blue and the C-terminal domain is dark blue. 10D8 is yellow, 85RF45.1 is red and 32F3 and its scFv are orange. Each figure also shows the electron density in pale blue, shown at a contour threshold of 1.0.

the resultant electron density maps were sufficiently detailed to allow us to build a model for the central domain of Pfs48/45. This domain contains the epitope for 10D8, also allowing us to build a model for the 10D8 Fab (Fig. 2, Supplementary Tables 1 and 2). However, while electron density could be observed for the N-terminal domain of Pfs48/45, this region was not sufficiently well resolved to allow a model to be built.

The hinge between the variable and constant domains of Fabs is flexible. In case this flexibility reduced the degree of order within the crystals, we also generated scFv constructs for 85RF45.1, 32F3 and 10D8. These were purified, mixed with full-length Pfs48/45 and complexes were subjected to crystallisation trials. Crystals formed for the complex containing 32F3 scFv and a dataset was collected to 2.13 Å resolution, allowing structure determination by molecular replacement. This higher resolution data allowed us to build an improved model for the central domain of Pfs48/45 (Fig. 2). However, the density for the N-terminal domain was still too poorly resolved to allow building of a complete molecular model of this domain. Analysis of packing within the different crystals suggested that, in each case, the N-terminal domain lies within a substantial cavity within the crystal lattice, and the lack of packing against other regions of Pfs48/45 contributed to disorder of this domain, even within a well-ordered lattice.

Recent improvements in protein structure prediction from AlphaFold2[30] provided a solution, allowing us to interpret the electron density maps for the N-terminal domain. AlphaFold2 correctly predicted the architecture of the central and C-terminal domains (with root mean square deviation of 1.26 Å for the central domain and 0.43 Å for the core 753/1044 atoms of the C-terminal domain), albeit with differences in loop structure (accession code Q8I6T1 at alphafold.ebi.ac.uk) (Supplementary Fig. 2). We therefore compared the AlphaFold2 model for the N-terminal domain with the electron density for this domain within the three different maps. The most complete electron density for this region of Pfs48/45 was obtained from crystals of full-length Pfs48/45 bound to 10D8. The AlphaFold2 model was therefore docked as single rigid body into this electron density and was rebuilt to fit the density. This generated a model for 122 residues of the 150 residue-long N-terminal domain, with a root mean square deviation of 1.63 Å from the AlphaFold2 model. The N-terminus and loops 62–68 and 163–168 were unresolved. This model was then used to guide building of the observed portions of the N-terminal domain in full-length Pfs48/45 bound to 10D8 and 85RF45.1 (95 residues were built) and Pfs48/45-D2 + 3 bound to both 10D8 and 32F3 (20 residues were built). Through this approach, we generated the first molecular models of full-length Pfs48/45 derived from experimental crystallographic data, with an AlphaFold2 model used to guide building of the N-terminal domain (Fig. 2, Supplementary Tables 1–3).

Pfs48/45 forms a dynamic three-domain disc-shaped architecture

We next used our structures, together with molecular dynamics simulations, to understand the organisation of Pfs48/45. Our three structures of Pfs48/45 all reveal a flattened, disc-like architecture (Fig. 3A), with a 20 residue long linker before the GPI anchor. This membrane-attachment site emerges from a flat surface of the molecule, suggesting that the disc may lie, on average, parallel to the membrane, with all three domains equally exposed at the gametocyte surface.

The availability of three structures of Pfs48/45 allowed us to assess its degree of flexibility within crystals. We docked the most complete model of the N-terminal domain, from the structure of full-length Pfs48/45 bound to 10D8, onto the fragments of the domain observed in the two other crystal forms and the three resultant models for the full Pfs48/45 ectodomain were aligned (Fig. 3B). While the relative positions of the central and C-terminal domains were largely unchanged across these molecules, the position of the N-terminal domain changed, due to flexibility in the linker joining the N-terminal and central domains. Through these movements, the separation between the N- and C-terminal domains varied.

https://doi.org/10.1038/s41467-022-33379-6

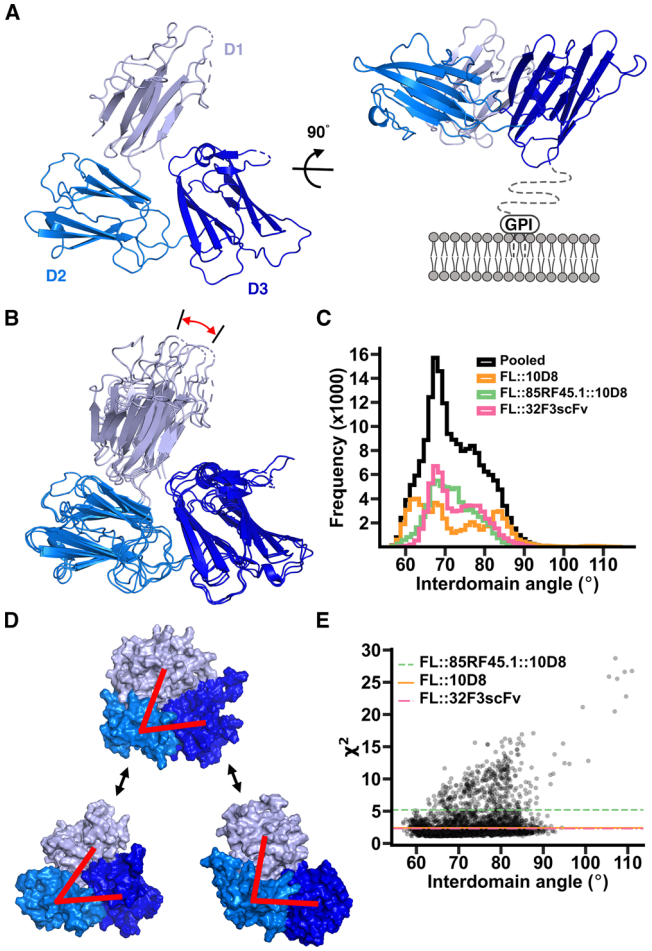


Fig. 3 | Structure and dynamics of Pfs48/45. A The structure of Pfs48/45 taken from that of full-length Pfs48/45 bound to 10D8, viewed from two different directions. The N-terminal (D1), central (D2) and C-terminal domains (D3) are light, medium and dark blue. The right-hand panel also indicates the membrane and the 20 residue long linker not included in our constructs. **B** Comparison of the three full-length Pfs48/45 structures. Three structures were constructed by taking the models of Pfs48/45 from the three different crystal structures of Pfs48/45 bound to different antibody combinations and aligning the structure of the N-terminal domain onto the fragments of the domain built into the electron density. These composite structures have been aligned based on the central and C-terminal domains, showing the motion of the N-terminal domain, highlighted by the red arrow. **C** Histogram showing the observed interdomain angles in full-length Pfs48/45 during atomistic molecular dynamics simulations. To define these angles two lines were drawn, linking the centres of mass of the N-terminal and central domains and those of the central and C-terminal domains. The angle shown is that between these lines. The three coloured histograms refer to outcomes of three independent simulations which started with the models in panel (**B**) with histograms being constructed by pooling five replicates for each starting structure. The black histogram is the pooled distribution of all replicates in all three starting models. **D** The average angle observed in the simulations from (**C**). is shown top-centre with the most closed observed model bottom-left and the most open model bottom-right. Red lines are as described in **C**. **E** shows the fitting of models from the molecular dynamic simulations to data from small angle X-ray scattering. Each point represents a different model and the χ^2 gives the quality of fit to the scattering curve. Horizontal lines show the χ^2 for the fit corresponding to interdomain angles found in the three crystal structures, as seen in **B**.

To further assess the degree of motion within Pfs48/45, we used molecular dynamics simulations. Three separate simulations were run, with our three independent structures of Pfs48/45, each modified to include the aligned N-terminal domain, used as three distinct starting points. In each case, we simulated the system for 500 ns with 5 repeats. Analysis of these simulations revealed motion of the N-terminal domain in excess of that seen in the three crystal structures, leading to further separation of the N- and C-terminal domains. The two peaks seen in some of the individual simulations were not observed in a pooled simulation, suggesting that each individual simulation had not converged, with pooled simulations reaching a more representative endpoint. We plotted two lines though the structure; one linking the centres of mass of the N-terminal and central domains and one linking the centres of mass of the central and C-terminal domains. The angle between these lines was used as a measure of the degree of opening of the complex. This varied from 56.4° to 114.9°, with an average of 72.8° and a standard deviation of 7.3° (Fig. 3C, D, Supplementary Fig. 3). The major peak varied from 56.4° to 92.1°, although we also observed a single instance of extreme opening to 112.7° (Supplementary Fig. 3A). This compares with angles of 60.3°, 63.1° and 67.5° in the three crystal structures.

To compare the fit of our molecular dynamic simulations to experimental data, we subjected Pfs48/45 to small angle X-ray scattering (Supplementary Fig. 4). We then sampled the molecular dynamic simulation trajectories to generate a PDB file for every 50th frame and fitted each of these to the scattering data using FoxS[31] thereby generating a χ^2 which allows us to quantify the fit. For comparison, we also fit the input models used for simulations, with domain positions matching those found in the three crystals. We then plotted the χ^2 against the interdomain angle (Fig. 3E). This analysis shows that no single angle fits most closely to the SAXS data, with a broad range of models, with interdomain angle 56° from 93°, all fitting with a lower χ^2 than any of the crystal structures. Indeed 66% of the models derived from simulation fitted the SAXS scattering data more closely than the three crystal structures. This confirms that Pfs48/45 in solution is a dynamic molecule, sampling a wide range of interdomain angles, through movement of the N-terminal domain relative to the central and C-terminal domains. This flexibility may allow Pfs48/45 to adopt different conformations on binding to partners, such as Pfs230 and will also leave all three domains exposed to antibody recognition.

Mapping transmission-reducing antibody epitopes to each domain of Pfs48/45

We next combined crystallography, electron microscopy and molecular dynamics simulations to assess how different monoclonal antibodies bind to Pfs48/45. In addition to visualising antibodies 32F3 and 85RF45.1 (Fig. 1), our crystal structures revealed the epitope for 10D8 (Fig. 2), an antibody with weak transmission-reducing potential which binds to the central domain of Pfs48/45[12]. We combined the structure of Pfs48/45 bound to 10D8 with those of the C-terminal domain bound to 32F3 and 85RF45.1 to generate a composite model, showing the structure of Pfs48/45 bound to these three antibodies (Fig. 4A).

To understand the degree of flexibility of these three epitopes we analysed the molecular dynamics simulations presented in Fig. 3, specifically assessing the motion of residues directly in contact with each of the three antibodies. This revealed that the epitope for 32F3 is significantly more flexible than those for 85RF45.1 and 10D8 (Fig. 4B), predominantly due to motion of the 357–369 loop, which becomes ordered on 32F3 binding due to direct interactions with the antibody (Supplementary Fig. 3). This indicates that the slower association-rate for 32F3, compared to the association rates for either 85RF45.1 or 10D8, is due to this flexibility and to the need for the epitope to adopt the correct structure on antibody binding. In contrast, off-rate is likely to be determined by the degree of complementarity of the epitope and paratope when forming an interaction. Indeed, 32F3 has the slowest off-rate (Supplementary Fig. 1), which correlates with a larger buried surface area for the epitope (1107 Å^2 for 32F3, 875 Å^2 for 85RF45.1 and 715 Å^2 for 10D8) and a greater number of direct interactions (Supplementary Fig. 3)[12].

https://doi.org/10.1038/s41467-022-33379-6

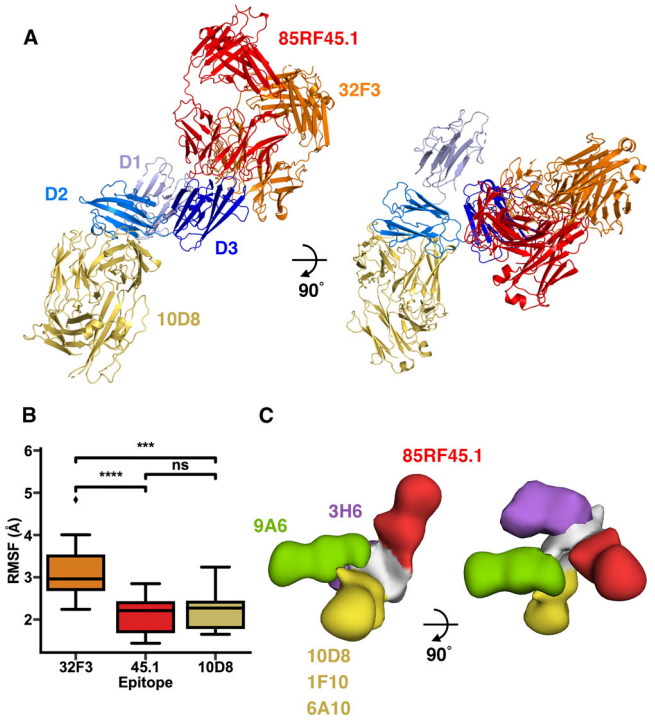


Fig. 4 | distribution of epitopes for transmission-blocking antibodies across Pfs48/45. A The structure of Pfs48/45 bound to 10D8 is used as a template, and combined with structures of the D3 domain of Pfs48/45 bound to either 85RF45.1 or 32F3 to generate a model, showing the relative locations of the three antibody epitopes. **B** A measure of the flexibility of the residues which form the epitopes for 32F3, 85RF45.1 and 10D8, as derived from molecular dynamics simulations. Each epitope was analysed in all 15 replicates and significance levels are based on a two-sided Mann–Whitney U tests where *** indicates $p < 0.0005$ and **** indicates $p < 0.00005$. The box bounds are interquartile ranges, and the lines within boxes are the median values. Whiskers extend to 1.5-fold the interquartile range. For 32F3 vs 85RF45.1 $p = 0.00008$; for 32F3 vs 10D8 $p = 0.00031$; for 10D8 vs 85RF45.1 $p = 0.58974$. **C** A composite model, derived from negative-stain electron microscopy structures. In each case, the complexes contained Pfs48/45 bound to 85RF45.1, together with either 9A6, 3H6, 10D8, 1F10 and 6A10. These were imaged by negative stain electron microscopy, and models were fitting into the resultant envelopes. These models were used to align the envelopes, allowing us to derive a composite model, showing the location for the five epitopes.

Finally, we used negative stain electron microscopy to visualise the approximate location on Pfs48/45 of the epitopes of four more antibodies. 1F10 and 6A10 are in the same competition group on the central domain as 10D8 while 9A6 is in a different competition group, binding to the same domain[12]. In contrast, 3H6 binds to the N-terminal domain. While 1F10 and 6A10 show some transmission-reducing activity, 9A6 and 3H6 do not[12]. In each case, we assembled complexes containing Pfs48/45, 85RF45.1 Fab and one other antibody Fab, with 85RF45.1 included to provide a clear marker which could be used to align and position each additional antibody. We then used negative stain electron microscopy to image the complexes and single particle analysis to determine a low-resolution structure. The Pfs48/45:85RF45.1 complex and an additional Fab model were then docked into these structures, taking into account to which domain the additional Fab bound to determine the organisation of the complex and the approximate location of the Fab. These five models were then aligned on Pfs48/45 and assembled together to show the locations of the five antibodies (Fig. 4C, Supplementary Fig. 5, Supplementary Table 4). The locations of 10D8, 1F10 and 6A10, which are part of the same competition group, were superimposable. In contrast, 9A6 and 3H6 adopted different locations on Pfs48/45, with both protruding in approximately the same plane as that shared by the three domains of Pfs48/45.

These findings suggest that both accessibility and also unknown functional properties of Pfs48/45 contribute to the transmission-blocking efficacy of antibodies which target Pfs48/45. When attached to the membrane, through a C-terminal GPI anchor, we would expect the Pfs48/45 disc to be on average arranged horizontal to the membrane plane (Fig. 3A). In this orientation, the epitope for 85RF45.1 will be most exposed on the membrane surface with the 32F3 epitope exposed to a lesser degree and the 10D8 epitope least accessible. This correlates with the order of transmission-blocking activity of these three antibodies, with 85RF45.1 as most potent and 10D8 as least potent. However, the epitopes for 9A6 and 3H6 emerge in the plane of Pfs48/45 and we would expect them to be more regularly exposed on the membrane surface than the epitope for 10D8, and yet they are not transmission-blocking[12]. It is also notable that neither 3H6 or 9A6 stain gametocytes, suggesting these epitopes to be buried in that context[12]. A possible explanation for this is that the region of Pfs48/45 bound by these antibodies may be occluded by its binding partners, such as Pfs230. Further studies of Pfs48/45 function are required to determine whether this is the case.

Antibodies targeting the N-terminal and central domains of Pfs48/45 substantially contribute to transmission-blocking activity of sera

The structure of Pfs48/45 indicates that each of its three domains will be exposed on the gametocyte surface (Fig. 3) and antibodies binding to each domain have been shown to stain gametocytes[12]. However, the most effective known transmission-blocking monoclonal antibodies bind to the C-terminal domain[8,25]. We therefore aimed to determine the degree to which antibodies against the N-terminal and central domains contribute to transmission-blocking activity, using a combination of immunisation and antibody depletion experiments.

We first produced protein consisting of the N-terminal and central domains of Pfs48/45 (Pfs48/45-D1 + 2) and used this to immunise mice. IgG was purified from these mice and the titres of antibodies targeting Pfs48/45-D1 + 2 were determined by end-point ELISA (Fig. 5A). The transmission-blocking activity of these antibodies was then assessed using a standard membrane feeding assay. Antibodies were added to blood containing *Plasmodium falciparum* gametocytes, fed to mosquitos, and the number of ookinetes formed in mosquito midguts was counted. Total IgG, purified from mice immunised with either 0.1 µg or 1 µg of Pfs48/45-D1 + 2 was tested at a concentration of 750 µg/ml, causing almost complete inhibition (96 and 100% reductions resulting from 0.1 µg and 1 µg immunisations, respectively) in oocyte numbers (Fig. 5B). When tested at 375 and 188 µg/ml, the total IgG from the 1 µg group also showed significant inhibition of 92 and 68%, respectively (Fig. 5C). Therefore, protein immunogens containing just the N-terminal and central domains of Pfs48/45 are able to induce highly potent transmission-blocking antibodies.

We next used a depletion approach to assess the contribution that antibodies which target the N-terminal and central domains of Pfs48/45 make to the transmission-blocking activity of IgG from mice immunised with full-length Pfs48/45. Total IgG from mice immunised with 5 µg of full-length Pfs48/45 was tested at a concentration of 750 µg/ml and showed 100% transmission-blocking activity (Fig. 5D, E). This IgG was depleted of antibodies which bind to the C-terminal domain of Pfs48/45 using an affinity column and was tested for transmission-blocking activity (Fig. 5D, E). Depletion of C-terminal domain (D3) specific antibodies from purified IgG was confirmed by ELISA, using Pfs48/45-D3 as the coating antigen, with no signal observed in the depleted IgG compared to the original un-depleted IgG (Fig. 5D).

At all three total IgG concentrations tested, the sera depleted of C-terminal domain-targeting antibodies showed no significant difference to the sera prior to depletion (Fig. 5E). This shows that the majority of the antibodies with transmission-blocking activity, induced

https://doi.org/10.1038/s41467-022-33379-6

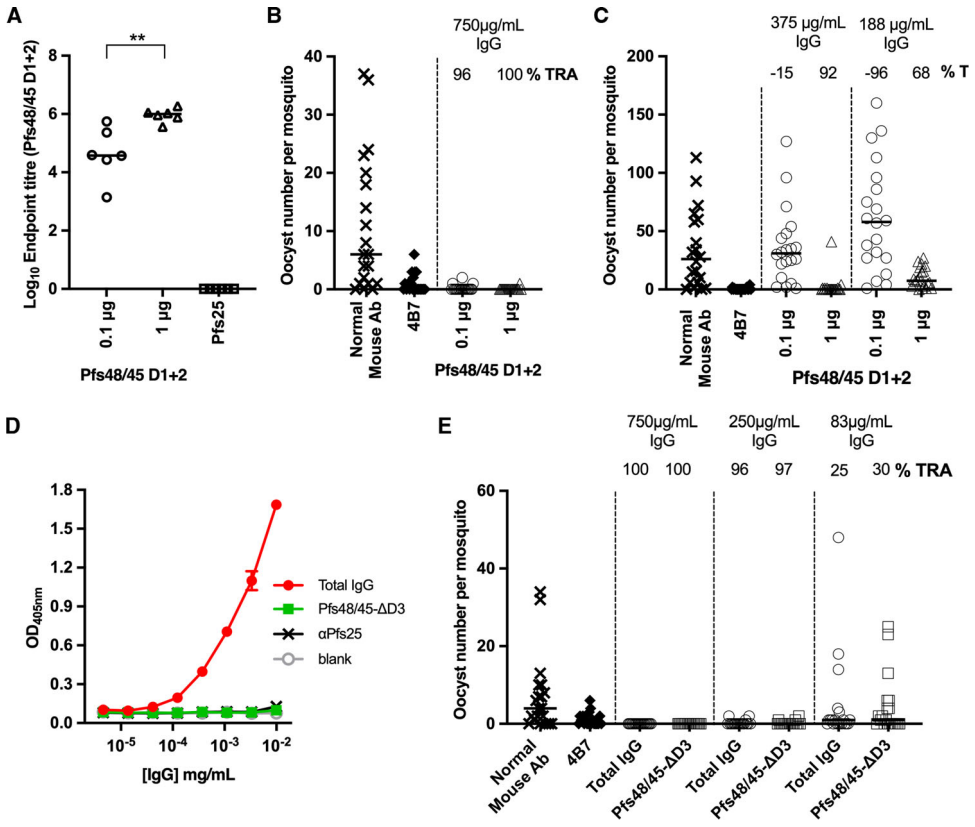


Fig. 5 | Transmission-blocking antibodies target the N-terminal and central domains of Pfs48/45. A–**C** Female CD1 mice ($n = 6$ mice) were immunised twice with either 0.1 µg or 1 µg of Pfs48/45-D1 + 2, or 1 µg of Pfs25. Sera was collected three weeks after the final dose for analysis. **A** IgG titres as measured by endpoint ELISA using Pfs48/45-D1 + 2 as the coating antigen. Each symbol is for the serum sample from an individual mouse; lines represent the median of each group. Mann–Whitney two-tailed test was performed to compare the two Pfs48/45-D1 + 2 groups ($p = 0.0043$). **B**, **C** Transmission-blocking efficacy of IgG induced by immunisation with Pfs48/45-D1 + 2. Total IgG was purified from the pooled serum of each group (3 weeks post-boost) and mixed with *P. falciparum* NF54 cultured gametocytes at 750 µg/mL (**B**) and in a separate experiment at 375 µg/mL and 83 µg/mL (**C**) and fed to *A. stephensi* mosquitoes ($n = 20$ mosquitoes per test group of pooled IgG in a single feed experiment). IgG from naive mice was used as a negative control ("normal mouse Ab"); the transmission-blocking anti-Pfs25 mAb 4B7 was used as a positive control at a concentration of 94 µg/mL. Data points represent the number of oocysts in individual mosquitoes 8 days post-feeding; horizontal lines show the arithmetic mean. **D**, **E** Female CD1 mice were immunised three times with 5 µg of Pfs48/45-FL. Three weeks after the final dose sera were collected and total IgG purified. Total IgG was depleted of Pfs48/45-D3-targeting IgG using a column coupled with Pfs48/45-D3, resulting in Pfs48/45-ΔD3. **D** Pfs48/45-D3 specific ELISA showing lack of recognition of Pfs48/45-D3 by Pfs48/45-ΔD3. Pfs25-specific IgG (αPfs25) was included as a negative control. Data presented as mean +/− standard deviation. **E** Transmission-blocking efficacy of total IgG from Pfs48/45-FL or Pfs48/45-ΔD3 fed to *A. stephensi* mosquitoes at concentrations of 750 µg/mL, 250 µg/mL and 83 µg/mL ($n = 20$ mosquitoes per test group of pooled IgG in a single feed experiment). IgG from naive mice was used as a negative control ("normal mouse Ab"); transmission-blocking anti-Pfs25 mAb 4B7 was used as a positive control. Data points represent the number of oocysts in individual mosquitoes; lines show the arithmetic mean (TRA, transmission reducing activity [% inhibition in mean oocyst count per mosquito]).

through immunisation with full-length Pfs48/45, bind to either the N-terminal or central domains.

Discussion

The structure of the Pfs48/45 ectodomain reveals that its three domains are arranged as a disc. The location of the C-terminal GPI-anchor attachment site on one of its flatter surfaces will cause the disc to lie approximately parallel to the membrane plane, with all three domains accessible on the cell surface, albeit with the flexible linker of 20 residues between the ordered section allowing considerable movement. The presence of a flexible linker between the N-terminal and central domains also causes the disc to be dynamic, due to opening and closing of the gap between the N- and C-terminal domains. How this structure and dynamics relate to the function of Pfs48/45 is currently unclear. Pfs48/45 on gametocyte surfaces is part of a larger protein complex including other proteins essential for gamete fusion, including Pfs230[32] and the PfCCp proteins[33,34]. The molecular architecture of this complex and the mechanistic roles of its components during gamete fusion, are currently not known. The structure of Pfs48/45 makes it tempting to speculate that the open, dynamic membrane-distal surface of the disc, as presented on gametocytes, acts as a platform that allows it to bind important interaction partners such as Pfs230 and the PfCCp proteins. To test this will require further study, both to determine how this complex is arranged and to understand which parts of Pfs48/45 are occluded by binding partners.

The structure of Pfs48/45 also has implications for our understanding of what makes an effective transmission-blocking antibody. A number of criteria might determine the potency of such an antibody, including the location of its epitope relative to functionally important sites on the protein, its binding kinetics and the accessibility of its epitope when presented on the parasite surface. Our lack of mechanistic insight into the role of Pfs48/45 during gamete fusion currently makes it impossible to determine whether the efficacy of antibodies correlates with their effects on protein function. However, our structural characterisation of antibody epitopes, coupled with kinetic data and measurements of transmission-blocking activity, suggest that epitope accessibility is a key feature of highly potent antibodies. In particular, comparison of 85RF45.1 and 32F3 reveals that these two antibodies have overlapping epitopes on the C-terminal domain and

have very similar affinities. The flexible nature of the 32F3 epitope, compared with the more rigid epitope for 85RF45.1 affects their kinetics of binding to soluble Pfs48/45 protein, with 85RF45.1 binding and dissociating more rapidly. However, unlike antigens involved in erythrocyte invasion[35], Pfs48/45 is exposed for significant periods of time on the gamete surface, making it unlikely that fast association-rates are important for antibody efficacy. Instead, the most likely cause of differences in potency of 85RF45.1 and 32F3 is the degree to which their epitopes are exposed on the gametocyte surface. While 32F3 approaches from the side of the disc, 85RF45.1 binds on the more exposed, membrane-distal surface, allowing greater accessibility to its epitope in the context of the gametocyte surface. Indeed, transmission-blocking antibody 10D8 has lower efficacy than either 85RF45.1 and 32F3 and binds to an epitope in a less exposed position on the Pfs48/45 surface. However, exposure and accessibility does not provide the full story. Antibodies 3H6 and 9A6 bind to more exposed surfaces of Pfs48/45 than 10D8 and yet are not transmission blocking and do not bind to gametocytes, most likely due to occlusion by Pfs48/45 binding partners. Future studies are needed to reveal how its interactions with other binding partners modulates its activity and affect vaccine design decisions.

Finally, our studies of full-length Pfs48/45 have consequences for vaccine design. Many current approaches for Pfs48/45 have focused on the C-terminal domain, which is easier to produce than full-length Pfs48/45 and contains the epitopes for the most potent known transmission-blocking antibodies. However, the structure of Pfs48/45 shows that all three of its constituent domains will be exposed when membrane anchored. Our depletion studies also show that the majority of transmission-blocking antibodies raised through immunisation of mice with full-length Pfs48/45 bind to the N-terminal and central domains. There are still many uncertainties about the role of Pfs48/45 during gamete fusion, including when Pfs48/45 assembles into a complex with other gametocyte surface proteins or which epitopes are exposed when it is part of this complex. However, our current data suggest a vaccine design strategy which includes full-length Pfs48/45, to take advantage of epitopes for transmission-blocking antibodies that target the N-terminal and central domains, perhaps first generating versions of Pfs48/45 that are thermally stabilised to prevent dynamic movement. It also indicates the importance of correct presentation of Pfs48/45 during vaccination. Immunisation with soluble, or incorrectly oriented Pfs48/45, will induce antibodies that target epitopes poorly accessible on the gametocyte surface. In contrast, a vaccine particle which presents Pfs48/45 in an orientation consistent with that on the gametocyte surface is more likely to induce antibodies, like 85RF45.1, with potent transmission-blocking potential.

Methods

Pfs48/45 expression and purification

The sequences for all Pfs48/45 protein variants (PlasmoDB:PF3D7_1346700: Pfs48/45 full-length residues 27–427, Pfs48/45-D1 + 2 residues 27–288, Pfs48/45-D2 + 3 residues 159–428 and Pfs48/45-D3 residues 291–428) were codon optimised for expression in Drosophila melanogaster (GeneArt Life Technologies), with a Drosophila BiP signal peptide introduced at the N-terminus and the four amino acids EPEA (C-Tag) inserted at the C-terminus. All N-glycosylation sites were left intact. The sequence was subcloned into the *Drosophila* S2 expression vector pExpres2.2 (ExpreS2ion Biotechnologies). Polyclonal *Drosophila* S2 stable cell lines were generated by non-viral transfection into ExpreS2 *Drosophila* S2 cells (Expres2ions Biotechnologies) with selection via G418 resistance. Recombinant Pfs48/45-D1 + 2, Pfs48/45-D2 + 3 and Pfs48/45-D3 were purified by initially concentrating the cell culture supernatant using a Tangential Flow Filtration system with a Pellicon 3 Ultracel 3 kDa membrane (Merck Millipore, UK). Pfs48/45 full-length cell culture supernatant was concentrated as above but using a Pellicon 3 Ultracel 10 kDa membrane (Merck Millipore, UK). The concentrated supernatants were then loaded onto CaptureSelect™ C-tag affinity columns (Thermo Fisher Scientific) equilibrated in Tris-buffered saline and bound proteins were eluted with 20 mM Tris–HCl, 2 M $MgCl_2$, pH 7.4. Protein containing fractions containing were then pooled and subjected to size-exclusion chromatography using a Superdex 200 16/600 PG column (Cytiva). Proteins were aliquoted and flash-frozen in liquid nitrogen for later use.

Expression and purification of antibodies and scFvs

32F3, 85RF45.1, 10D8, 1F10, 3H6, 6A10 and 9A6 were expressed from hybridoma cell lines, purified using protein A/G and ficin cleaved to generate Fab fragments[12]. To produce scFv constructs, the sequence of 32F3 scFv was designed with the variable region of the heavy chain (from residue DVKLV to residue TLTVS) followed by an 18-residue linker (GGSSRSSSSGGGGSGGGG) and the variable region of light chain (from residue QIVLS to KLELK). The sequence of the 85RF45.1 scFv contains the variable region of the light chain (from residue QFVLS to residue KLTVT), followed by the 18-residue linker and the variable region of the heavy chain (from residue EVQLV to residue MVTVS). The sequence of 10D8 scFv contains the variable region of the light chain (from residue DIVMS to residue TKLEI), followed by the 18-residue linker and the variable region of the heavy chain (from residue EVMLV to residue GTSVT). Codon optimised synthetic genes (ThermoFisher) of 10D8 scFv, 32F3 scFv, and 85RF45.1 scFv were PCR amplified and subcloned into a pHLsec vector containing a C-terminal hexa-histidine tag. These were expressed in HEK293 cells. Cultures were harvested six days after transfection. The supernatant was filtered and incubated with Excel Ni Sepharose resin (GE Healthcare) at 4 °C for 30 min. The mixture was then applied onto a gravity flow column. The column was washed with 20 mM Tris pH 8.0, 150 mM NaCl, 10 mM imidazole and eluted in the same buffer also containing 250 mM imidazole. Elution fractions were collected and concentrated to run gel filtration with superdex 200 increase 10/300 (GE Healthcare) in 20 mM Tris pH 8.0, 150 mM NaCl.

Crystallisation and structure determination

Complexes containing different combinations of full-length Pfs48/45, the C-terminal domain of Pfs48/45 (Pfs48/45-D3) and the central and C-terminal domains (Pfs48/45-D2 + 3), with one or more Fab fragment or scFv, were prepared by mixing different components in stoichiometric ratios. Complexes were purified by size exclusion chromatography on two Superdex Increase 200 10/300 columns (GE Healthcare) in series into 10 mM HEPES, 150 mM NaCl, pH 7.2. Peak fractions containing pure complex were then concentrated and used to set up vapour diffusion crystallisation trials in sitting drops by mixing 100 nl of protein complex with 100 nl of well solutions.

Crystals formed in the following conditions. For Pfs48/45-D3:32F3, initial crystals formed at a protein concentration of 11.1 mg/ml at 293 K in 0.04 M potassium dihydrogen phosphate, 16% (w/v) polyethylene glycol (PEG) 8000 and 20% (v/v) glycerol. Crystallisation conditions were further optimised by increasing the concentration of polyethylene glycol 8000 to 25% (w/v) and omitting potassium dihydrogen phosphate during crystallisation. For data collection, these crystals were transferred into well solution containing 30% (v/v) glycerol and flash-cooled in liquid nitrogen. For Pfs48/45-D2 + 3:10D8:32F3, crystals formed at a protein concentration of 10.5 mg/ml at 293 K in 10% (w/v) PEG monomethyl ether 5000, 0.1 M 2-(N-morpholino) ethanesulfonic acid pH 6.5 and 12% (v/v) 1-propanol. Initial crystals were further optimised by micro-seeding into the same condition. For data collection, crystals were transferred into well solution containing 25% (v/v) 2-methyl-2,4-pentanediol and flash-cooled in liquid nitrogen. For Pfs48/45-FL:10D8, the best crystals grew at a protein concentration of 10.7 mg/ml at 277 K in 10% (w/v) PEG 1000 and 10% (w/v) PEG 8000. For data collection, crystals were

transferred into well solution containing 25% (v/v) 2-methyl-2,4-pentanediol and flash-cooled in liquid nitrogen. Crystals for Pfs48/45-FL:85RF45.1:10D8 formed at a protein concentration of 13.4 mg/ml at 277 K in 0.1 M Lithium sulfate, 0.1 M HEPES pH 7.0 and 20% w/v Polyvinylpyrrolidone. For data collection, crystals were transferred into well solution containing 30% (v/v) glycerol and flash-cooled in liquid nitrogen. For Pfs48/45FL:32F3 scFv, crystals formed at protein concentration of 12 mg/ml at 291 K in 0.1 M MES pH 6.0, 0.2 M $CaCl_2$, 20% (w/v) PEG6000 and for data collection crystals were transferred into well solution containing 25% (v/v) glycerol.

Data were collected on the following beamlines. Data for Pfs48/45-D3:32F3 and for Pfs48/45-FL:10D8 were collected at IO4 (Diamond Light Source, UK); data for Pfs48/45-D2 + 3:10D8:32F3 were collected at IO4-1 (Diamond Light Source, UK); data for Pfs48/45-FL:85RF45.1:10D8 were collected at IO3 (Diamond Light Source, UK) and data for Pfs48/45-FL:32F3 scFv were collected at IO4 (Diamond Light Source, UK). The data for Pfs48/45-FL:85RF45.1:10D8, Pfs48/45-D3:32F3, Pfs48/45-D2 + 3:10D8:32F3 and Pfs48/45-FL:10D8 were processed with the CCP4i2 programme suite[36] using the Xia2/DIALS v3.3 pipeline[37]. For Pfs48/45FL:85RF45.1:10D8, two datasets from the same crystal were merged together. Data for Pfs48/45-FL:32F3 scFv were processed with XDS. All structures were determined by molecular replacement using Phaser-MR v2.8.3[38] with model building in Coot 0.9.4[39] and refinement using phenix.refine[40] and Buster v2.1[41]. All structure figures were prepared using PyMol v2.5.2 (Schroedinger LLC).

Surface plasmon resonance analysis

Surface plasmon resonance experiments were carried out using a Biacore T200 instrument (GE Healthcare), to analyse binding of Pfs48/45 to 85RF45.1, 32F3 and 10D8. Pfs48/45 was buffer-exchanged into 20 mM HEPES pH 7.2, 300 mM NaCl, 0.05% Tween-20. Individual mAbs were immobilised on a CM5-chip (GE Healthcare) pre-coupled to Protein A/G (Thermo Fisher) and Pfs48/45 was injected over the chip surface at a flow rate of 30 μl/min, with 240 s association time and 600 s dissociation time. For all three antibodies, we injected a series of samples, forming two-fold dilution series, with a top concentration of 7.8 nM for 85RF45.1 and 125 nM for 32F3 and 10D8. After each injection, the chip surface was regenerated with 10 mM Glycine, pH 2.0 for 120 s at 10 μl/min, followed by a regeneration period of 180 s. Data were analysed using the BIAevaluation software 2.0.3 (GE Healthcare).

Molecular dynamics simulations

To prepare structures for simulation, we first completed models for the missing regions in the Pfs48/45-FL:85RF45.1:10D8 and Pfs48/45-FL:32F3 scFv structures by grafting the model for the N-terminal domain from the Pfs48/45-FL:10D8 structure on the existing partial N-terminal domain models in the former two structures. Next, we modelled missing loops in each structure individually using MODELLER (v. 10.2)[42]. We allowed refinement only in the missing loops and generated 1000 decoys corresponding to residues 45–428 for each structure and picked the final models using the SOAP[43] scoring function. We used OpenMM (v. 7.7)[44] to run atomistic molecular dynamics simulations for each starting structure. To avoid non-physiological interactions due to truncation, we capped N- and C-termini using acetyl and N-methyl groups, respectively. Next, we protonated the models at a pH of 7.5, soaked them in truncated octahedral water boxes with a padding distance of 1 nm, and added NaCl to an ionic strength of 150 mM to neutralise charges. We parameterised the systems using the Amber14-SB force field[45] and modelled water molecules using the TIP3P-FB model. Non-bonded interactions were calculated using the particle mesh Ewald[46] method using a cut-off of distance of 0.9 nm, with an error tolerance of 0.0005. Water molecules and heavy atom-hydrogen bonds were rigidified using the SETTLE[47] and SHAKE[48] algorithms, respectively. We used hydrogen mass repartitioning[49] to allow for 4 fs time steps. Simulations were run using the mixed-precision CUDA platform in OpenMM using the Middle Langevin Integrator and the Monte-Carlo Barostat. We equilibrated systems using a multi-step protocol: (i) energy minimisation over 10,000 steps, (ii) heating of the NVT ensembles from 100 K to 300 K over 200 ps, (iii) 200 ps simulation of the NPT ensembles at 300 K, (iii) cooling of the NVT ensembles from 300 K to 100 K over 200 ps, (iv) energy minimisation over 10,000 steps, (v) heating of the NVT ensembles from 100 K to 300 K over 200 ps, and (vi) 5 ns simulation of the NPT ensembles at 300 K. Following this, we ran production simulations of NPT ensembles for 500 ns. We used MDTraj (v. 1.9.6)[50] for analyses of resulting trajectories.

Small angle X-ray scattering analysis

Data for size exclusion chromatography-coupled small angle X-ray scattering (SEC-SAXS) of full-length Pfs48/45 were collected at the EMBL P12 BioSAXS beamline[51] located at the PETRA III storage ring (DESY, Hamburg), at a wavelength of 1.24 Å using a Pilatus 2 M detector (Dectris, Baden-Daettwil, Switzerland) with a sample-detector distance of 3 m. For data collection, Pfs48/45 was concentrated to 3.25 mg/ml and injected over a Superdex Increase 3.2/300 column (GE Healthcare) equilibrated with 20 mM Hepes, 150 mM NaCl, pH 7.0. Frames were collected with an exposure time of 1 s. Data were processed with the CHROMIXS[52] and ATSAS v3.0.3[53] software suites, subtracting averaged buffer frames from averaged frames corresponding to the Pfs48/45 peak. The radius of gyration Rg was calculated using AutoRg in PRIMUS[54], the distance distribution function P(r) and the maximum particle diameter Dmax were calculated using GNOM[55]. Volumetric representations were created by first calculating 20 ab initio bead models using DAMMIF[56], followed by averaging of these models with DAMAVER[57] and subsequent refinement against the original data with DAMMIN[58]. From the resulting model, an envelope was calculated using Situs, and models of Pfs48/45 were each fitted into this envelope using Chimera[59]. Figures were prepared with PyMol (Schroedinger LLC). Assessment of the fit of models derived from molecular dynamics to SAXS data was conducted using FoxS[31]. We we removed all N-methyl and acetyl caps as well as hydrogen atoms from the models and FoxS was used with the default settings.

Negative stain electron microscopy and image processing

Each Fab was mixed with full-length Pfs48/45 and 85RF45.1 Fab at the ratio of 1:1:1 and incubated on ice for 30 min in 20 mM Tris pH 7.5, 150 mM NaCl. The mixture was then injected onto Superdex 200 Increase 10/300 column (Cytiva). Fractions corresponding to Pfs48/45 bound to two Fabs (around 150 kD) were collected and concentrated. Purified complexes were flash-frozen using liquid nitrogen and stored at −80 °C.

For negative staining, purified complexes of Pfs48/45:85RF45.1 Fab with one additional Fab, from 1F10 Fab, 3H6 Fab, 6A10 Fab, 9A6 Fab or 10D8 Fab were diluted to 10 μg/ml. Carbon support 200-mesh copper grids were glow-discharged at 15 mA for 20 s. 5 μl protein sample was applied onto the grids for 60 s and blotted. Then the grids were washed twice with 20 mM Tris pH 7.5, 150 mM NaCl buffer and stained twice with 2% uranyl acetate for 60 s and blotted.

100–120 micrographs were collected for each sample on Talos F200c transmission electron microscope with a Ceta 16 M CMOS camera (Thermo Scientific). Nominal magnification was ×45,000 for the 1F10 dataset (pixel size 2.30 Å), and ×57,000 for the other datasets (pixel size 1.8 Å). The defocus was set at −5 μm for the 1F10 dataset, and at −1.5 μm for other datasets.

Data were first processed using SIMPLE v3.0[60]. The micrographs had their contrast transfer functions estimated, and particles were automatically picked based on 2D references generated from manually picked particles. 10,000 to 30,000 particles were selected after two rounds of 2D classification and selected to build initial 3D models using

SIMPLE. The particles were then exported to RELION v3.1[61,62] followed by a round of 2D classification. Around 10k particles were selected for 3D classification using the initial 3D model generated from SIMPLE. 3D refinement was then conducted on one of the classes from the 3D classification.

Immunisation of mice

Animal experiments and procedures were performed according to the UK Animals (Scientific Procedures) Act Project License (PA7D20B85) and approved by the Oxford University Local Ethical Review Body. Six- to eight-week-old female CD-1 mice (Envigo RMS Inc.), housed in specific-pathogen-free environments, were vaccinated with full-length Pfs48/45, Pfs48/45-D1 + 2, or Pfs25, into each leg via the intramuscular route, using a prime-boost regime with a 3-week interval, with blood being collected 3 weeks after the final dose. Immunisations were formulated in adjuvant prior to vaccination by mixing Alhydrogel® adjuvant 2% (Invivogen; 85 µg per dose) with antigen in TBS, and incubated at RT for 1 h before injection. Sera were obtained from whole blood by leaving samples overnight at 4 °C to clot, followed by 10 min centrifugation at 16,000 × *g* in a benchtop centrifuge at room temperature.

Enzyme-linked immunosorbent assays

For endpoint ELISAs, Nunc-Immuno Maxisorp 96-well plates (Thermo Scientific) coated with 2 µg/ml of antigen in PBS (Sigma) overnight at 4 °C. Plates were washed with PBS-Tween (0.05%) and blocked with 10% Casein Block (Thermo Scientific). Diluted samples for endpoint ELISAs were added to the top row of the plate in duplicate, and used to generate a threefold serial dilution. Plates were incubated for 2 h at room temperature and then washed as before. Goat anti-mouse whole IgG conjugated to alkaline phosphatase (Sigma) was added for 1 h at room temperature. Following a final wash, plates were developed by adding *p*-nitrophenylphosphate (Sigma) at 1 mg/mL in diethanolamine buffer (Sigma) and OD_{405nm} was read on an ELx808 microtitre plate reader (Biotek). The endpoint titre is defined as the x-axis intercept of the dilution curve at an absorbance value (±three standard deviations) greater than the OD for a serum sample from a naive mouse. If a 1:250 dilution of a serum sample did not develop a signal above background, the IgG titre in this sample was considered to be 0. Sera from a pool of mice with a strong anti Pfs48/45-D1 + 2 response were included on each plate as an internal control, and all plates developed until this control reached an endpoint titre of 1,350,000–1,650,000. The sera of mice immunised with Pfs25 was included as a negative control.

To validate the depletion of Pfs48/45-D3 specificity from purified IgG, ELISAs were conducted as above with Pfs48/45-D3 being used as the coating antigen and IgG being serially diluted threefold down the plate from an initial concentration of 0.01 mg/mL. Raw $OD405_{nm}$ is presented. Pfs25-specific IgG was included as a negative control.

Standard membrane feeding assays

The ability of vaccine-induced antibodies to block the development of *P. falciparum* strain NF54 was evaluated by SMFA as previously described[63]. The percentage of mature stage V gametocytes was adjusted to 0.15 ± 0.05% and the male-female ratio was stable (almost always 1 male: 2–3 female). These were mixed with purified IgG at the concentrations (diluted in PBS) shown in the figures and then fed to 4–6 days old starved female *Anopheles stephensi* (SDA 500) via a parafilm membrane. The mosquitoes were maintained at 26 °C and 80% relative humidity. After 8 days, midguts from 20 mosquitoes per group were dissected, oocysts counted, and the number of infected mosquitoes recorded. Percent reduction in infection intensity was calculated relative to the respective control IgG tested (Normal mouse IgG) in the same assay. The monoclonal antibody 4B7, specific for Pfs25 was included as a positive control at 94 µg/ml.

Statistical analysis

Comparison of endpoint ELISA data between sera of the two groups of mice was performed by a Mann–Whitney test. The 95% confidence intervals (95% CI), and *p*-values of SMFA results were calculated using a zero-inflated negative binomial random effects model described previously[64].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Coordinates and structure factors derived in this study have been deposited in the Protein Data Bank under accession codes 7ZWF, 7ZWI, 7ZWM, 7ZXF and 7ZXG. Source data for all graphs generated in this study are provided in "source data.xlsx". Source data are provided with this paper.

References

1. Sinnis, P. & Fidock, D. A. The RTS,S vaccine—a chance to regain the upper hand against malaria? *Cell* **185**, 750–754 (2022).
2. Datoo, M. S. et al. Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. *Lancet* **397**, 1809–1818 (2021).
3. Duffy, P. E. Making a good malaria vaccine better. *Trends Parasitol.* **38**, 9–10 (2022).
4. Draper, S. J. et al. Malaria vaccines: recent advances and new horizons. *Cell Host Microbe* **24**, 43–56 (2018).
5. Duffy, P. E. Transmission-blocking vaccines: harnessing herd immunity for malaria elimination. *Expert Rev. Vaccines* **20**, 185–198 (2021).
6. van Dijk, M. R. et al. A central role for P48/45 in malaria parasite male gamete fertility. *Cell* **104**, 153–164 (2001).
7. Rener, J., Graves, P. M., Carter, R., Williams, J. L. & Burkot, T. R. Target antigens of transmission-blocking immunity on gametes of plasmodium falciparum. *J. Exp. Med.* **158**, 976–981 (1983).
8. Vermeulen, A. N. et al. Sequential expression of antigens on sexual stages of Plasmodium falciparum accessible to transmission-blocking antibodies in the mosquito. *J. Exp. Med.* **162**, 1460–1476 (1985).
9. Outchkourov, N. S. et al. Correctly folded Pfs48/45 protein of *Plasmodium falciparum* elicits malaria transmission-blocking immunity in mice. *Proc. Natl Acad. Sci. USA* **105**, 4301–4305 (2008).
10. Chowdhury, D. R., Angov, E., Kariuki, T. & Kumar, N. A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*. *PLoS ONE* **4**, e6352 (2009).
11. Kapulu, M. C. et al. Comparative assessment of transmission-blocking vaccine candidates against *Plasmodium falciparum*. *Sci. Rep.* **5**, 11193 (2015).
12. Lennartz, F. et al. Structural basis for recognition of the malaria vaccine candidate Pfs48/45 by a transmission blocking antibody. *Nat. Commun.* **9**, 3822 (2018).
13. Graves, P. M. et al. Naturally occurring antibodies to an epitope on *Plasmodium falciparum* gametes detected by monoclonal antibody-based competitive enzyme-linked immunosorbent assay. *Infect. Immun.* **56**, 2818–2821 (1988).
14. Roeffen, W. et al. Association between anti-Pfs48/45 reactivity and *P. falciparum* transmission-blocking activity in sera from Cameroon. *Parasite Immunol.* **18**, 103–109 (1996).
15. Bousema, T. et al. The dynamics of naturally acquired immune responses to *Plasmodium falciparum* sexual stage antigens Pfs230 & Pfs48/45 in a low endemic area in Tanzania. *PLoS ONE* **5**, e14114 (2010).

16. Jones, S. et al. Naturally acquired antibody responses to recombinant Pfs230 and Pfs48/45 transmission blocking vaccine candidates. *J. Infect.* **71**, 117–127 (2015).
17. Stone, W. J. et al. Naturally acquired immunity to sexual stage *P. falciparum* parasites. *Parasitology* **143**, 187–198 (2016).
18. de Jong, R. M. et al. Monoclonal antibodies block transmission of genetically diverse *Plasmodium falciparum* strains to mosquitoes. *NPJ Vaccines* **6**, 101 (2021).
19. Niederwieser, I., Felger, I. & Beck, H. P. Limited polymorphism in *Plasmodium falciparum* sexual-stage antigens. *Am. J. Trop. Med. Hyg.* **64**, 9–11 (2001).
20. Kocken, C. H. et al. Minimal variation in the transmission-blocking vaccine candidate Pfs48/45 of the human malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **69**, 115–118 (1995).
21. Theisen, M., Jore, M. M. & Sauerwein, R. Towards clinical development of a Pfs48/45-based transmission blocking malaria vaccine. *Expert Rev. Vaccines* **16**, 329–336 (2017).
22. Carter, R., Coulson, A., Bhatti, S., Taylor, B. J. & Elliott, J. F. Predicted disulfide-bonded structures for three uniquely related proteins of *Plasmodium falciparum*, Pfs230, Pfs48/45 and Pf12. *Mol. Biochem. Parasitol.* **71**, 203–210 (1995).
23. Gerloff, D. L., Creasey, A., Maslau, S. & Carter, R. Structural models for the protein family characterized by gamete surface protein Pfs230 of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **102**, 13598–13603 (2005).
24. Templeton, T. J. & Kaslow, D. C. Identification of additional members define a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol. Biochem. Parasitol.* **101**, 223–227 (1999).
25. Roeffen, W. et al. *Plasmodium falciparum*: production and characterization of rat monoclonal antibodies specific for the sexual-stage Pfs48/45 antigen. *Exp. Parasitol.* **97**, 45–49 (2001).
26. Carter, R., Bushell, G., Saul, A., Graves, P. M. & Kidson, C. Two apparently nonrepeated epitopes on gametes of *Plasmodium falciparum* are targets of transmission-blocking antibodies. *Infect. Immun.* **50**, 102–106 (1985).
27. Carter, R., Graves, P. M., Keister, D. B. & Quakyi, I. A. Properties of epitopes of Pfs 48/45, a target of transmission blocking monoclonal antibodies, on gametes of different isolates of *Plasmodium falciparum*. *Parasite Immunol.* **12**, 587–603 (1990).
28. Outchkourov, N. et al. Epitope analysis of the malaria surface antigen pfs48/45 identifies a subdomain that elicits transmission blocking antibodies. *J. Biol. Chem.* **282**, 17148–17156 (2007).
29. Kundu, P. et al. Structural delineation of potent transmission-blocking epitope I on malaria antigen Pfs48/45. *Nat. Commun.* **9**, 4458 (2018).
30. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
31. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
32. Kumar, N. Target antigens of malaria transmission blocking immunity exist as a stable membrane bound complex. *Parasite Immunol.* **9**, 321–335 (1987).
33. Pradel, G. et al. A multidomain adhesion protein family expressed in *Plasmodium falciparum* is essential for transmission to the mosquito. *J. Exp. Med.* **199**, 1533–1544 (2004).
34. Simon, N., Kuehn, A., Williamson, K. C. & Pradel, G. Adhesion protein complexes of malaria gametocytes assemble following parasite transmission to the mosquito. *Parasitol. Int.* **65**, 27–30 (2016).
35. Alanine, D. G. W. et al. Human antibodies that slow erythrocyte invasion potentiate malaria-neutralizing antibodies. *Cell* **178**, 216–228.e21 (2019).
36. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D. Biol. Crystallogr* **67**, 235–242 (2011).
37. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D. Struct. Biol.* **74**, 85–97 (2018).
38. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr* **40**, 658–674 (2007).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D. Biol. Crystallogr* **66**, 486–501 (2010).
40. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D. Biol. Crystallogr* **66**, 213–221 (2010).
41. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D. Biol. Crystallogr* **68**, 368–380 (2012).
42. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **86**, 2 9 1–2 9 37 (2016).
43. Dong, G. Q., Fan, H., Schneidman-Duhovny, D., Webb, B. & Sali, A. Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics* **29**, 3158–3166 (2013).
44. Eastman, P. et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
45. Tian, C. et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020).
46. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
47. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
48. Rychaert, J., 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.* **23**, 327–341 (1977).
49. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
50. McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
51. Blanchet, C. E. et al. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr* **48**, 431–443 (2015).
52. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
53. Manalastas-Cantos, K. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr* **54**, 343–355 (2021).
54. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS—a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).
55. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr* **25**, 495–503 (1992).
56. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr* **42**, 342–346 (2009).
57. Volkov, V. V. & Svergun, D. I. Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr* **36**, 860–864 (2003).

58. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
59. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
60. Elmlund, D. & Elmlund, H. SIMPLE: software for ab initio reconstruction of heterogeneous single-particles. *J. Struct. Biol.* **180**, 420–427 (2012).
61. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
62. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
63. Miura, K. et al. Qualification of standard membrane-feeding assay with *Plasmodium falciparum* malaria and potential improvements for future assays. *PLoS ONE* **8**, e57909 (2013).
64. Miura, K. et al. Transmission-blocking activity is determined by transmission-reducing activity and number of control oocysts in *Plasmodium falciparum* standard membrane-feeding assay. *Vaccine* **34**, 4145–4151 (2016).

Acknowledgements

This work was funded by a Medical Research Council project grant (MR/R001138/1) to S.B. and M.K.H. This work was funded in part by the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 733273). M.K.H. is a Wellcome Investigator. The SMFA activity was supported by the intramural programme of the National Institute of Allergy and Infectious Disease/NIH. The authors would like to thank Ed Lowe and the beamline scientists at beamlines I03, I04 and I04.1 at the Diamond synchrotron for help with crystallographic data collection, and the beamline staff at beamline P12 at DESY for help with SAXS data collection. Electron microscopy data was collected at the COSMIC facility in Oxford and we thank Teige Matthews-Palmer, Rishi Matadeen and Joseph Caesar for support. We are grateful to Simon Draper and Lloyd King for facilitating use of the project license for the animal work.

Author contributions

K.K., F.L., D.M., S.B. and M.K.H. conceived and planned the study and wrote the manuscript. D.M. and D.D. produced Pfs48/45 and fragments. A.M. performed animal studies. D.D. and D.M. conducted endpoint ELISAs. K.M. and C.A.L. performed SMFAs. M.J. provided 32F3 and 85RF45.1. F.L. prepared Fab fragments, purified and crystallised complexes and collected X-ray diffraction data. K.K. prepared scFv, purified and crystallised complexes and collected X-ray diffraction data. K.K., F.L. and M.K.H. built, refined and analysed structures. F.L. performed and analysed SPR and SAXS experiments. B.G. performed and analysed molecular dynamics simulations. K.K performed negative stain sample preparation, data collection and analysis. All authors read and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-33379-6.

Correspondence and requests for materials should be addressed to Sumi Biswas or Matthew K. Higgins.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

References

Abanades, Brennan et al. (2021). "ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation". In: *Bioinformatics* 38, pp. 1877–1880.

Abanades, Brennan et al. (2022). "ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins". In: *Communications Biology* 6.

Abhinandan, K. R. and Andrew C. R. Martin (2008). "Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains." In: *Molecular immunology* 45 14, pp. 3832–9.

Adams, Nicholas M., Simon Grassmann, and Joseph C. Sun (2020). "Clonal expansion of innate and adaptive lymphocytes". In: *Nature Reviews Immunology* 20, pp. 694–707.

Adolf-Bryfogle, Jared et al. (2017). "RosettaAntibodyDesign (RAbD): A general framework for computational antibody design". In: *PLoS Computational Biology* 14.

"AI will transform science — now researchers must tame it" (2023). In: *Nature* 621, pp. 658–658.

Akbar, Rahmad et al. (2021). "A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding". In: *Cell Reports* 34.

Alanine, Daniel G. W. et al. (2019). "Human Antibodies that Slow Erythrocyte Invasion Potentiate Malaria-Neutralizing Antibodies". In: *Cell* 178, 216–228.e21.

Alberts, Bruce et al. (2022). *Molecular biology of the cell*. W.W. Norton Company.

Alford, Rebecca F. et al. (2017). "The Rosetta all-atom energy function for macromolecular modeling and design". In: *bioRxiv*.

Almagro, Juan Carlos et al. (2011). "Antibody modeling assessment". In: *Proteins: Structure* 79.

Almagro, Juan Carlos et al. (2014). "Second antibody modeling assessment (AMA-II)". In: *Proteins: Structure* 82.

Alt, Frederick W. and David Baltimore (1982). "Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions." In: *Proceedings of the National Academy of Sciences of the United States of America* 79 13, pp. 4118–22.

Amaral, Marta et al. (2017). "Protein conformational flexibility modulates kinetics and thermodynamics of drug binding". In: *Nature Communications* 8.

Andersen, Hans Christian (1980). "Molecular dynamics simulations at constant pressure and/or temperature". In: *Journal of Chemical Physics* 72, pp. 2384–2393.

Åqvist, Johan et al. (2004). "Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm". In: *Chemical Physics Letters* 384, pp. 288–294.

Bai, Hongjun et al. (2011). "Predicting kinetic constants of protein–protein interactions based on structural properties". In: *Proteins: Structure* 79.

Barducci, Alessandro, Giovanni Bussi, and Michele Parrinello (2008). "Well-tempered metadynamics: a smoothly converging and tunable free-energy method". In: *Physical review letters* 100.

Bateman, Alex et al. (2022). "UniProt: the Universal Protein Knowledgebase in 2023". In: *Nucleic Acids Research* 51, pp. D523–D531.

Berek, Claudia, A. Berger, and Mara Apel (1991). "Maturation of the immune response in germinal centers". In: *Cell* 67, pp. 1121–1129.

Berendsen, Herman J. C. et al. (1984). "Molecular dynamics with coupling to an external bath". In: *Journal of Chemical Physics* 81, pp. 3684–3690.

Berman, Helen M. et al. (2000). "The Protein Data Bank". In: *Nucleic acids research* 28 1, pp. 235–42.

Blackler, Ryan J. et al. (2022). "Antigen binding by conformational selection in near-germline antibodies". In: *The Journal of Biological Chemistry* 298.

Bonati, Luigi, GiovanniMaria Piccini, and Michele Parrinello (2021). "Deep learning the slow modes for rare events sampling". In: *Proceedings of the National Academy of Sciences* 118.

Bork, Peer, Liisa Holm, and Chris Sander (1994). "The immunoglobulin fold. Structural classification, sequence patterns and common core." In: *Journal of molecular biology* 242 4, pp. 309–20.

Boyd, Scott D. et al. (2010). "Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements". In: *The Journal of Immunology* 184, pp. 6986–6992.

Branduardi, Davide, Giovanni Bussi, and Michele Parrinello (2012). "Metadynamics with adaptive gaussians". In: *Journal of chemical theory and computation* 8 7, pp. 2247–54.

Braun, Efrem et al. (2019). "Best Practices for Foundations in Molecular Simulations [Article v1.0]." In: *Living journal of computational molecular science* 1 1.

Briney, Bryan S. et al. (2018). "Commonality despite exceptional diversity in the baseline human antibody repertoire". In: *Nature* 566, pp. 393–397.

Bussi, Giovanni and Alessandro Laio (2020). "Using metadynamics to explore complex free-energy landscapes". In: *Nature Reviews Physics* 2, pp. 200–212.

Callaway, Ewen (2015). "The revolution will not be crystallized: a new method sweeps through structural biology". In: *Nature* 525, pp. 172–174.

Canutescu, Adrian A. and Roland L. Dunbrack (2003). "Cyclic coordinate descent: A robotics algorithm for protein loop closure". In: *Protein Science* 12.

Castagnos, François, Martin Mihelich, and Charles Dognin (2022). "A Simple Log-based Loss Function for Ordinal Text Classification". In: *International Conference on Computational Linguistics*.

Chaudhury, Sidhartha, Sergey Lyskov, and Jeffrey J. Gray (2010). "PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta". In: *Bioinformatics* 26 5, pp. 689–91.

Chernyshev, Mark et al. (2023). "Vaccination of SARS-CoV-2-infected individuals expands a broad range of clonally diverse affinity-matured B cell lineages". In: *Nature Communications* 14.

Chin, Mark, Claire Marks, and Charlotte M. Deane (2021). "Humanization of antibodies using a machine learning approach on large-scale repertoire data". In: *Bioinformatics* 37, pp. 4041–4047.

Chinery, Lewis et al. (2022). "Paragraph - Antibody Paratope prediction using Graph Neural Networks with minimal feature vectors". In: *bioRxiv*.

Chiu, See Hong and Lei Xie (2015). "Toward high-throughput predictive modeling of protein binding/unbinding kinetics". In: *bioRxiv*.

Choi, Yoonjoo and Charlotte M. Deane (2010). "FREAD revisited: Accurate loop structure prediction using a database search algorithm". In: *Proteins: Structure* 78.

Chothia, Cyrus et al. (1987). "Canonical structures for the hypervariable regions of immunoglobulins." In: *Journal of molecular biology* 196 4, pp. 901–17.

Cock, Peter J. A. et al. (2009). "Biopython: freely available Python tools for computational molecular biology and bioinformatics". In: *Bioinformatics* 25, pp. 1422–1423.

Cunningham, Orla et al. (2021). "Polyreactivity and polyspecificity in therapeutic antibody development: risk factors for failure in preclinical and clinical development campaigns". In: *mAbs* 13.

Dama, James F., Michele Parrinello, and Gregory A. Voth (2014). "Well-tempered metadynamics converges asymptotically." In: *Physical review letters* 112 24, p. 240602.

Dara, Suresh et al. (2021). "Machine Learning in Drug Discovery: A Review". In: *Artificial Intelligence Review* 55, pp. 1947–1999.

Darden, Thomas A., Darrin M. York, and Lee G. Pedersen (1993). "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems". In: *Journal of Chemical Physics* 98, pp. 10089–10092.

Del Alamo, Diego et al. (2022). "Sampling alternative conformational states of transporters and receptors with AlphaFold2". In: *eLife* 11.

Dizicheh, Zahra Bahrami, I-Ling Chen, and Patrick Koenig (2023). "VHH CDR-H3 conformation is determined by VH germline usage". In: *Communications Biology* 6.

Dobson, Claire Louise et al. (2016). "Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo". In: *Scientific Reports* 6.

Dong, Guang Qiang et al. (2013). "Optimized atomic statistical potentials: assessment of protein interfaces and loops". In: *Bioinformatics* 29 24, pp. 3158–66.

Du, Zongyang et al. (2021). "The trRosetta server for fast and accurate protein structure prediction". In: *Nature Protocols* 16, pp. 5634–5651.

Dunbar, James and Charlotte M. Deane (2015). "ANARCI: antigen receptor numbering and receptor classification". In: *Bioinformatics* 32, pp. 298–300.

Dunbar, James et al. (2013). "ABangle: characterising the VH-VL orientation in antibodies." In: *Protein engineering, design & selection : PEDS* 26 10, pp. 611–20.

Dunbar, James et al. (2014). "SAbDab: the structural antibody database". In: *Nucleic Acids Research* 42, pp. D1140–D1146.

Eastman, Peter K. et al. (2016). "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics". In: *PLoS Computational Biology* 13.

Evans, Denis J. and Brad Lee Holian (1985). "The Nose–Hoover thermostat". In: *Journal of Chemical Physics* 83, pp. 4069–4074.

Evans, Richard et al. (2021). "Protein complex prediction with AlphaFold-Multimer". In: *bioRxiv*.

Faraggi, Eshel et al. (2018). "Comparing NMR and X-ray protein structure: Lindemann-like parameters and NMR disorder". In: *Journal of Biomolecular Structure and Dynamics* 36, pp. 2331–2341.

Fernández-Quintero, Monica L. et al. (2019a). "CDR-H3 loop ensemble in solution – conformational selection upon antibody binding". In: *mAbs* 11, pp. 1077–1088.

Fernández-Quintero, Monica L. et al. (2019b). "Characterizing the Diversity of the CDR-H3 Loop Conformational Ensembles in Relationship to Antibody Binding Properties". In: *Frontiers in Immunology* 9.

Fernández-Quintero, Monica L. et al. (2020a). "Antibodies exhibit multiple paratope states influencing VH–VL domain orientations". In: *Communications Biology* 3.

Fernández-Quintero, Monica L. et al. (2020b). "Antibody CDR loops as ensembles in solution vs. canonical clusters from X-ray structures". In: *mAbs* 12.

Fernández-Quintero, Monica L. et al. (2020c). "Local and Global Rigidification Upon Antibody Affinity Maturation". In: *Frontiers in Molecular Biosciences* 7.

Fernández-Quintero, Monica L. et al. (2021). "Germline-Dependent Antibody Paratope States and Pairing Specific VH-VL Interface Dynamics". In: *Frontiers in Immunology* 12.

Fey, Matthias and Jan Eric Lenssen (2019). "Fast Graph Representation Learning with PyTorch Geometric". In: *ArXiv* abs/1903.02428.

Fischer, Michael B. et al. (1998). "Dependence of germinal center B cells on expression of CD21/CD35 for survival." In: *Science* 280 5363, pp. 582–5.

Foote, J. Lindsley. and Gregory Paul Winter (1992). "Antibody framework residues affecting the conformation of the hypervariable loops." In: *Journal of molecular biology* 224 2, pp. 487–99.

Fraser, James S. et al. (2011). "Accessing protein conformational ensembles using room-temperature X-ray crystallography". In: *Proceedings of the National Academy of Sciences* 108, pp. 16247–16252.

Gilfillan, Susan, Christophe Benoist, and Diane Mathis (1995). "Mice Lacking Terminal Deoxynucleotidyl Transferase: Adult Mice with a Fetal Antigen Receptor Repertoire". In: *Immunological Reviews* 148.

Gitlin, Alexander D., Ziv Shulman, and Michel C. Nussenzweig (2014). "Clonal selection in the germinal centre by regulated proliferation and hypermutation". In: *Nature* 509, pp. 637–640.

Goldstein, Leonard D. et al. (2019). "Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies". In: *Communications Biology* 2.

Goodfellow, Ian J., Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. `http://www.deeplearningbook.org`. Cambridge, MA, USA: MIT Press.

Gordon, Gemma L. et al. (2023). "A comparison of the binding sites of antibodies and single-domain antibodies". In: *Frontiers in Immunology* 14.

Gowers, Richard J. et al. (2016). "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations". In.

Greener, Joe G. et al. (2021). "A guide to machine learning for biologists". In: *Nature Reviews Molecular Cell Biology* 23, pp. 40–55.

Greiff, Victor et al. (2017). "Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development." In: *Cell reports* 19 7, pp. 1467–1478.

Guest, Johnathan D. et al. (2020). "An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants." In: *Structure*.

Guloglu, Bora and Charlotte M. Deane (2023). "Specific attributes of the VL domain influence both the structure and structural variability of CDR-H3 through steric effects". In: *Frontiers in Immunology* 14.July, pp. 1–14.

Guo, Haoyang et al. (2022). "AlphaFold2 models indicate that protein sequence determines both structure and dynamics". In: *Scientific Reports* 12.

H'enin, J'erome et al. (2022). "Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]". In: *Living Journal of Computational Molecular Science*.

Hamelryck, Thomas and Bernard Manderick (2003). "PDB file parser and structure class implemented in Python". In: *Bioinformatics* 19 17, pp. 2308–10.

Henzler-Wildman, Katherine A. and Dorothee Kern (2007). "Dynamic personalities of proteins". In: *Nature* 450, pp. 964–972.

Heo, Lim, Collin F Arbour, and Michael Feig (2019). "Driven to near-experimental accuracy by refinement via molecular dynamics simulations". In: *Proteins: Structure* 87, pp. 1263–1275.

Hiranuma, Naozumi et al. (2020). "Improved protein structure refinement guided by deep learning based accuracy estimation". In: *Nature Communications* 12.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Computation* 9, pp. 1735–1780.

Hopf, Thomas A. et al. (2014). "Sequence co-evolution gives 3D contacts and structures of protein complexes". In: *eLife* 3.

Hopkins, Chad W. et al. (2015). "Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning." In: *Journal of chemical theory and computation* 11 4, pp. 1864–74.

Hou, Minghua et al. (2023). "Protein multiple conformations prediction using multi-objective evolution algorithm". In: *bioRxiv*.

Huang, Jing et al. (2016). "CHARMM36m: an improved force field for folded and intrinsically disordered proteins". In: *Nature Methods* 14, pp. 71–73.

Hummer, Alissa M., Brennan Abanades, and Charlotte M. Deane (2022). "Advances in computational structure-based antibody design." In: *Current opinion in structural biology* 74, p. 102379.

Hummer, Alissa M. et al. (2023). "Investigating the Volume and Diversity of Data Needed for Generalizable Antibody-Antigen G Prediction". In: *bioRxiv*.

Inoue, Takeshi et al. (2020). "Exit from germinal center to become quiescent memory B cells depends on metabolic reprograming and provision of a survival signal". In: *The Journal of Experimental Medicine* 218.

Izaguirre, Jesús A. et al. (2001). "Langevin stabilization of molecular dynamics". In: *Journal of Chemical Physics* 114, pp. 2090–2098.

Jackson, Katherine J. L. et al. (2004). "Exonuclease activity and P nucleotide addition in the generation of the expressed immunoglobulin repertoire". In: *BMC Immunology* 5, pp. 19–19.

Jacobson, M. et al. (2002). "On the role of the crystal environment in determining protein side-chain conformations." In: *Journal of molecular biology* 320 3, pp. 597–608.

Jaffe, David B. et al. (2022). "Functional antibodies exhibit light chain coherence". In: *Nature* 611, pp. 352–357.

Jain, Tushar et al. (2017). "Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning". In: *Bioinformatics* 33, pp. 3758–3766.

Jamasb, Arian, Pietro Lio', and Tom L. Blundell (2020). "Graphein - a Python Library for Geometric Deep Learning and Network Analysis on Protein Structures and Interaction Networks". In: *bioRxiv*.

Janin, Joël and Francis Rodier (1995). "Protein–protein interaction at crystal contacts". In: *Proteins: Structure* 23.

Jarasch, Alexander et al. (2015). "Developability assessment during the selection of novel therapeutic antibodies." In: *Journal of pharmaceutical sciences* 104 6, pp. 1885–1898.

Jeliazkov, Jeliazko R. et al. (2017). "Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification". In: *Frontiers in Immunology* 9.

Jimenez, Ralph et al. (2002). "Flexibility and molecular recognition in the immune system". In: *Proceedings of the National Academy of Sciences of the United States of America* 100, pp. 92–97.

Jin, Wengong et al. (2021). "Iterative Refinement Graph Neural Network for Antibody Sequence-Structure Co-design". In: *ArXiv* abs/2110.04624.

Jing, Bowen et al. (2023). "EigenFold: Generative Protein Structure Prediction with Diffusion Models". In: *ArXiv*.

Jing, Xiaoyang and Jinbo Xu (2021). "Fast and effective protein model refinement using deep graph neural networks". In: *Nature Computational Science* 1, pp. 462–469.

Jolly, Christopher J. et al. (1996). "The targeting of somatic hypermutation." In: *Seminars in immunology* 8 3, pp. 159–68.

Jorgensen, William L. et al. (1983). "Comparison of simple potential functions for simulating liquid water". In: *Journal of Chemical Physics* 79, pp. 926–935.

Jubb, Harry C et al. (2017). "Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures". In: *Journal of Molecular Biology* 429, pp. 365–371.

Julian, Mark C. et al. (2017). "Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability". In: *Scientific Reports* 7.

Jumper, John M. et al. (2021). "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596, pp. 583–589.

Kabsch, Wolfgang and Christian A. Sander (1983). "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22.

Kelow, Simon P., Jared Adolf-Bryfogle, and Roland L. Dunbrack (2020). "Hiding in plain sight: structure and sequence analysis reveals the importance of the antibody DE loop for antibody-antigen binding". In: *mAbs* 12.

Khalid, Zunera et al. (2022). "IgNAR antibody: Structural features, diversity and applications." In: *Fish & shellfish immunology*.

Kidd, Marie J et al. (2016). "DJ Pairing during VDJ Recombination Shows Positional Biases That Vary among Individuals with Differing IGHD Locus Immunogenotypes". In: *The Journal of Immunology* 196, pp. 1158–1164.

Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980.

Koshland, Daniel E. (1958). "Application of a Theory of Enzyme Specificity to Protein Synthesis." In: *Proceedings of the National Academy of Sciences of the United States of America* 44 2, pp. 98–104.

Kotagiri, Prasanti et al. (2021). "B cell receptor repertoire kinetics after SARS-CoV-2 infection and vaccination". In: *Cell Reports* 38, pp. 110393–110393.

Kovaltsuk, Aleksandr et al. (2018). "Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires". In: *The Journal of Immunology* 201, pp. 2502–2509.

Krishnamurthy, Vijay M. et al. (2008). "Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding." In: *Chemical reviews* 108 3, pp. 946–1051.

Kryshtafovych, Andriy et al. (2021). "Critical assessment of methods of protein structure prediction (CASP)—Round XIV". In: *Proteins: Structure* 89, pp. 1607–1617.

Laio, Alessandro and Michele Parrinello (2002). "Escaping free-energy minima". In: *Proceedings of the National Academy of Sciences of the United States of America* 99, pp. 12562–12566.

Lauer, Timothy M et al. (2012). "Developability index: a rapid in silico tool for the screening of antibody aggregation propensity." In: *Journal of pharmaceutical sciences* 101 1, pp. 102–15.

Leach, Andrew R. (2001). *Molecular Modelling: Principles and Applications*. Prentice Hall.

Leaver-Fay, Andrew et al. (2011). "ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules." In: *Methods in enzymology* 487, pp. 545–74.

LeCun, Yann, Yoshua Bengio, and Geoffrey E. Hinton (2015). "Deep Learning". In: *Nature* 521, pp. 436–444.

Ledvina, Polly S. et al. (1998). "Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: Equilibrium, kinetic, and crystallographic studies". In: *Protein Science* 7.

Lee, Gyu Rie et al. (2019). "GalaxyRefine2: simultaneous refinement of inaccurate local regions and overall protein structure". In: *Nucleic Acids Research* 47, W451–W455.

Lee, Jae Hyeon et al. (2023). "EquiFold: Protein Structure Prediction with a Novel Coarse-Grained Structure Representation". In: *bioRxiv*.

Leem, Jinwoo et al. (2016). "ABodyBuilder: Automated antibody structure prediction with data–driven accuracy estimation". In: *mAbs* 8, pp. 1259–1268.

Leem, Jinwoo et al. (2018). "Antibody side chain conformations are position-dependent". In: *Proteins: Structure* 86, pp. 383–392.

Lefranc, Marie-Paule and Gérard Lefranc (2001). "The Immunoglobulin FactsBook". In.

Lefranc, Marie-Paule et al. (2003). "IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains." In: *Developmental and comparative immunology* 27 1, pp. 55–77.

Leimkuhler, Benedict J. and Charles Matthews (2016). "Efficient molecular dynamics using geodesic integration and solvent–solute splitting". In: *Proceedings. Mathematical, Physical, and Engineering Sciences / The Royal Society* 472.

Li, Tong et al. (2015). "Rigidity Emerges during Antibody Evolution in Three Distinct Antibody Systems: Evidence from QSFR Analysis of Fab Fragments". In: *PLoS Computational Biology* 11.

Li, Weizhong and Adam Godzik (2006). "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences". In: *Bioinformatics* 22 13, pp. 1658–9.

Lide, David R. (1991). *Handbook of Chemistry and Physics*. 72nd ed. Boca Raton, FL: CRC Press.

Lin, Zeming et al. (2022). "Evolutionary-scale prediction of atomic level protein structure with a language model". In: *bioRxiv*.

Lindorff-Larsen, Kresten et al. (2011). "How Fast-Folding Proteins Fold". In: *Science* 334, pp. 517–520.

Liu, Dong C. and Jorge Nocedal (1989). "On the limited memory BFGS method for large scale optimization". In: *Mathematical Programming* 45, pp. 503–528.

Liu, Xianggen et al. (2020). "Deep geometric representations for modeling effects of mutations on protein-protein binding affinity". In: *PLoS Computational Biology* 17.

Liu, Zhihai et al. (2015). "PDB-wide collection of binding data: current status of the PDBbind database". In: *Bioinformatics* 31 3, pp. 405–12.

Lu, Lenette L. et al. (2017). "Beyond binding: antibody effector functions in infectious diseases". In: *Nature Reviews Immunology* 18, pp. 46–61.

Lu, Ruei-Min et al. (2020). "Development of therapeutic antibodies for the treatment of diseases". In: *Journal of Biomedical Science* 27.

Ma, Puyi, Da-Wei Li, and Rafael Brüschweiler (2023). "Predicting protein flexibility with AlphaFold". In: *Proteins: Structure* 91, pp. 847–855.

Maier, James et al. (2015). "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB." In: *Journal of chemical theory and computation* 11 8, pp. 3696–713.

Makabe, Koki et al. (2008). "Thermodynamic Consequences of Mutations in Vernier Zone Residues of a Humanized Anti-human Epidermal Growth Factor Receptor Murine Antibody, 528*". In: *Journal of Biological Chemistry* 283, pp. 1156–1166.

Marks, Claire and Charlotte M. Deane (2017a). "Antibody H3 Structure Prediction". In: *Computational and Structural Biotechnology Journal* 15, pp. 222–231.

— (2020). "How repertoire data are changing antibody science". In: *The Journal of Biological Chemistry* 295, pp. 9823–9837.

Marks, Claire, Jiye Shi, and Charlotte M. Deane (2018). "Predicting loop conformational ensembles". In: *Bioinformatics* 34, pp. 949–956.

Marks, Claire et al. (2017b). "Sphinx: merging knowledge-based and ab initio approaches to improve protein loop prediction". In: *Bioinformatics* 33, pp. 1346–1353.

Martinez-Valdez, H ctor et al. (1996). "Human germinal center B cells express the apoptosis-inducing genes Fas, c-myc, P53, and Bax but not the survival gene bcl-2". In: *The Journal of Experimental Medicine* 183, pp. 971–977.

McGibbon, Robert T. et al. (2014). "MDTraj: a modern, open library for the analysis of molecular dynamics trajectories". In: *bioRxiv*.

Milstein, Cesar (1999). "The hybridoma revolution: an offshoot of basic research." In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 21 11, pp. 966–73.

Mishra, Arjun K. and Roy A. Mariuzza (2018). "Insights into the Structural Basis of Antibody Affinity Maturation from Next-Generation Sequencing". In: *Frontiers in Immunology* 9.

Misiura, Mikita M et al. (2021). "DLPacker: Deep Learning for Prediction of Amino Acid Side Chain Conformations in Proteins". In: *bioRxiv*.

Mitchell, Laura S. and Lucy J. Colwell (2018). "Comparative analysis of nanobody sequence and structure data". In: *Proteins* 86, pp. 697–706.

Mitsunaga, Erin M and Michael Paul Snyder (2020). "Deep Characterization of the Human Antibody Response to Natural Infection Using Longitudinal Immune Repertoire Sequencing". In: *Molecular & Cellular Proteomics* 19, pp. 278–293.

Miyamoto, Shuichi and Peter A. Kollman (1992). "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". In: *Journal of Computational Chemistry* 13.

Moal, Iain H. and Paul A. Bates (2012). "Kinetic Rate Constant Prediction Supports the Conformational Selection Mechanism of Protein Binding". In: *PLoS Computational Biology* 8.

Mock, Marissa et al. (2023). "AI can help to speed up drug discovery — but only if we give it the right data". In: *Nature* 621, pp. 467–470.

Mullard, Asher (2022). "2021 FDA approvals". In: *Nature Reviews Drug Discovery*.

Muller, Meinard (2007). *Information retrieval for music and motion*. Springer Verlag, pp. 68–84.

Muramatsu, Masamichi et al. (2000). "Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme". In: *Cell* 102, pp. 553–563.

Murphy, Ryan L. et al. (2018). "Janossy Pooling: Learning Deep Permutation-Invariant Functions for Variable-Size Inputs". In: *ArXiv* abs/1811.01900.

Needleman, Saul B. and Christian D. Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." In: *Journal of molecular biology* 48, pp. 443–53.

Norman, Richard A. et al. (2019). "Computational approaches to therapeutic antibody design: established methods and emerging trends". In: *Briefings in Bioinformatics* 21, pp. 1549–1567.

North, Benjamin, Andreas Lehmann, and Roland L. Dunbrack (2011). "A new clustering of antibody CDR loop conformations." In: *Journal of molecular biology* 406 2, pp. 228–56.

Nowak, Jaroslaw et al. (2016). "Length-independent structural similarities enrich the antibody CDR canonical class model". In: *mAbs* 8, pp. 751–760.

Olsen, Tobias Hegelund, Fergus Boyles, and Charlotte M. Deane (2021). "Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences". In: *Protein Science : A Publication of the Protein Society* 31, pp. 141–146.

Olsen, Tobias Hegelund, Iain H. Moal, and Charlotte M. Deane (2022). "AbLang: an antibody language model for completing antibody sequences". In: *Bioinformatics Advances* 2.

Outeiral, Carlos, Daniel A. Nissley, and Charlotte M. Deane (2022). "Current structure predictors are not learning the physics of protein folding". In: *Bioinformatics* 38, pp. 1881–1887.

Ovchinnikov, Victor et al. (2018). "Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies". In: *eLife* 7.

Palma, Francesco Di and Anna Tramontano (2017). "Dynamics behind affinity maturation of an anti-HCMV antibody family influencing antigen binding". In: *FEBS Letters* 591.

Pan, Albert C. et al. (2013). "Molecular determinants of drug-receptor binding kinetics." In: *Drug discovery today* 18 13-14, pp. 667–73.

Parrinello, Michele and Aneesur Rahman (1981). "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied Physics* 52, pp. 7182–7190.

Paszke, Adam et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Neural Information Processing Systems*.

Patel, Jagdish Suresh and F. Marty Ytreberg (2017). "Fast Calculation of Protein–Protein Binding Free Energies Using Umbrella Sampling with a

Coarse-Grained Model". In: *Journal of Chemical Theory and Computation* 14, pp. 991–997.
Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine Learning in Python". In: *ArXiv* abs/1201.0490.
Peng, Zhangzhi et al. (2023). "Generative Diffusion Models for Antibody Design, Docking, and Optimization". In: *bioRxiv*.
Phillips, Angela M. et al. (2021). "Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies". In: *eLife* 10.
Piana, Stefano and Alessandro Laio (2007). "A bias-exchange approach to protein folding." In: *The journal of physical chemistry. B* 111 17, pp. 4553–9.
Ponder, Jay W. et al. (2010). "Current status of the AMOEBA polarizable force field." In: *The journal of physical chemistry. B* 114 8, pp. 2549–64.
Price, Nicholas C. et al. (2011). *Principles and problems in physical chemistry for biochemists*. Oxford Univ. Press.
Puthenveetil, Robbins and Olga Vinogradova (2019). "Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments". In: *The Journal of Biological Chemistry* 294, pp. 15914–15931.
Raghu, Maithra et al. (2016). "On the Expressive Power of Deep Neural Networks". In: *International Conference on Machine Learning*.
Rajewsky, Klaus, Irmgard Förster, and Ana Cumano (1987). "Evolutionary and somatic selection of the antibody repertoire in the mouse." In: *Science* 238 4830, pp. 1088–94.
Raybould, Matthew I. J. et al. (2019a). "Five computational developability guidelines for therapeutic antibody profiling". In: *Proceedings of the National Academy of Sciences of the United States of America* 116, pp. 4025–4030.
Raybould, Matthew I. J. et al. (2019b). "Thera-SAbDab: the Therapeutic Structural Antibody Database". In: *Nucleic Acids Research* 48, pp. D383–D388.
Raybould, Matthew I. J. et al. (2023). "Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling". In: *bioRxiv*.
Regep, Cristian et al. (2017). "The H3 loop of antibodies shows unique structural characteristics". In: *Proteins* 85, pp. 1311–1318.
Robinson, Sarah A. et al. (2021). "Epitope profiling using computational structural modelling demonstrated on coronavirus-binding antibodies". In: *PLoS Computational Biology* 17.
Rock, Edwin P. et al. (1994). "CDR3 length in antigen-specific immune receptors". In: *The Journal of Experimental Medicine* 179, pp. 323–328.
Roth, David B (2014). "V(D)J Recombination: Mechanism, Errors, and Fidelity." In: *Microbiology spectrum* 2 6.
Rouet, Romain et al. (2018). "Next-Generation Sequencing of Antibody Display Repertoires". In: *Frontiers in Immunology* 9.
Roy, Ambrish et al. (2011). "A Protocol for Computer-Based Protein Structure and Function Prediction". In: *Journal of Visualized Experiments : JoVE*.
Ruffolo, Jeffrey A., Jeremias Sulam, and Jeffrey J. Gray (2021). "Antibody structure prediction using interpretable deep learning". In: *Patterns* 3.
Ruffolo, Jeffrey A. et al. (2022). "Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies". In: *Nature Communications* 14.
Ryckaert, Jean-Paul, Giovanni Ciccotti, and Herman J. C. Berendsen (1977). "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints:

Molecular Dynamics of n-Alkanes". In: *Journal of Computational Physics* 23, pp. 327–341.
Salzberg, Steven L. (2018). "Open questions: How many genes do we have?" In: *BMC Biology* 16.
Satorras, Victor Garcia, Emiel Hoogeboom, and Max Welling (2021). "E(n) Equivariant Graph Neural Networks". In: *ArXiv* abs/2102.09844.
Sauer, Marion et al. (2020). "Multi-state design of flexible proteins predicts sequences optimal for conformational change". In: *PLoS Computational Biology* 16.
Scarselli, Franco et al. (2009). "The Graph Neural Network Model". In: *IEEE Transactions on Neural Networks* 20, pp. 61–80.
Schasfoort, Richard B M and Anna J. Tudos (2008). "Handbook of surface plasmon resonance". In.
Schmidt, Aaron G. et al. (2012). "Preconfiguration of the antigen-binding site during affinity maturation of a broadly neutralizing influenza virus antibody". In: *Proceedings of the National Academy of Sciences* 110, pp. 264–269.
Schmidtke, Peter et al. (2011). "Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design." In: *Journal of the American Chemical Society* 133 46, pp. 18903–10.
Schneider, Constantin et al. (2021). "DLAB: deep learning methods for structure-based virtual screening of antibodies". In: *Bioinformatics* 38, pp. 377–383.
Schoeder, Clara T. et al. (2021). "Modeling Immunity with Rosetta: Methods for Antibody and Antigen Design". In: *Biochemistry* 60, pp. 825–846.
Schoop, Andreas and Fabian Dey (2015). "On-rate based optimization of structure-kinetic relationship–surfing the kinetic map." In: *Drug discovery today. Technologies* 17, pp. 9–15.
Schreyer, Adrian and Tom L. Blundell (2009). "CREDO: A Protein–Ligand Interaction Database for Drug Discovery". In: *Chemical Biology & Drug Design* 73.
Sender, Ron et al. (2023). "The total mass, number, and distribution of immune cells in the human body". In: *Proceedings of the National Academy of Sciences of the United States of America* 120.
Senior, Melanie R (2023). "Fresh from the biotech pipeline: fewer approvals, but biologics gain share". In: *Nature Biotechnology* 41, pp. 174–182.
Shaw, David E. et al. (2021). "Anton 3: Twenty Microseconds of Molecular Dynamics Simulation Before Lunch". In: *SC21: International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–11.
Shehata, Laila et al. (2019). "Affinity Maturation Enhances Antibody Specificity but Compromises Conformational Stability." In: *Cell reports* 28 13, 3300–3308.e4.
Shrake, Andrew and John A. Rupley (1973). "Environment and exposure to solvent of protein atoms. Lysozyme and insulin." In: *Journal of molecular biology* 79 2, pp. 351–71.
Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." In: *Science* 228 4705, pp. 1315–7.
Smith, S L (1996). "Ten years of Orthoclone OKT3 (muromonab-CD3): a review." In: *Journal of transplant coordination : official publication of the North American Transplant Coordinators Organization* 6 3, 109–19, quiz 120–1.
Sormanni, Pietro, Francesco Antonio Aprile, and Michele Vendruscolo (2018). "Third generation antibody discovery methods: in silico rational design." In: *Chemical Society reviews* 47 24, pp. 9137–9157.

Spindler, Nadja et al. (2014). "Structural Basis for the Recognition of Human Cytomegalovirus Glycoprotein B by a Neutralizing Human Antibody". In: *PLoS Pathogens* 10.

Spoendlin, Fabian C. et al. (2023). "Improved computational epitope profiling using structural models identifies a broader diversity of antibodies that bind to the same epitope". In: *Frontiers in Molecular Biosciences* 10.

Srivastava, Ashutosh et al. (2018). "Role of Computational Methods in Going beyond X-ray Crystallography to Explore Protein Structure and Dynamics". In: *International Journal of Molecular Sciences* 19.

Srivastava, Nitish et al. (2014). "Dropout: a simple way to prevent neural networks from overfitting". In: *J. Mach. Learn. Res.* 15, pp. 1929–1958.

Steckbeck, Jonathan D. et al. (2005). "Kinetic Rates of Antibody Binding Correlate with Neutralization Sensitivity of Variant Simian Immunodeficiency Virus Strains". In: *Journal of Virology* 79, pp. 12311–12320.

Sultan, Mohammad M and Vijay S. Pande (2017). "tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables." In: *Journal of chemical theory and computation* 13 6, pp. 2440–2447.

Tennenhouse, Ariel et al. (2023). "Computational optimization of antibody humanness and stability by systematic energy-based ranking." In: *Nature biomedical engineering*.

Teplyakov, Alexey et al. (2016). "Structural diversity in a human antibody germline library". In: *mAbs* 8, pp. 1045–1063.

Townsend, Catherine et al. (2016). "Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions". In: *Frontiers in Immunology* 7.

Vaisman-Mentesh, Anna et al. (2020). "The Molecular Mechanisms That Underlie the Immune Biology of Anti-drug Antibody Formation Following Treatment With Monoclonal Antibodies". In: *Frontiers in Immunology* 11.

Van Gunsteren, Wilfred F. and Herman J. C. Berendsen (1988). "A Leap-frog Algorithm for Stochastic Dynamics". In: *Molecular Simulation* 1, pp. 173–185.

Váradi, Mihály et al. (2021). "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models". In: *Nucleic Acids Research* 50, pp. D439–D444.

Vascon, Filippo et al. (2020). "Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design". In: *Computational and Structural Biotechnology Journal* 18, pp. 1774–1789.

Virtanen, Pauli et al. (2019). "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17, pp. 261–272.

Wang, Jing et al. (2022). "A highly accurate metadynamics-based Dissociation Free Energy method to calculate protein–protein and protein–ligand binding potencies". In: *Scientific Reports* 12.

Webb, Ben M. and Andrej Sali (2016). "Comparative Protein Structure Modeling Using MODELLER". In: *Current Protocols in Bioinformatics* 54, pp. 5.6.1–5.6.37.

Weikl, Thomas R. and Fabian Paul (2014). "Conformational selection in protein binding and function". In: *Protein Science* 23.

Weitzner, Brian D., Roland L. Dunbrack, and Jeffrey J. Gray (2015). "The origin of CDR H3 structural diversity." In: *Structure* 23 2, pp. 302–11.

Weitzner, Brian D. et al. (2017). "Modeling and docking of antibody structures with Rosetta". In: *Nature Protocols* 12, pp. 401–416.

Willis, Jordan R. et al. (2013). "Human Germline Antibody Gene Segments Encode Polyspecific Antibodies". In: *PLoS Computational Biology* 9.

Winship, Christopher and Robert D. Mare (1984). "Regression Models With Ordinal Variables". In: *American Sociological Review* 49, p. 512.

Wong, Sergio E., Benjamin D. Sellers, and M. Jacobson (2011). "Effects of somatic mutations on CDR loop flexibility during affinity maturation". In: *Proteins: Structure* 79.

Wong, Wing Ki, Jinwoo Leem, and Charlotte M. Deane (2019). "Comparative Analysis of the CDR Loops of Antigen Receptors". In: *Frontiers in Immunology* 10.

Wong, Wing Ki et al. (2018). "SCALOP: sequence-based antibody canonical loop structure annotation". In: *Bioinformatics* 35, pp. 1774–1776.

Wong, Wing Ki et al. (2020). "Ab-Ligity: identifying sequence-dissimilar antibodies that bind to the same epitope". In: *mAbs* 13.

Wu, Tai Te and Elvin A. Kabat (1970). "An Analysis of the Sequences of the Variable Regions of Bence Jones Proteins and Myeloma Light Chains and Their Implications for Antibody Complementarity". In: *The Journal of Experimental Medicine* 132, pp. 211–250.

Yang, Jianyi et al. (2019). "Improved protein structure prediction using predicted interresidue orientations". In: *Proceedings of the National Academy of Sciences* 117, pp. 1496–1503.

Yaseen, Ashraf et al. (2016). "FLEXc: protein flexibility prediction using context-based statistics, predicted structural features, and sequence information". In: *BMC Bioinformatics* 17.

Yuan, Chao, Hao Chen, and Daisuke Kihara (2012). "Effective inter-residue contact definitions for accurate protein fold recognition". In: *BMC Bioinformatics* 13, pp. 292–292.

Zhu, Kai et al. (2014). "Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction". In: *Proteins: Structure* 82.