Structural studies of interactions between three human pathogens and their host



Samuel F. Gérard

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Abstract

This thesis describes three distinct projects which investigate important aspects of the pathogens that cause acquired immunodeficiency syndrome, Buruli ulcer and intestinal amoebiasis in humans. Through this work, host-pathogen interactions are studied using electron cryo-microscopy.

The first chapter describes how tubular assemblies of the human immunodeficiency virus 1 capsid protein interact with cyclophilin A, a host cellular factor important for regulation of viral restriction. The capsid protein forms asymmetric and curved hexamers, as opposed to previous planar models, and cyclophilin A binds on curved assemblies through a novel non-canonical interface. Hexamer plasticity contributes to variations in local curvature on the capsid, which modulate binding to cyclophilin A and possibly other host factors.

The second chapter shows how the toxin mycolactone, secreted by *Mycobacterium ulcerans*, interacts with the Sec translocon, a protein conducting channel required for translocation of most membrane and secreted proteins. This first structure of the inhibited Sec translocon reveals an unexpected mode of inhibitor binding, in which mycolactone wedges open the lateral gate of Sec at the cytosolic side. This traps the translocon in a partially activated state, in which the lateral gate is partly open but the plug helix seals the channel. The position of mycolactone prevents engagement of signal peptides, blocking complete translocon opening and subsequent protein translocation. Sec appears to be in a dynamic equilibrium, in which lateral gate opening and closing allow recruitment of substrates but also allosteric inhibitors.

The third chapter describes the structure of the 260-kDa surface antigen of *Entamoeba histolytica*, a key protein during invasion of intestinal cells. The structure is unusual and includes domains with folds involved in carbohydrate binding, lipid binding and transport, and protein-protein interactions.

Overall, these studies provide new insights into processes that may be targeted by drugs or vaccines in the fight against infectious diseases.

Acknowledgements

I am extremely grateful to my main supervisor, Pr Matt Higgins, for his support over the course of my DPhil studies. He let me explore and express my own ideas, and knew how to guide me at the right time in my projects. His encouragements to present data at various meetings and conferences and his trust allowed me to develop and thrive as a scientist. I am thankful for all his advice and the constructive and funny discussions we had over the past few years.

I am also very grateful to my supervisor, Pr Peijun Zhang, who took me as a rotation student and introduced me to cryo-EM and tomography. Special thanks to Dr Tao Ni, for his kindness and the scientific discussions we had.

Thanks to all lab members, with whom I had great fun!

I was also very lucky to collaborate with Pr Rachel Chambers, Dr Belinda Hall, Pr Phil Biggin and Dr Afroditi Maria Zaki. I would like to acknowledge other scientists for their help and advice, including Dr Adam Costin, Dr Justin Deme, Dr Edward Löwe, Dr Errin Johnson, Dr Joseph Caesar, Pr Susan Lea, Pr Christian Siebold, Pr Nicole Zitzman, Pr Graham Clark and Pr Richard Farndale.

I extend my thanks to the Wellcome Trust foundation for funding my studies, Louise Samson for her kindness and help with administrative work, and Merton College for its generosity and support.

I also thank all my friends for the fun time spent together in Oxford and back home.

I could not have undertaken this journey without my parents, who keep supporting me in all my endeavours, and my brothers, with whom I have shared invaluable moments. I also present my sincere thankfulness to my grandmother, for all the memories we shared together.

Last, this work would not have been possible without my wife and brilliant scientist, Céline, who gave me unconditional love and support over the past twelve years. I am so proud of what we have built together, and excited about what the future will bring us. Our son, Elian, is a blessing, and I thank him for the joy he brings and for letting us sleep at night.

Publications arising from this work

Ni T, **Gérard S**, Zhao G, Dent K, Ning J, Zhou J, Shi J, Anderson-Daniels J, Li W, Jang S, Engelman AN, Aiken C, Zhang P. (2020). Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A. *Nat Struct Mol Biol.*, 27(9):855-862. doi: 10.1038/s41594-020-0467-8.

Gérard SF, Hall BS, Zaki AM, Corfield KA, Mayerhofer PU, Costa C, Whelligan DK, Biggin PC, Simmonds RE, Higgins MK. (2020). Structure of the Inhibited State of the Sec Translocon. *Mol Cell.*, 79(3):406-415.e7. doi: 10.1016/j.molcel.2020.06.013.

Some material from these papers is reproduced in modified form in this thesis. The work of other authors is acknowledged where appropriate. These publications appear in Appendix B.

Contents

Abstract i

Acknowledgements ii

Publications arising from this work iii

List of Figures ix

List of Tables xii

List of Abbreviations xiii

1 Helical processing of HIV-1 capsid protein bound to cyclophilin A 1

1.1 Summary . . . 1

1.2 Introduction . . . 2

1.2.1 HIV burden worldwide . . . 2

1.2.2 Life cycle and replication of HIV-1 . . . 2

1.2.3 Formation of HIV-1 mature capsid . . . 3

1.2.4 Structural features of HIV-1 mature capsid . . . 4

1.2.4.1 Structures of HIV-1 CA building blocks . . . 5

1.2.4.2 Local curvature adopted by CA assemblies . . . 8

1.2.5 Roles of HIV-1 CA at the host-pathogen interface . . . 8

1.2.5.1 A protective shell that maintains viral genome integrity 8

1.2.5.2 A binding platform for cellular host factors . . . 9

1.2.6 CypA interaction with HIV-1 CA . . . 12

1.2.7 Aims . . . 15

1.3 Materials and methods . . . 15

1.3.1 Sample preparation and data collection . . . 15

1.3.2 Helical reconstruction of tubular assemblies . . . 16

1.3.2.1 Preparation and sorting of tubes . . . 16

1.3.2.2 Determination of helical symmetry . . . 16

1.3.2.3 Classifications and refinement of tubular assemblies . . 25

1.3.3 Localised reconstructions of subregions of interest . . . 26

1.3.4 Model building, refinement and data availability 28
1.3.5 Analysis of conservation of CypA residues 28
1.4 Results 28
1.4.1 Helical reconstruction of HIV-1 CA tubes in complex with CypA 28
1.4.2 Localised reconstruction of CA hexamer 32
1.4.3 Localised reconstruction and classification of CypA bound to CA 35
1.5 Discussion and future directions 39

References **42**

2 Structure of the inhibited state of the Sec translocon **49**
2.1 Summary 49
2.2 Introduction 50
2.2.1 The importance of translocation in protein biogenesis 50
2.2.1.1 Protein synthesis by the ribosome factory 50
2.2.1.2 The endoplasmic reticulum: a central organelle for biogenesis of secretory and membrane proteins 52
2.2.1.3 Protein targeting to the ER: segregation of secretory and membrane proteins 54
2.2.2 The central role of the protein translocation channel 55
2.2.2.1 Identification of the protein conducting channel 55
2.2.2.2 An evolutionary conserved complex 56
2.2.2.3 A dynamic channel 58
2.2.3 Study of Sec61 protein inhibition 62
2.2.3.1 A shared mechanism of action? 62
2.2.3.2 Mycolactone: a unique inhibitor 64
2.2.4 Project aims 66
2.3 Materials and methods 67
2.3.1 Synthetic mycolactone 67
2.3.2 Purification of ribosome-translocon complexes (RTCs) in detergent 67
2.3.3 Extraction of RTCs in nanodiscs and diisobutylene/maleic acid lipid particles 68
2.3.4 SDS-PAGE and immunoblotting 69
2.3.5 Grid preparation and data acquisition 70
2.3.6 Image processing 71
2.3.7 Homology modelling and refinement of atomic models 72
2.3.8 *In vitro* translation/translocation assays 73
2.3.9 Liquid chromatography mass spectrometry 73
2.3.10 Forward genetic screen, viability assays and purification of microsomes from mycolactone-resistant cell lines 74

2.3.11 Molecular dynamics simulations 74
2.4 Results 75
2.4.1 Preparation of mammalian RTCs for cryo-EM 75
2.4.2 Optimisation of grid support 76
2.4.3 Cryo-EM processing of RTCs in the presence and absence of mycolactone 78
2.4.4 Structure of the inhibited state of Sec61 80
2.4.5 Differences between the inhibited Sec translocon and other states 81
2.4.6 Mode of binding of mycolactone to Sec61 83
2.4.7 Mutational studies of Sec61 resistance to mycolactone 86
2.5 Discussion and perspectives 89
2.5.1 An unexpected mode of binding 89
2.5.2 Mycolactone prevents engagement of signal peptides 89
2.5.3 Implications for the dynamics of translocation 90
2.5.4 Mode of action of other inhibitors 91
2.5.5 Perspectives 92

References **93**

3 Structural studies of the 260-kDa antigen of Entamoeba histolytica 102
3.1 Summary 102
3.2 Introduction 103
3.2.1 Intestinal infection by *Entamoeba histolytica* 103
3.2.2 The 260-kDa surface antigen of *Entamoeba histolytica* 104
3.2.2.1 Identification of the 260-kDa antigen 104
3.2.2.2 Functions of the 260-kDa antigen 105
3.2.3 Interacting partners of the 260-kDa antigen 109
3.2.3.1 The intermediate subunit of 150 kDa (IgL) 109
3.2.3.2 Other potential extracellular interacting partners . . . 111
3.2.4 Prospect for a vaccine against amoebiasis 111
3.2.5 Aims 113
3.3 Materials and methods 113
3.3.1 Construct design and molecular cloning 113
3.3.1.1 Constructs of heavy, light and intermediate subunits . 114
3.3.1.2 Constructs of fragments of antibodies 114
3.3.1.3 Constructs for bacterial refolding of HgL putative CRD and LgL 115
3.3.1.4 Construct of HgL putative VWD fused to glutathione S-transferase 115
3.3.2 Protein expression, purification and modification 116

3.3.2.1 Protein expression and purification from Expi293F cells 116
3.3.2.2 Bacterial refolding of HgL putative CRD and LgL . . . 116
3.3.2.3 Bacterial expression of GST-fused HgL VWD 117
3.3.2.4 SDS-PAGE and immunoblotting 118
3.3.2.5 Protein methylation 118
3.3.2.6 Biotinylation of Fab fragments 118
3.3.2.7 Limited proteolysis and mass spectrometry analysis . . 119
3.3.3 Purification of the 260-kDa antigen from amoebae 120
3.3.3.1 Culture of trophozoites in axenic conditions 120
3.3.3.2 Purification of the 260-kDa antigen 122
3.3.4 Functional studies of the 260-kDa antigen 123
3.3.4.1 Interaction with antibodies and ligands by surface plasmon resonance 123
3.3.4.2 Enzyme-linked immunosorbent assay of HgL VWD binding to collagens 124
3.3.4.3 Size-exclusion chromatography multi-angle light scattering of the recombinant lectin heterodimer 125
3.3.4.4 SAXS analysis of constructs of the heavy chain (HgL) 125
3.3.5 X-ray crystallography of recombinant proteins 126
3.3.5.1 Crystallisation trials 126
3.3.5.2 Crystallisation of HgL VWD, CP33 scFv and Fab . . . 126
3.3.5.3 Data collection and structure determination 127
3.3.6 Cryo-EM of the 260-kDa and 150-kDa antigens 128
3.3.6.1 Cryo-EM of the native 260-kDa antigen 128
3.3.6.2 Cryo-EM of the cysteine-rich region of the 260-kDa antigen 132
3.3.6.3 Cryo-EM of IgL 133
3.4 Results 134
3.4.1 Recombinant expression of fragments of the amoebic antigen and antibodies for X-ray crystallography 134
3.4.1.1 Structural homology modelling, bioinformatic analyses and construct design 134
3.4.1.2 Protein production and modification of amoebic constructs 137
3.4.1.3 Use of antibody fragments as chaperones for crystallisation 143
3.4.1.4 Crystal structure of HgL VWD 148
3.4.1.5 Purification of the heterodimer of LgL in complex with HgL 149
3.4.2 Purification of the native amoebic antigen 151
3.4.2.1 Establishment of axenic culture of *Entamoeba histolytica* 151

3.4.2.2 Isolation of the native 260-kDa antigen 152
3.4.3 Cryo-EM studies of the native 260-kDa antigen 154
3.4.3.1 Negative staining 154
3.4.3.2 Optimisation of specimen for data acquisition 154
3.4.3.3 Cryo-EM processing 156
3.4.3.4 Partial structure of the 260-kDa heterodimer 157
3.4.3.5 Preliminary cryo-EM studies of HgL C-terminus 166
3.4.3.6 Preliminary cryo-EM studies of the recombinant intermediate chain 168
3.4.3.7 Functional studies of the 260-kDa antigen 170
3.5 Discussion and perspectives 175
3.5.1 The 260-kDa antigen has a distinctive architecture 175
3.5.2 Robust expression of the amoebic antigen in mammalian cells . 176
3.5.3 Putative functions of the 260-kDa antigen 177

References **181**

General Conclusion **191**

Appendices **194**

A Tables and Figures **195**

B Publications **201**

List of Figures

1.1 Simplified life cycle of HIV-1 retrovirus. 3
1.2 Maturation of HIV-1 virions. 4
1.3 A model of the mature capsid. 5
1.4 Structural features of HIV-1 CA protomer and capsomers. 6
1.5 Dimer and trimer interfaces between capsomers. 7
1.6 HIV-1 capsid is a binding platform for cellular host factors. 11
1.7 Current structural knowledge on CypA binding to HIV-1 CA assemblies. 13
1.8 Helical parameters and the 2D lattice. 17
1.9 Fourier transform of the projection of a helix. 18
1.10 Differences between continuous and discontinuous helices in Fourier space. 20
1.11 Indexing of Bessel orders in Fourier space. 21
1.12 Determination of screw symmetry. 23
1.13 Validation of screw symmetry by IHRSR. 25
1.14 Sorting of CA tubes for 3D reconstruction. 29
1.15 Examples of determination of screw symmetry. 30
1.16 3D reconstruction from helical parameters. 31
1.17 Localised reconstruction of a hexamer. 33
1.18 A curved hexamer. 34
1.19 Density subtraction at the dimer interface. 36
1.20 Resolution and binding mode of CypA at the dimer interface. 37
1.21 Binding of CypA at the dimer interface. 38
1.22 Localised reconstruction of the trimer interface. 39
1.23 CypA binding is governed by local curvature. 41

2.1 Eukaryotic protein synthesis by the ribosome. 51
2.2 Conventional protein secretory pathway in eukaryotes. 53
2.3 Co-translational and post-translational translocations. 55
2.4 Structural features of the Sec translocon. 57
2.5 Dynamics of Sec61 during co-translational translocation. 59
2.6 Dynamics of Sec61 during post-translational translocation. 61
2.7 Gating inhibitors of Sec61α. 63
2.8 Mycolactone and Buruli ulcer. 65

2.9 How does inhibition work? 67
2.10 Purification of RTCs from canine microsomes in the presence and absence of mycolactone. 76
2.11 Optimisation of grid support for cryo-EM data collection. 77
2.12 Cryo-EM processing of RTCs in the absence of mycolactone. 79
2.13 Cryo-EM processing of RTCs in the presence of mycolactone. 80
2.14 Structural features of the inhibited state of Sec61. 81
2.15 Comparison of Sec61 inhibited conformation with other states adopted during translocation. 82
2.16 Insertion trials of RTCs into lipids. 84
2.17 Mycolactone binding site. 86
2.18 Sec61α resistance mutations to mycolactone and their effect on mycolactone binding. 87
2.19 Model of inhibition of co-translational translocation. 91

3.1 Infection by *Entamoeba histolytica* 104
3.2 Schematic representation of the 260-kDa surface heterodimer. 105
3.3 Colonic mucins and their O-glycans. 106
3.4 Mapping of antibodies that affect adherence and similarities with integrins. 108
3.5 A bigger surface complex. 110
3.6 Selective biotinylation of Fabs 119
3.7 Parasite culture of *Entamoeba histolytica* HM1:IMSS 121
3.8 Constructs of HgL for recombinant expression. 135
3.9 LgL shares homology with β-trefoil domains. 136
3.10 Constructs of IgL for recombinant expression. 137
3.11 Protein production in Expi293F cells and modification. 139
3.12 Limited proteolysis of HgL_05 and HgL_08 coupled with MS. 140
3.13 New constructs based on limited proteolysis. 142
3.14 Refolding trials of HgL putative CRD and LgL predicted β-trefoil domains. 143
3.15 Purification of fragments of antibodies. 144
3.16 CP33 binding to HgL. 145
3.17 Structures of CP33 Fab and scFv. 146
3.18 Complex formation with antibodies for crystallisation trials. 147
3.19 Limited proteolysis of scFv-bound HgL_13. 147
3.20 Crystal structure of HgL VWD. 149
3.21 Recombinant expression of LgL:HgL heterodimer in Expi293F cells. . . 150
3.22 Purification of the 260-kDa native antigen in detergent. 152
3.23 Purification of the 260-kDa native antigen in lipids. 153
3.24 Negative staining after immuno-affinity chromatography. 154
3.25 Sample optimisation for cryo-EM data collection. 155

3.26 Processing pipeline and resolution estimates. 157
3.27 Structure of the N-terminal lipoprotein-like domain. 159
3.28 HgL VWD in the cryo-EM reconstruction. 160
3.29 Structure of the cysteine-rich pseudo-domains. 161
3.30 An elongated protrusion at the cysteine-rich region. 163
3.31 Structure of the light chain. 165
3.32 Construct design of HgL C-terminus and complex formation with CP33 Fab. 166
3.33 Cryo-EM studies of the C-terminal half of HgL in complex with CP33 Fab. 168
3.34 EM studies of IgL_01 with and without Fabs. 169
3.35 Cryo-EM study of IgL_01. 170
3.36 ELISA-type binding assay of HgL VWD to collagen. 171
3.37 Interaction with complement proteins and IgL. 172
3.38 SPR of the 260-kDa antigen with a panel of mAbs. 173
3.39 SAXS analysis of HgL. 174
3.40 Putative functions of the 260-kDa antigen. 178

A.1 SEC-MALS of HgL_05 and HgL_09. 200
A.2 Purification trials of GST-fused LgL. 200

List of Tables

2.1 Refinement and model statistics. 73

3.1 Current crystallographic statistics. 128
3.2 Cryo-EM settings for imaging of the 260-kDa antigen. 130
3.3 Cryo-EM settings for imaging of HgL_20. 132
3.4 Cryo-EM settings for imaging of IgL. 133

A.1 List of primers. 197
A.2 Crystallisation trials. 198
A.3 Mass spectrometry of in-gel tryptic digest of native HgL. 199

List of Abbreviations

2D . . . . . two-dimensional

3D . . . . . three-dimensional

A-site . . . . . aminoacyl-site

AIDS . . . . . acquired immune deficiency syndrome

ALA . . . . . amoebic liver abscess

AspN . . . . . flavastacin

BSA . . . . . bovine serum albumin

CA . . . . . capsid protein

CHAPS . . . . cholamidopropyl dimethylammonio 1-propanesulfonate

CHO . . . . . Chinese hamster ovary

CMC . . . . . critical micelle concentration

CRD . . . . . carbohydrate recognition domain

Cryo-EM . . . electron cryo-microscopy

Cryo-ET . . . electron cryo-tomography

CTD . . . . . C-terminal domain

CTF . . . . . contrast transfer function

CypA . . . . . cyclophilin A

DIBMA . . . . diisobutylene-maleic acid

DMSO . . . . . dimethylsulfoxide

dNTP . . . . . deoxynucleoside triphosphate

DTT . . . . . dithiothreitol

E. coli . . . . . *Escherichia coli*

E-site . . . . . exit-site

eEF . . . . . elongation factor

EMDB . . . . electron microscopy data bank

ER . . . . . endoplasmic reticulum

Fab . . . . . fragment antigen binding

Gag . . . . . HIV-1 group specific antigen

Gal . . . . . galactose

GalNAc . . . . N-acetylgalactosamine

GluC . . . . . staphylococcus aureus protease V8

GPI . . . . . glycosylphosphatidylinositol

GST . . . . . glutathione S-transferase

GTP . . . . . guanosine triphosphate

GTPase . . . . guanosine triphosphate hydrolase

HEK . . . . . human embryonic kidney

HgL . . . . . heavy subunit of the 260-kilodalton amoebic antigen

HIV . . . . . human immunodeficiency virus

IgL . . . . . intermediate subunit of 150 kilodalton

IHRSR . . . . iterative helical real-space reconstruction

IMAC . . . . . immobilised metal ion affinity chromatography

kDa . . . . . kilodalton

LC-MS . . . . liquid chromatography mass spectrometry

LgL . . . . . light subunit of the 260-kilodalton amoebic antigen

LMNG . . . . lauryl maltose neopentyl glycol

MA . . . . . matrix protein

mAb . . . . . monoclonal antibody

MIDAS . . . . metal ion-dependent adhesion site

MR . . . . . molecular replacement

mRNA . . . . messenger ribonucleic acids

MS . . . . . mass spectrometry

MSP . . . . . membrane protein scaffold

NC . . . . . nucleocapsid

Ni-NTA . . . . nickel-nitriloacetic acid

NMR . . . . . nuclear magnetic resonance

NTD . . . . . N-terminal domain

OG . . . . . octyl β-D-glucopyranoside

P-site . . . . . peptidyl-site

PCR . . . . . polymerase chain reaction

PDB . . . . . protein data bank

PEG . . . . . polyethylene glycol

prl . . . . . protein localisation

PTC . . . . . peptidyl-transferase centre

Rg . . . . . radius of gyration

RMSD . . . . . root mean square deviation

RNA . . . . . ribonucleic acid

RNC . . . . . ribosome-nascent chain complex

RTC . . . . . ribosome-translocon complex

RU . . . . . response unit

scFv . . . . . single-chain variable fragment

SDS-PAGE . . sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEC . . . . . size-exclusion chromatography

SP . . . . . small peptide linker

SPR . . . . . surface plasmon resonance

SR . . . . . signal recognition particle receptor

SRP . . . . . signal recognition particle

TEV . . . . . tobacco etch virus

TM . . . . . transmembrane

tRNA . . . . . transfer ribonucleic acids

VWD . . . . . von Willebrand-like domain

WT . . . . . wild type

1

Helical processing of HIV-1 capsid protein bound to cyclophilin A

1.1 Summary

Viruses are obligate intracellular microbes that take advantage of host cellular resources to replicate and survive. By doing so, pathogenic viruses, such as the human immunodeficiency virus (HIV), trigger significant damage to their host. Successful replication of HIV relies on its mature capsid, whose integrity is regulated by interactions with diverse cellular host factors. Among these, cyclophilin A (CypA) is important for infection and trafficking and regulates restriction of viral replication. In this first chapter, electron cryo-microscopy (cryo-EM) structures of tubular assemblies of HIV-1 capsid protein (CA) in the presence of CypA show that CypA binds simultaneously to a canonical and a newly described non-canonical interface on CA tubes. Remarkably, the structure of CA hexamer is curved and asymmetric, and this plasticity contributes to formation of local curvature on HIV-1 capsid, which is sensed by host factors such as CypA.

1.2 Introduction

1.2.1 HIV burden worldwide

HIV is a retrovirus which infects human immune cells, primarily $CD4^+$ T cells, and impairs their function [1]. Over time and untreated, infection leads to continuous depletion of the immune system and eventually causes acquired immunodeficiency syndrome (AIDS). People affected by AIDS have a weak immune system and are therefore more vulnerable to opportunistic infections and cancers. Since the start of the pandemic and by the end of 2021, 36.3 million deaths from AIDS-related illnesses were estimated by the Joint United Nations Programme on HIV/AIDS.

There is currently no cure or vaccine against HIV. Instead, current treatments consist of a combination of anti-retroviral drugs which inhibit HIV replication. However, there is evidence of resistance caused by mutations of the retrovirus that modulate the effect of current drugs on viral replication [2]. Therefore, diseases caused by HIV remain a global issue and the development of novel anti-retroviral therapies is desirable to avert AIDS-related deaths and prevent infections.

1.2.2 Life cycle and replication of HIV-1

HIV-1 is the most virulent and infective type of HIVs, and accounts for the majority of infections. During the early phase of HIV-1 replication cycle (Figure 1.1), attachment of a mature retrovirus to host cell surface receptors precedes fusion of viral and cellular membranes, eventually followed by delivery of viral content into the cytoplasm of the host cell. The mature conical capsid that contains viral ribonucleic acid (RNA) and replication proteins is released and trafficked towards the cell nucleus. There, reverse-transcribed viral RNA integrates into its host genome, allowing production of viral proteins in the cytoplasm and release of new virions in the extracellular environment, which mature and infect new target cells.

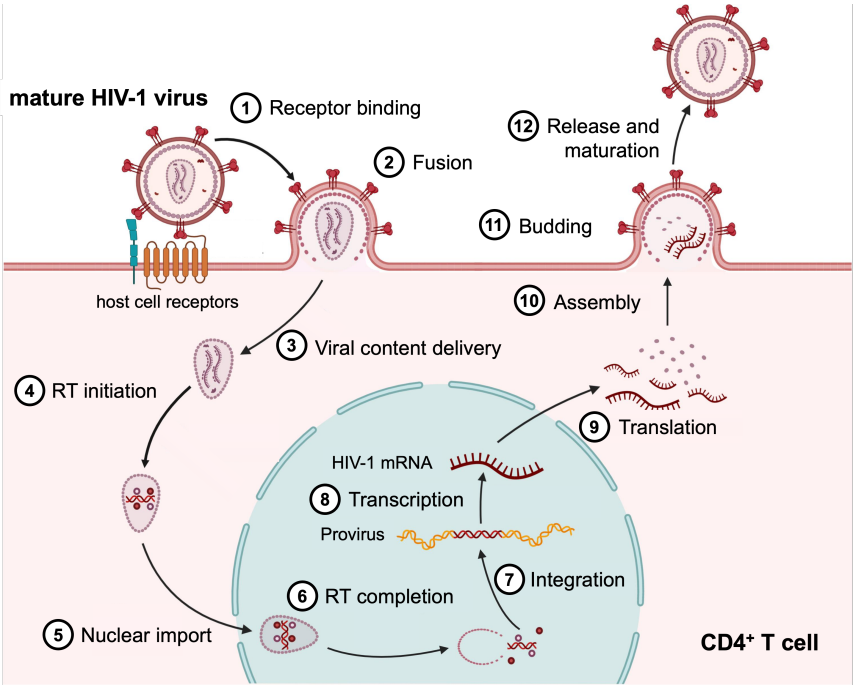


Figure 1.1: Simplified life cycle of HIV-1 retrovirus. A mature virus binds to cell surface receptors of a $CD4^+$ T cell (step 1). After membrane fusion (step 2) and delivery of the viral content (step 3), the capsid enclosing viral RNA and accessory proteins houses reverse transcription (RT, step 4) and is shuttled and imported to the nucleus (step 5). There, reverse-transcription is completed (step 6) and viral DNA is integrated into the genome of the host cell (step 7). Transcription and translation by the host cell (steps 8 and 9) lead to production of HIV proteins that assemble into immature viruses that bud out of the infected cell (steps 10 and 11). Eventually, immature viruses are released in the extracellular environment where they mature (step 12) to repeat a new cycle of infection (picture taken from [3]).

1.2.3 Formation of HIV-1 mature capsid

During infection, HIV-1 viral assembly undergoes a series of dramatic morphological changes [4]. These are driven by processing of the HIV-1 group specific antigen (Gag). Gag consists of four domains: matrix (MA, 132 residues), capsid (CA, 231 residues), nucleocapsid (NC, 54 residues) and p6, joined by smaller peptide linkers SP1 (between CA-NC) and SP2 (between NC-p6, Figure 1.2.a).

Gag is processed by the viral protease and cleaved proteins are rearranged to form mature and infectious viral particles (Figures 1.2.b-c). These display MA at the inner surface of their membranes, whereas CA proteins assemble to form a capsid shell which encapsulates viral integrase, reverse transcriptase and NC in complex with viral RNA.

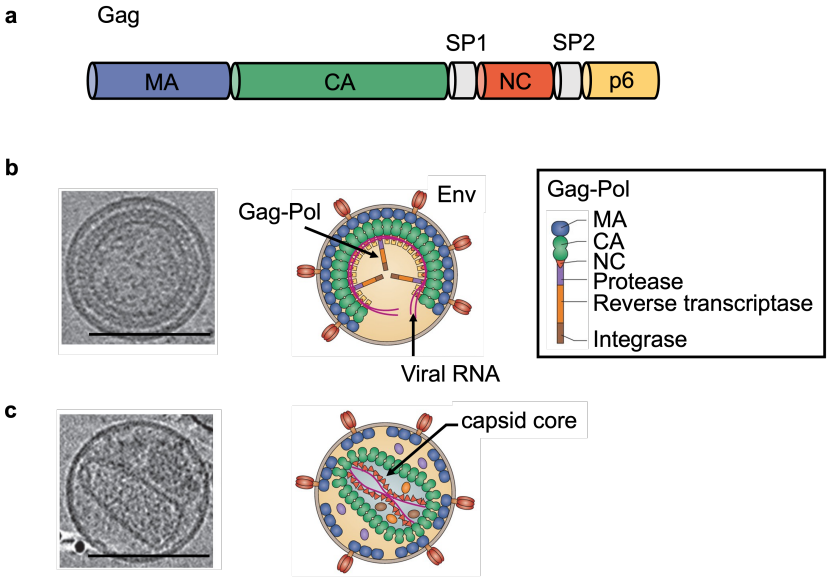


Figure 1.2: Maturation of HIV-1 virons. a) Domain organisation of HIV-1 Gag. The four domains (MA, CA, NC and p6) and linkers (SP1 and SP2) are coloured. b) Cryo-electron tomogram and schematic representation of an immature particle. Scale bar is equal to 100 nm. c) Cryo-electron tomogram and schematic representation of a mature particle. Scale bar is equal to 100 nm. This figure is adapted from [4].

1.2.4 Structural features of HIV-1 mature capsid

HIV-1 mature capsid is assembled by 1,000-1,500 copies of the same capsid protein CA. CA is essential for infection, with many single-point mutations resulting in non-infectious particles with often aberrant capsid morphologies [5–7]. In high salt, CA-NC was shown to form cones *in vitro*. Using electron microscopy, cone angles were measured and based on Euler's theorem and structural analogy with element carbon, CA was proposed to form a surface array of approximately 200-250 hexamers and 12 pentamers (capsomers), which assemble into a fullerene cone [7] (Figure 1.3), with an average length of 103 nm and average diameters of 25 nm and 52 nm on opposite ends [8].

Remarkably, this peculiar structure is asymmetric, with notorious changes in local curvatures. Various studies have provided structural details of CA assemblies, and more recent works have focused on the mechanistic of their arrangement into an asymmetric capsid.

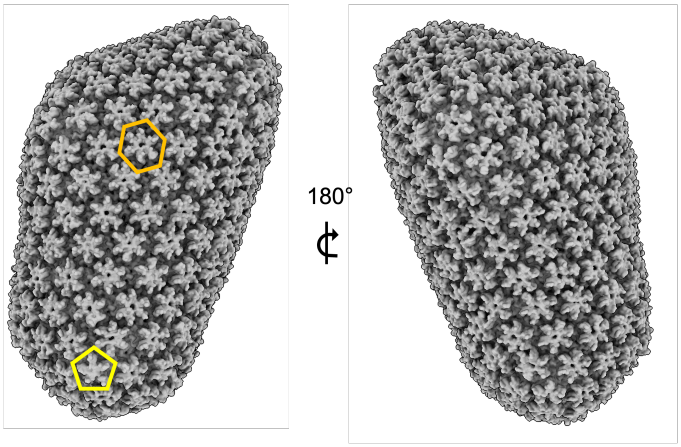


Figure 1.3: Model of HIV-1 mature capsid. The capsid is asymmetric and composed of CA, arranged into hexamers and pentamers (protein data bank (PDB) code 3J3Y). A hexamer and a pentamer are highlighted in orange and yellow, respectively.

1.2.4.1 Structures of HIV-1 CA building blocks

At the monomeric level, nuclear magnetic resonance (NMR) and crystal structures of CA domains [9, 10] and antibody-bound full-length CA [9, 11] first revealed a 231-residue alpha-helical protein that consists of an arrowhead-shaped 7-helix N-terminal domain (NTD) joined to a globular 4-helix C-terminal domain (CTD) by a flexible linker (Figure 1.4.a). A β-hairpin formed by the first 13 residues of CA_{NTD} forms a salt bridge [12] that is essential for capsid assembly and viral infectivity [13, 14]. CA_{NTD} also includes a CypA-binding loop involved in the interaction with the cytoplasmic host factor CypA [10].

CA is able to form dimers and higher-order oligomers in solution. For instance, purified HIV-1 CA forms tubular assemblies of various length at neutral pH and high salt concentrations *in vitro*, suggesting weak hydrophobic interactions as a driving force for oligomerisation. In particular, low-resolution cryo-EM studies of these tubular assemblies highlighted the formation of hexamers, with CA_{NTDs} involved in intra-hexamer interactions, and CA_{CTDs} at the interface between two adjacent hexamers [15]. This model was improved by docking of NMR and crystal structures into a 9 Å map obtained by electron crystallography of two-dimensional (2D) crystals [16].

Later, crosslinks between CA protomers and mutations at CA_{CTD} to prevent association between capsomers allowed determination of crystal structures of CA hexamer [17] and pentamer [18] at higher resolution. In both structures, CA_{CTDs} surround an inner

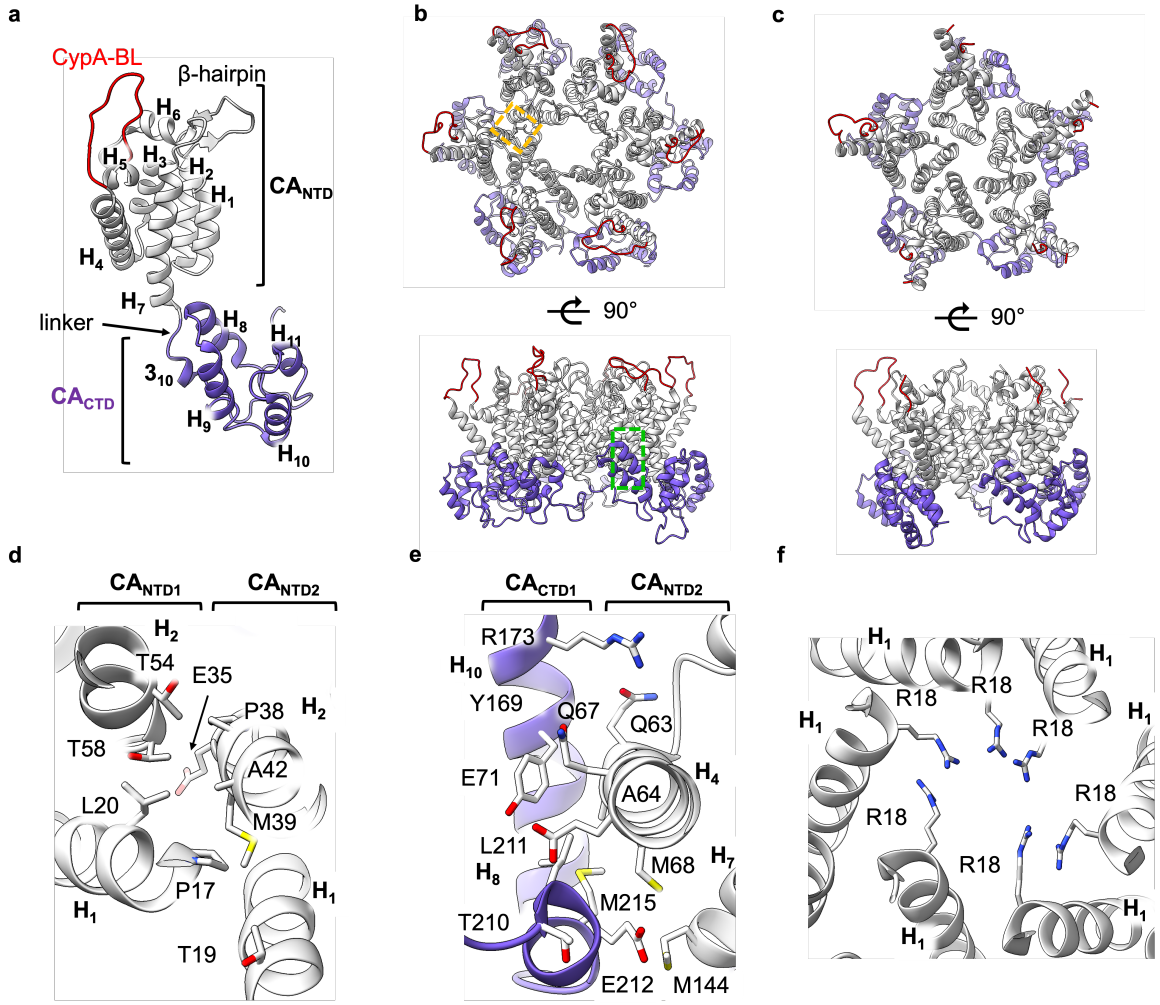


Figure 1.4: Structural features of HIV-1 CA protomer and capsomers. a) Ribbon representation of HIV-1 CA monomer (PDB code 3J34). CA_{NTD}, CA_{CTD} and CypA binding loop (BL) are coloured in grey, slate blue and red, respectively. The same colour code is used from b) to g). b) Ribbon representation of CA hexamer (PDB code 3J34, top and side views). The regions for the CA_{NTD}-CA_{NTD} and CA_{CTD}-CA_{NTD} interfaces of two adjacent CA monomers are outlined in orange and green, respectively. c) Ribbon representation of CA pentamer (PDB code 3P05, top and side views). d) Ribbon representation of the CA_{NTD}-CA_{NTD} interface between adjacent monomers of a hexamer (PDB code 3J34). Most interactions are made by small aliphatic and hydrophobic residues. e) Ribbon representation of the CA_{CTD}-CA_{NTD} interface between adjacent monomers of a hexamer (PDB code 3J34). f) Ribbon representation of the ring formed by arginine residues 18 of six CA_{NTDs} of a hexamer (PDB code 3J34).

ring, formed by association of the first three helices of each CA_{NTD} into 18-helix (hexamer, Figure 1.4.b) and 15-helix bundles (pentamer, Figure 1.4.c). The inner rings use identical aliphatic residues which mediate hydrophobic interactions between CA_{NTDs} (Figure 1.4.d). In addition, hydrogen bonding interactions between CA_{CTD} and CA_{NTD} of adjacent protomers were identified in both capsomers (Figure 1.4.e). Outer CA_{CTDs} were found

to be more mobile than rigid inner CA_{NTDs}, suggesting a mechanism of local curvature formation shared by pentamers and hexamers. Last, the central core of capsomers is positively charged and formed by five or six arginine (R18) residues (Figure 1.4.f).

Then, the crystal structure of a hexagonal lattice formed by seven wild-type (WT) hexamers [19], together with intermediate-resolution cryo-EM structures of *in vitro* tubular assemblies [20] and electron cryo-tomography (cryo-ET) structures of intact capsids [21] highlighted key interfaces formed between capsomers at higher resolution (Figure 1.5.a-b). In particular, in addition to the dimer interface formed by CA_{CTDs} of adjacent hexamers, a new interface between three adjacent hexamers was shown to be essential for capsid assembly, rigid-body tilting and twisting between capsomers. Both dimer and trimer interfaces are defined by hydrophobic interactions between CA_{CTDs} (Figure 1.5.b-c).

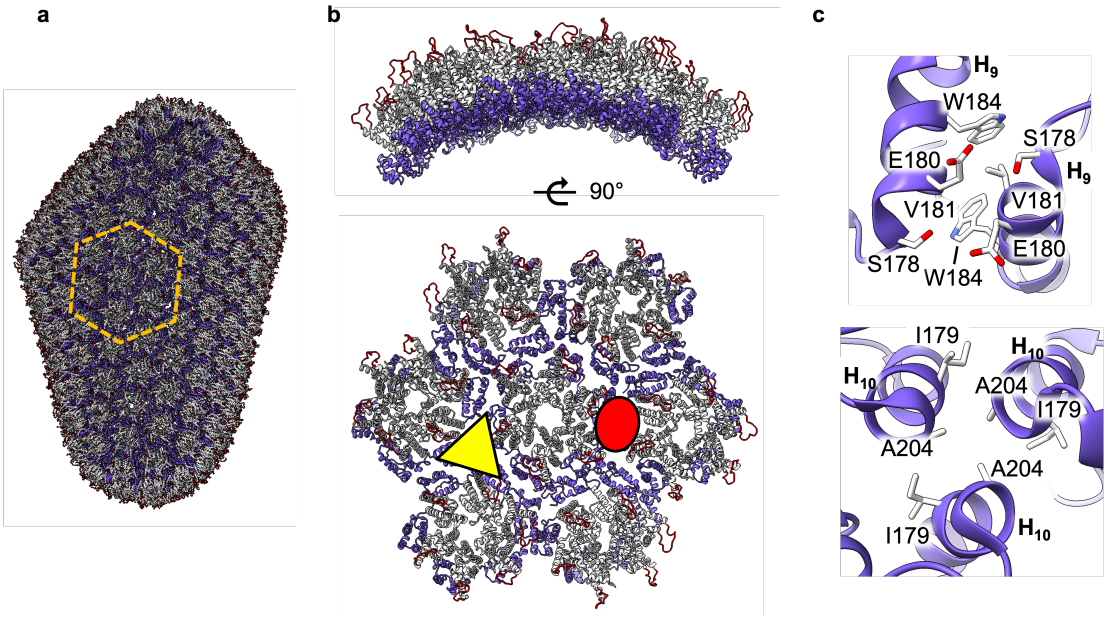


Figure 1.5: Dimer and trimer interfaces between capsomers. a) Ribbon representation of a model of HIV-1 CA core (PDB code 3J3Y). A heptamer of hexamers is delimited in orange. CA_{NTD}, CA_{CTD} and CypA binding loop are coloured in grey, slate blue and red, respectively. This colour code is used in b) and c). b) Ribbon representation of HIV-1 CA heptamer of hexamers (PDB code 3J34, top and side views). The side view shows how these arrange on a tubular assembly. Locations of trimer and dimer interfaces between three and two adjacent hexamers are highlighted in yellow and red, respectively. c) Ribbon representation of the dimer (top) and trimer (bottom) interfaces between adjacent hexamers (PDB code 3J34). Residues at the interface are mainly hydrophobic.

In total, several CA-CA interfaces on the HIV-1 mature capsid were shown to be

essential for its stability. These include the six- or five-fold symmetric CA_{NTD}-CA_{NTD} interface at the centre of hexamers or pentamers, respectively, the inter-molecular CA_{NTD}-CA_{CTD} interface between adjacent monomers, and inter-capsomer two-fold and three-fold CA_{CTD}-CA_{CTD} interfaces.

1.2.4.2 Local curvature adopted by CA assemblies

In the mature capsid, the CA lattice bends to adopt variable curvatures. It has been proposed that variations in the orientations of CA_{CTDs} between planar capsomers at the dimer and trimer interfaces account for local curvature [18, 22, 23]. This appears to be driven by movement of the flexible linker between CA_{NTD} and CA_{CTD} of each protomer.

However, several caveats to current structures limit our understanding of capsid assembly. First, all high-resolution crystal structures obtained for capsomers are planar. Second, the current details provided by cryo-EM structures of tubular assemblies and cryo-ET structures of intact capsids are limited due to their moderate resolution. Therefore, high resolution structures of capsomers in curved assemblies are needed to provide molecular details of local geometry on native capsids. Furthermore, such structures have functional relevance since it is the curved capsid lattice, rather than CA protomers, which forms the binding platform for various host factors during viral infection [24, 25].

1.2.5 Roles of HIV-1 CA at the host-pathogen interface

From entry into a target cell to delivery of the viral genome into the host nucleus, infection is highly dependent on HIV-1 capsid. The structural integrity of the shell is tightly regulated by a variety of host factors and depends on reverse transcription of viral RNA. For example, some host factors interact with the capsid and mediate its disassembly, in a process called uncoating. Since some of these interactions are important for infection, the capsid is an interesting target for development of antivirals.

1.2.5.1 A protective shell that maintains viral genome integrity

Successful replication implies that HIV-1 is able to transfer its genome into its host nucleus. To achieve this, the capsid encapsulates the viral genome and accessory proteins in a closed environment. There, it houses reverse-transcription of viral RNA

and provides a barrier that prevents detection and targeting of the viral genome by host cytoplasmic proteins.

Reverse transcription of a viral single-stranded RNA into a double-stranded DNA is crucial during infection. The reaction is carried out by HIV-1 reverse transcriptase and initiated inside the capsid. Completion of reverse transcription can occur in the closed capsid and after the capsid starts to uncoat [26]. As the reaction requires cytoplasmic deoxynucleoside triphosphates (dNTPs), these must be recruited inside the shell. This import is mediated by the electropositive channel formed by arginine residues at the centre of CA hexamers [27]. There, dNTPs interact with the pore with high affinity and are translocated through the channel, which can switch from closed to open conformations through displacement of the β-hairpin. This mechanism might be conserved in other lentiviral capsids, since they share a similar positively charged channel. Importantly, dNTP import and subsequent reverse transcription can be blocked *in vitro* using a small channel inhibitor [27].

The effects of reverse transcription on HIV-1 capsid stability has also been studied. For instance, inhibition of reverse transcription delayed capsid uncoating by increasing its stability [28, 29]. Likewise, reverse transcription into rigid filamentous DNA material was suggested to act as a mechanical initiator of uncoating, by increasing pressure inside the capsid [30]. Some cellular host proteins, such as the eukaryotic elongation factor (eEF) 1 [31], appear to be critical for efficient completion of reverse transcription, also suggesting at least partial capsid dissociation before termination. However, some studies suggest that reverse transcription may be completed after entry through the nuclear pore complex and not required for capsid disassembly there [32, 33]. Overall, HIV-1 capsid acts as a shell, at least during the initial stages of reverse transcription.

1.2.5.2 A binding platform for cellular host factors

Between host cell entry and genome delivery into the nucleus, HIV-1 capsid lies at the interface between host cytosol and viral content. The capsid provides a binding platform for cytoplasmic host proteins, which regulate its stability. Multiple studies

have identified such host proteins, aiming to analyse their impact on infection and their potential for drug design.

In the early steps of infection, various host factors in the cytoplasm associate with HIV-1 capsid (Figure 1.6.a-d,h). Among these, the best characterised is CypA, a cytoplasmic protein thoroughly described in section 1.3.6. Another host factor is the tripartite motif protein 5 (TRIM5), which blocks infection soon after viral entry into host cell cytoplasm [34, 35]. TRIM5 binds to HIV-1 capsid through a PRYSPRY motif [36] for the α isoform (TRIM5α) or the CypA domain obtained from fusion of TRIM5 and CypA (TRIM5Cyp) [37–39]. TRIM5α acts as a pattern recognition receptor specific to HIV-1 CA lattice and forms a hexagonal scaffold (Figure 1.6.d), which surrounds the capsid shell [40, 41]. Notably, interaction with TRIM5α activates innate immune signalling pathways against the capsid [42]. Nonetheless, the direct effect of TRIM5α on capsid stability remains debated. Formation of stable complexes of capsids bound to TRIM5α suggests that it may protect and stabilise the capsid lattice [24, 43]. In this case, restriction may be caused by TRIM5α ubiquitination and recruitment of proteasomes to clear viral particles [41, 44]. Conversely, some studies showed that TRIM5α also promotes premature capsid disassembly [40, 45–49], suggesting a direct role in capsid destabilisation.

Another important host factor is the polyanion inositol hexaphosphate (IP_6). IP_6 is recruited from the cytosol of host producer cells during assembly of HIV-1 immature viruses [51]. Before new infection and after maturation, IP_6 promotes assembly into a mature cone through binding to the ring formed by six arginines (residue 18) at the centre of each hexamer (Figure 1.6.g) [52]. This interaction stabilises HIV-1 capsid after delivery in the cytosol of host cells, whereas recruitment of other host factors may displace IP_6 and contribute to capsid uncoating [53].

In the cytoplasm, HIV-1 capsid is trafficked toward the nucleus. Some host proteins regulate delivery of the capsid cargo to the nuclear pore complex. For instance, motor complexes such as dynein and kinesin-1 lie at the interface between the capsid and microtubules. There, interactions with CA occur through adaptor proteins that directly bind to CA, such as fasciculation and elongation protein zeta-1 (FEZ1) and bicaudal D cargo adaptor 2 (BICD2) [54–57]. While dynein and kinesin-1 move along the

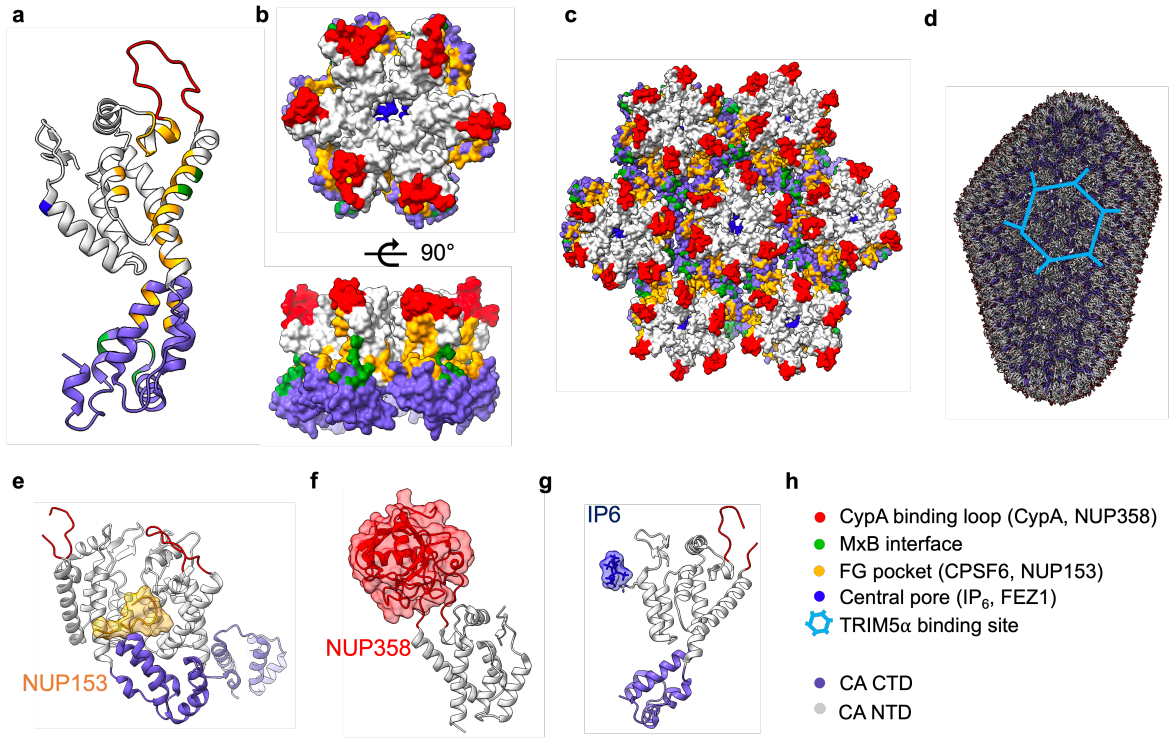


Figure 1.6: HIV-1 capsid is a binding platform for cellular host factors. a) Ribbon representation of HIV-1 CA monomer (PDB code 3J34). Residues that interact with different host factors (interfaces described in h) are in blue, orange, red and green. Other residues of CA_{NTD} and CA_{CTD} are in grey and slate blue, respectively. Figures a) to g) share the same colour code. b) Surface representation of HIV-1 CA hexamer (PDB code 3J34). c) Surface representation of HIV-1 CA heptamer of hexamers (PDB code 3J34). d) Ribbon representation of a model of HIV-1 CA core (PDB code 3J3Y). The binding interface for TRIM5α is shown in sky blue. e) Ribbon representation of HIV-1 CA bound to NUP153 (surface in orange) through the FG-binding pocket of CA (PDB code 4WYM). f) Ribbon representation of HIV-1 CA bound to NUP358 (surface in red) through CypA binding loop of CA (PDB code 4LQW). g) Ribbon representation of HIV-1 CA bound to IP_6 (surface in blue) through arginine 18 (PDB code 6H09). h) Colour code used in this figure. Figure inspired from [50]

microtubule highway and support trafficking of the capsid, both proteins exert an opposing force that contributes to capsid uncoating [58, 59].

At the nuclear membrane, delivery of the viral genome relies on interactions between nuclear pore complex components and the capsid. In particular, nucleoporins NUP153 and NUP358 assist nuclear entry through direct binding to CA. NUP153 binds to the CA_{NTD}-CA_{CTD} interface through a FG motif (Figure 1.6.e)[60, 61] and other nucleoporins (NUP62, NUP98, NUP214) that contain FG motifs bind to CA-NC complexes *in vitro* [61]. NUP358 also contains a cyclophilin domain that interacts with CA through the CypA binding loop (Figure 1.6.f)[62, 63]. Knockdowns of *Nup358* and *Nup153* are both associated with reduced infectivity.

Other host factors such as transportin-3 (TNPO3) and cleavage and polyadenylation specificity factor 6 (CPSF6) promote HIV-1 infection during and after nuclear entry. In the absence of TNPO3, CPSF6 was suggested to accumulate in the cytoplasm and prevent uncoating [64]. In the presence of TNPO3, CPSF6 is in the nucleus and its action in the cytoplasm is therefore inhibited, thus promoting uncoating, nuclear import and infectivity. The interplay between both proteins was further highlighted by reduced HIV-1 infectivity in TNPO3 mutants unable to bind to CPSF6 [65]. In addition, mutation of CA preventing CPSF6 binding almost abolished HIV-1 sensitivity to host factors involved in nuclear import, including TNPO3, NUP358 and NUP153 [66, 67]. Besides, recent studies also found evidence for a role of myxovirus resistance protein B (MxB) in inhibition of HIV-1 infection through binding to HIV-1 CA during and after nuclear entry [68–70].

Most of these host factors associate with interfaces provided by the capsid lattice rather than single CA protomers. However, most structures of host factors bound to CA were obtained for CA monomers or capsomers, rather than higher order assemblies. Therefore, further structural work in the presence of curved assemblies is required to understand how they affect capsid assembly.

1.2.6 CypA interaction with HIV-1 CA

CypA is a peptidyl-prolyl isomerase with a chaperone-like function, which regulates protein folding through catalysis of the *cis-trans* isomerisation of proline imidic peptide bonds. The 165 amino-acid protein is abundant in the cytoplasm and ubiquitous. The crystal structure of CypA reveals an eight-stranded antiparallel β-barrel core, which contains a hydrophobic binding pocket with isomerase activity for exposed proline substrates [71].

CypA was the first cellular host factor found to interact with HIV-1 capsid cores and individual CA protomers through a canonical interface. CypA associates with CA monomer at its CypA-binding loop between helices 4 and 5 of CA_{NTD}, an exposed region of the capsid [10]. The loop contains 13 residues, six of which (H87, A88, G89, P90, I91 and A92) are involved in binding to CypA (Figure 1.7.c-e). P90 acts as a substrate for CypA, which isomerises G89-P90 peptide bond [72]. Both G89V

and P90A point mutations [73] and treatment with CypA inhibitor cyclosporin A [74] abrogate CA binding to CypA. Nonetheless, the relevance of isomerisation on viral infectivity and capsid stability remains unclear, with contradictory findings that associate CypA with destabilisation of CA-NC assemblies *in vitro* [75] or, by contrast, stabilisation of viral cores *in vitro*.

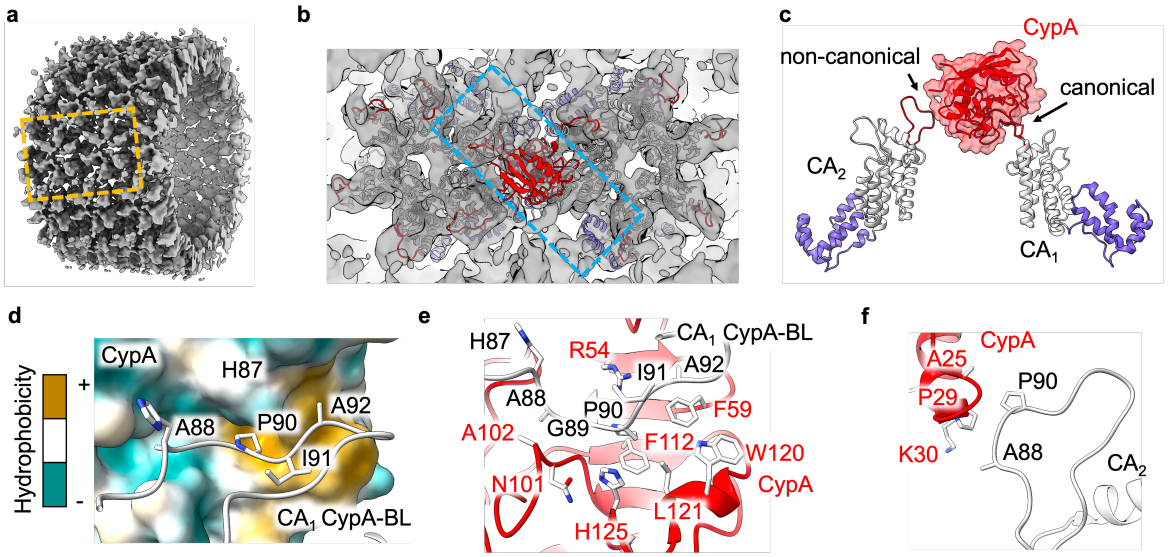


Figure 1.7: Current structural knowledge on CypA binding to HIV-1 CA assemblies. a) Helical reconstruction of *in vitro* tubular assemblies of CA in the presence of CypA (electron microscopy data bank (EMDB) code EMD-3075). Two adjacent hexamers and the CypA binding interface are delimited in orange. b) Close up on the region highlighted in a), with two hexamers (PDB code 3J34) and the CypA binding interface (delimited in blue, PDB code 5FJB) fitted in the cryo-EM density. CA_{NTD}, CA_{CTD} and CypA binding loops are coloured in grey, slate blue and red, respectively. CypA is coloured in red. c) Ribbon and surface representation focusing on the region highlighted in b), with the same colour code. CypA (red surface) contacts two CypA-binding loops of opposite CA monomers at the dimer interface between adjacent hexamers. d) Surface hydrophobicity representation focusing on the canonical binding site to CA (grey ribbon, PDB code 3J34). CypA hydrophobic pocket interacts with residues of HIV-1 CA CypA binding loop (BL). e) Ribbon representation of the canonical binding site of CypA (in red) with CA (in grey, PDB code 3J34). This interface was first characterised by crystallographic studies of CypA bound to CA_{NTD} [10]. f) Ribbon representation of the non-canonical binding site of CypA (in red) with CA (in grey, PDB code 3J34), a new interface suggested by cryo-EM studies of tubular assemblies at moderate resolution (8 Å).

Though CA-CypA binding affinity is comparable for CA monomer, dimer and hexamer [76, 77], it is higher in the presence of higher order assemblies such as native capsids or tubes, suggesting a favoured and more complex mode of binding to assembled capsids. Cryo-EM studies of tubular assemblies of HIV-1 CA in the presence of saturating amounts of CypA showed that all available CypA-binding loops are not simultaneously associated with CypA on the capsid [78]. Instead, CypA appears to bind preferentially

to regions of high curvature in the CA lattice (Figure 1.7.a-b). The same study suggested the presence of a second non-canonical CypA-binding site, through which CypA could bridge two monomers from two adjacent hexamers (Figure 1.7.c,f). However, resolution at this dimer interface was limited and the model was derived from molecular dynamics simulations. Intriguingly, mutations of residues involved in the proposed non-canonical binding site did not alter CypA affinity for CA tubes [77]. Therefore, the precise mode of binding of CypA to HIV-1 CA is only partially characterised.

It is difficult to unambiguously establish how CypA-CA interaction affects HIV-1 infectivity, since studies with different cell lines show varying results. For instance, CypA-CA interaction is associated with increased infectivity in $CD4^+$ T cells, dendritic cells and macrophages, whereas in HeLa cells CypA does not appear to modulate viral infectivity [79–82].

The exact influence of CypA on capsid stability remains disputed for now, with contradicting studies highlighting both stabilising and destabilising effects on CA assemblies. For instance, CypA was shown to increase stiffness of *in vitro* tubular assemblies [78], to stabilise HIV-1 CA cores [83] and to inhibit HIV-1 core disassembly *in vitro* [84]. In addition, loss of CypA binding to HIV-1 capsid was associated with faster viral core uncoating in the cytoplasm [63]. Conversely, CypA had a stabilising or destabilising effect on HIV-1 cores *in vivo*, depending on the studied cell type [81]. Remarkably, changes in CypA concentration induced formation of HIV-1 CA *in vitro* assemblies of different morphology [78]. Given that the concentration of cytoplasmic CypA is different from one cell type to another, this change might explain differences observed for CypA modulation of capsid stability.

Interestingly, CypA was suggested to prevent binding of other cellular factors, such as CPSF6 [85] and TRIM5α [86], which are able to disrupt HIV capsid assemblies. Likewise, capsid uncoating by TNPO3 was inhibited by addition of CypA [84]. Furthermore, a regulatory role of CypA was observed for host factors involved in nuclear entry such as NUP153 and NUP358, as import through these was reduced by CypA inhibition, suggesting a link between nuclear entry and CypA-dependent uncoating [87, 88]. Similarly, CypA is associated with MxB restriction [70]. Altogether, these findings

strongly support a role of CypA in the regulation of capsid uncoating through direct or indirect regulation of binding of other cellular factors to HIV-1 capsid.

1.2.7 Aims

The main aim of this project is to provide a higher resolution structure of curved tubular assemblies of HIV-1 CA in complex with CypA. This could explain how hexamers arrange along CA tubes and how they selectively accommodate CypA molecules. Besides, another aim of this chapter is to describe the steps during helical reconstruction and localised reconstruction for structure determination of heterogenous tubular assemblies at high resolution. The method could be applied to other helical assemblies and associated proteins.

1.3 Materials and methods

In high salt, HIV-1 CA forms curved tubular assemblies, which are suitable for cryo-EM studies. In this section, a background on helical symmetry is given and data processing of helical samples is detailed.

1.3.1 Sample preparation and data collection

Protein purification and sample preparation were performed by Dr Tao Ni (University of Oxford). CypA was expressed in *Escherichia coli* (*E. coli*) and purified as previously described [78]. Full-length WT CA was diluted to 2 mg/mL in high salt buffer (50 mM Tris pH 8.0, 1 M NaCl) and incubated at 37°C for 1 h to yield tubular assemblies. CypA was added to CA tubes at a molar ratio of 6:2.5 (CA:CypA). The sample was applied on a glow-discharged R2/2 Quantifoil grid and plunge-frozen in liquid ethane using a vitrobot (FEI, ThermoFisher Scientific).

Data collection was performed by Dr Tao Ni (University of Oxford) and Dr Kyle Dent (Electron Bio-Imaging Centre, Didcot). Grids were imaged on a 300 kV Krios G3 transmission electron microscope (ThermoFisher Scientific) and images were recorded on a K2 Summit (Gatan) direct electron detector using ArbitrEM (https://github.com/kyledent/ArbitrEM)

in SerialEM. 3535 movies were recorded, with a total accumulative electron dose of 40 electrons per Å^2 fractioned over 40 subframes and a total exposure time of 8 s.

1.3.2 Helical reconstruction of tubular assemblies

1.3.2.1 Preparation and sorting of tubes

Movies were motion-corrected with MotionCor2 [89]. Helical assemblies from dose-weighted micrographs were picked manually in EMAN2 [90] using *sxhelixboxer.py* with a box size of 500 × 500 pixels. Segments and their coordinates were extracted in EMAN2, flipped and aligned in iMAGIC [91] and the diameter of each helical assembly was determined in SPIDER [92]. Helices were first sorted into three groups based on their diameter. In each subset, Fourier transforms of all tubes were inspected and tubes that shared similar diffraction patterns were sorted into subgroups for symmetry determination.

1.3.2.2 Determination of helical symmetry

In this section, some background on helical symmetry is given. Moreover, determination and validation of helical parameters are presented for a subset of HIV-1 CA tubes in complex with CypA.

1.3.2.2.1 Parameters of a helix

A continuous helix is characterised by a pitch (P), *ie* the axial distance between two points separated by one turn around the helical axis, and a radius (r). Protein helical samples can be seen as discontinuous helices, in which subunits are placed in a regular manner along a continuous helix, with positions of two subunits related by a rotation around and a translation along the helical axis (Figure 1.8.a).

Between two neighbouring subunits, the axial distance and rotation around the helical axis are the rise (Δz) and the twist ($\Delta\Phi$), respectively (Figure 1.8.a). These two parameters are sufficient to define the screw symmetry of a discontinuous helix. Axial cut and unrolling of a cylinder surrounding such helix yields a characteristic 2D lattice, which is reminiscent of 2D lattices observed for 2D crystals (Figure 1.8.b). Helices consist of one strand (*1*-start) or multiple strands (n-start). Nonetheless,

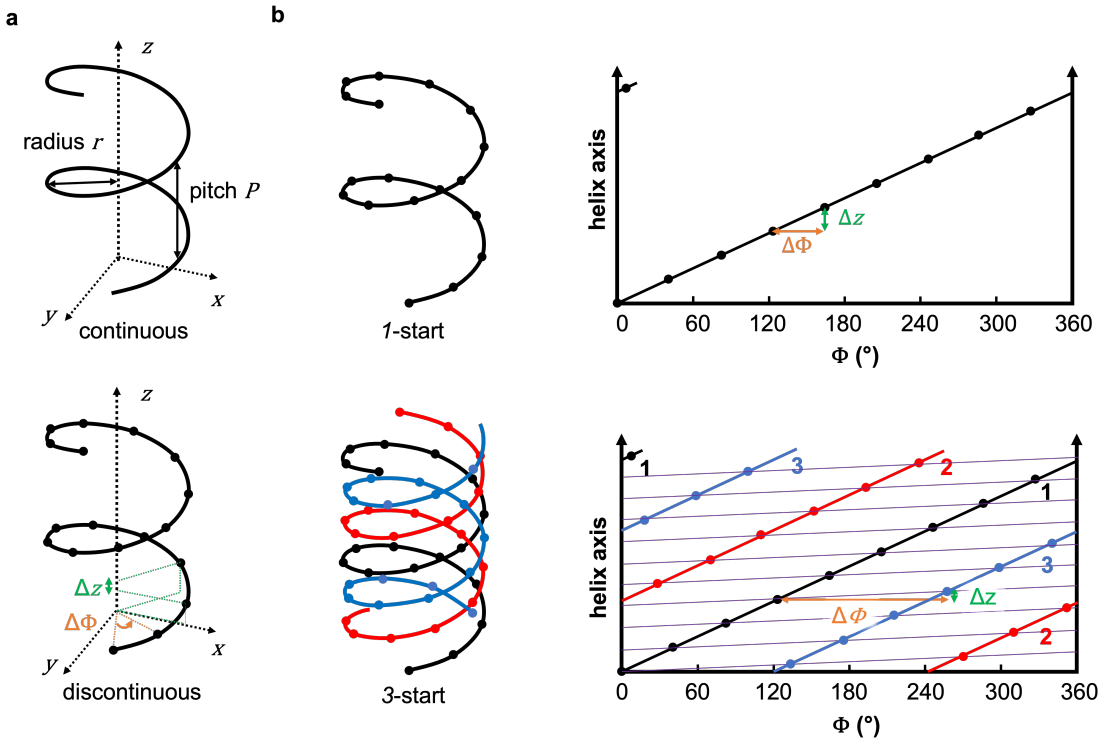


Figure 1.8: Helical parameters and the 2D lattice. a) Schematic representation of continuous (top) and discontinuous (bottom) helices. The pitch (P) and radius (r) are shown for a continuous helix. The rise (Δz) and twist ($\Delta\Phi$) are shown for a discontinuous helix. b) Schematic representation of *1*-start and *3*-start discontinuous helices, and their corresponding 2D lattice, where the azimuthal angle (Φ) varies between subunits at different helical heights. The 2D lattice is generated from the cut and unrolling of a cylinder that intersects all subunits along the helical axis. For a *1*-start helix, a single strand is observed. In the case of a *3*-start helix, three strands are distinguished. Each strand has the same rise and twist as that of the *1*-start helix shown above. However, the three *1*-start helices can be transformed into a single *1*-start helix (shown in purple) that intersects all subunits on each strand. In this case, the rise and twist are different but are sufficient to describe the screw symmetry of the *3*-start helix (figure adapted from [93]).

any n-start helix can be described by a single set of rise and twist that accounts for positions of all subunits in all strands.

1.3.2.2.2 Fourier transform of a helical object

Tubular assemblies are ideal samples for studies by electron microscopy. In fact, the first three-dimensional (3D) reconstruction obtained by electron microscopy was that of a helical phage tail, solved in 1968 by De Rosier and Klug [94]. A projection of a helix contains information for many views of the asymmetric unit in different orientations. In particular, helical symmetry can be determined from inspection of the diffraction pattern of a helix in Fourier space. Though they can both be described by a 2D lattice,

the diffraction patterns of 2D crystals and of a projection of a helical tube are related but different. Indeed, while 2D crystals give discrete diffraction spots, the Fourier transform of the projection of a helical object contains smeary dots. This is caused by rolling up of the helical lattice into a tube in real space. Indeed, in a projection of the rolled lattice, lattice points at the front and back of the cylinder appear closer at the edges than lattice points at the centre of the cylinder (Figure 1.9.a-c). In addition, while front and back of the flattened tube give identical diffraction patterns, rolling it up leads to shifts in diffraction patterns and differential intensities between both halves.

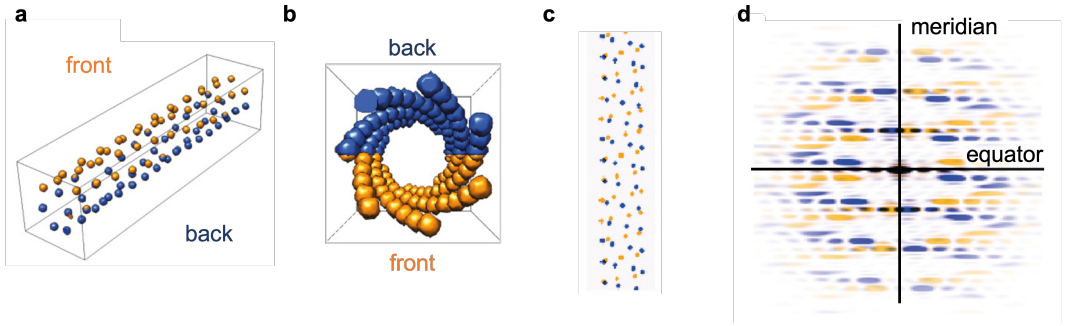


Figure 1.9: Fourier transform of the projection of a helix. a) Subunits of a tubular assembly seen from the side. Subunits are shown as beads. Subunits of the front and back halves are coloured in orange and blue, respectively. b) Subunits of the same tubular assembly seen from the helical axis. c) Projection of the front and back halves of the tube. d) Fourier transform of the projection shown in c). The contribution of each half to the diffraction pattern can be appreciated (figure taken from [95]).

Typically, the Fourier transform of the projection of a helical object consists of a pattern of horizontal smeary reflections, called layer lines (Figure 1.9.d). A reflection of highest intensity is found at the centre of the diffraction pattern. This origin is intersected by a layer line, called the equator, and a vertical axis, named the meridian, across which the transform shows mirror symmetry. Notably, contributions from front and back halves of the tube can be distinguished.

1.3.2.2.3 The Bessel function and the selection rule

The density distribution of a helix can be described as a sum of helical waves that repeat around the circumference of a cylinder surrounding the helix [96]. Therefore, the Fourier transform of a helical projection is the sum of the Fourier transforms of its constitutive waves. Each Fourier transform of a wave produces a layer line,

which is characterised by indices (n, l), where n corresponds to the start number of the helix, and l is the layer line order.

For a given helical wave, *ie* a given layer line (n, l), solutions of the Fourier transform are given by the Bessel function J_n. For instance, the meridian is derived from the Bessel function J_0. Bessel functions first reach a maximum, then oscillate and decline with increasing distance from the origin (Figure 1.10.b). Importantly, the first maximum reached by J_n gets further from the origin with increasing n^{th} orders, which explains the characteristic cross-shaped Fourier transform observed in the simple case of a continuous helix (Figure 1.10.c).

The Fourier transform of a discontinuous helix can be produced from convolution of a continuous helix of pitch P with a set of 'layer planes' that cross each subunit perpendicularly to the helical axis (Figure 1.10.a). On one hand, in the Fourier transform of a continuous helix, layer lines are spaced by $\frac{1}{P}$ along the meridian (Figure 1.10.c). On the other hand, layer planes in real space intersect the helical axis at heights of $m \times \Delta z$, where Δz is the subunit axial translation (helical rise), and m is an integer. The corresponding Fourier transform consists of points separated from the origin by m times $\frac{1}{\Delta z}$. Multiplication of both Fourier transforms yields the characteristic diffraction pattern observed for the projection of a discontinuous helix (Figure 1.10.c). In particular, it results in crosses on the meridian at each multiple of the rise $(m \times \Delta z)$, and consecutive layer lines are separated axially by $l = \frac{1}{c}$, where c is the repeat distance of the helix.

In such case, layer line orders l follow the selection rule: $\frac{l}{c} = \frac{n}{P} + \frac{m}{\Delta z}$, where n is the order of the Bessel function J_n, c is the repeat distance of the helix, P is the pitch of the helix and Δz is the helical rise [99]. Another way to write this formula is $l = tn + um$, where n is the Bessel order, t is the number of turns per repeat $(\frac{c}{P})$ and u is the number of subunits per repeat $(\frac{c}{\Delta z})$.

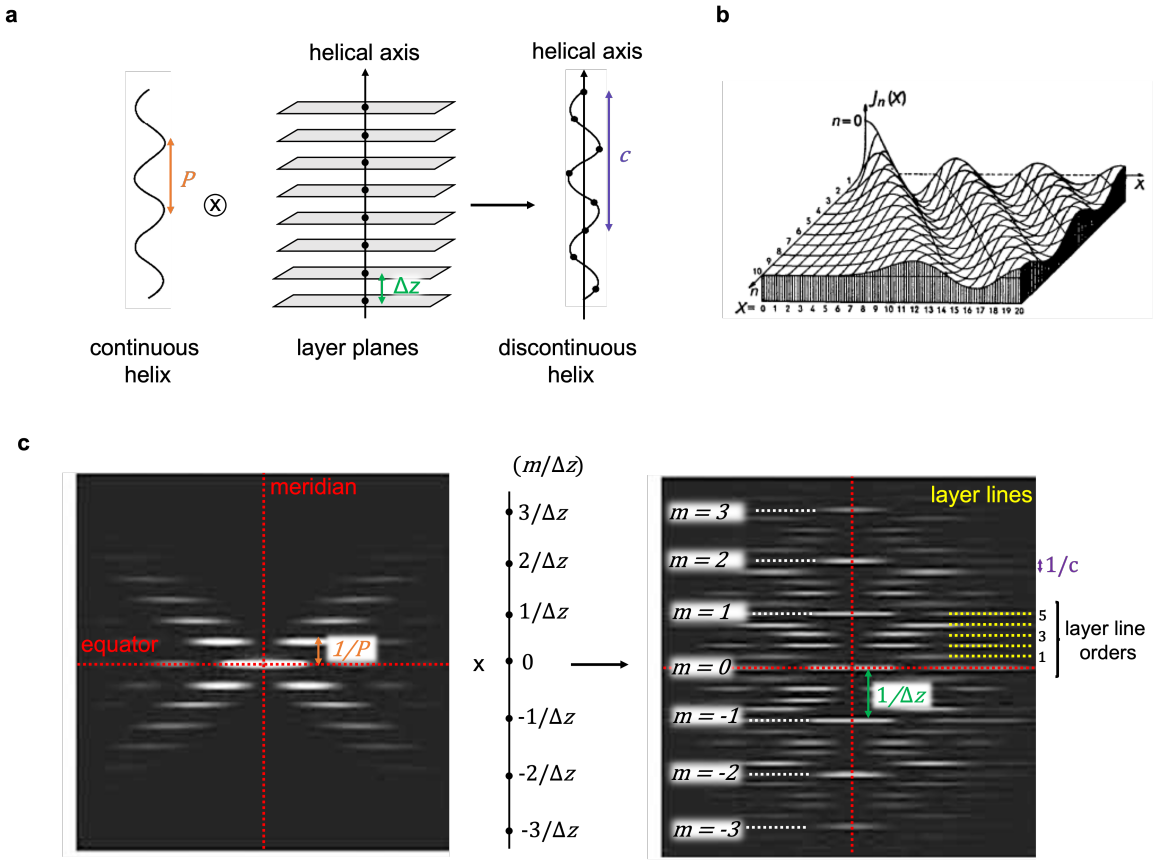


Figure 1.10: Differences between continuous and discontinuous helices in Fourier space. a) Schematic representation of formation of a discontinuous in real space from a continuous helix and helical subunits. A discontinuous helix results from convolution of a continuous helix and a set of layer planes that intersect consecutive asymmetric units of the helix perpendicularly to the helical axis with a height of Δz. The pitch (P) and distance repeat (c) of the continuous and discontinuous helices are shown in orange and purple, respectively. b) 3D plot of the Bessel function $J_n(X)$ for 10 orders n. With increasing n orders, the X distance of first maxima from the origin increase (plot taken from [97]). c) Description in Fourier space resulting from the convolution shown in a). The product of the Fourier transform of the continuous helix with that of each layer planes yield a more complex Fourier transform of a discontinuous helix, in which layer lines can be drawn with an inter-distance equal to the reciprocal of the repeat distance ($\frac{1}{c}$). Layer lines that cross the meridian have a distance to the equator proportional to the reciprocal of the rise ($m \times \frac{1}{\Delta z}$, figure adapted from [98]).

1.3.2.2.4 Indexing of helical symmetry

Since both rise and pitch are sufficient to describe helical symmetry, these need to be estimated before reconstruction and later refined to allow structure determination at high resolution. For many tubular samples, symmetry is unknown and requires attribution of the Bessel orders to their corresponding layer lines in Fourier space. This process, called indexing, consists in a series of steps and relies on rules and equations

that are detailed below, with the example of a tubular assembly of HIV-1 CA in complex with CypA from our dataset (Figure 1.11.a).

First, two vectors $(1,0)$ and $(0,1)$ starting from the origin and directed toward the centres of first order maxima of two different layer lines are drawn arbitrarily (Figure 1.11.b). They are characterised by Bessel orders $n_{(1,0)}$ and $n_{(0,1)}$, respectively. Likewise, their layer line orders are noted $l_{(1,0)}$ and $l_{(0,1)}$, respectively. These two vectors define a 2D lattice which should intersect layer lines at all intensities after mirror symmetry (Figure 1.11.c,d). Since pairs of lattice vectors are drawn arbitrarily to find the satisfactory 2D lattice, this process can be tedious.

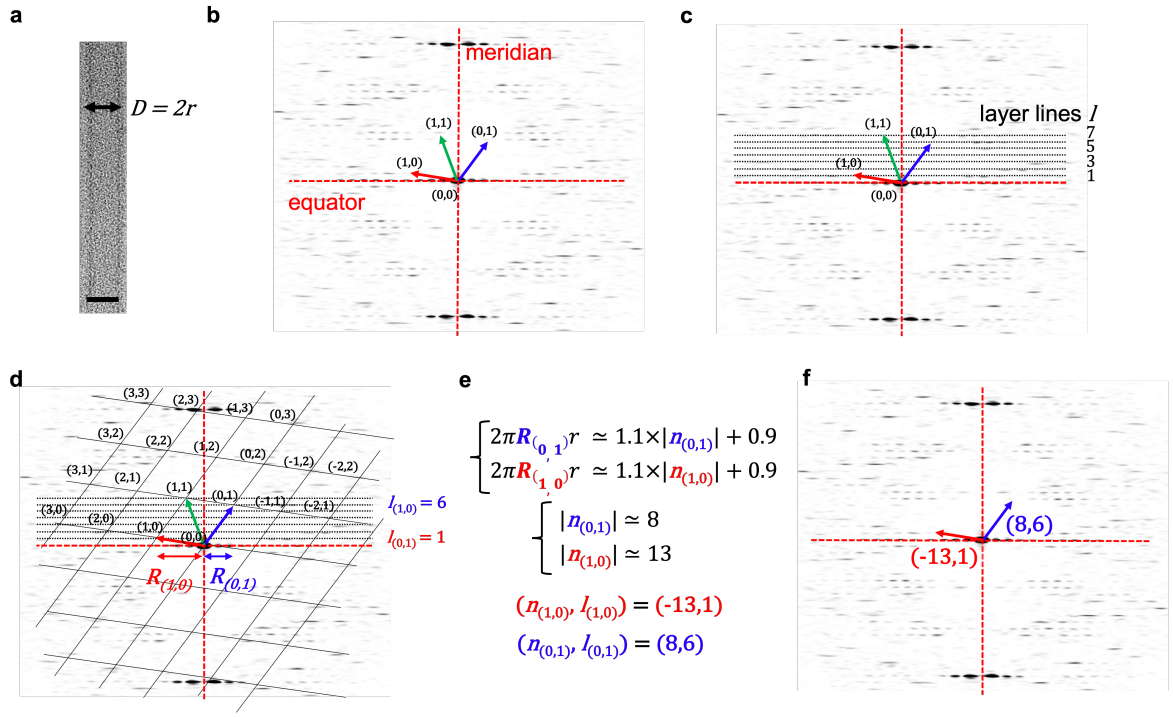


Figure 1.11: Indexing of Bessel orders in Fourier space. a) Projection of a helical tube of HIV-1 CA protein in complex with CypA, observed in IMOD. Scale bar is equal to 30 nm. D is the diameter, which can be estimated from this projection. b) Fourier transform of a), with black and white inverted for clarity. Meridian and equator are dashed in red. First, two lattice vectors are drawn (in red and blue). The third vector (in green) is a combination of the addition of both lattice vectors. c) The same Fourier transform, with layer lines added with their corresponding orders. d) The same Fourier transform, with the lattice drawn from both lattice vectors $(1,0)$ and $(0,1)$. Intersections in the lattice match with positions of intensities from all layer lines after application of mirror symmetry. One can draw a lattice with identical intersections using the green vector instead of the red or blue vector for instance. The distances on the meridian $R_{(1,0)}$ and $R_{(0,1)}$ can be measured. e) The Bessel approximation for the first maximum allows determination of Bessel orders $n_{(1,0)}$ and $n_{(0,1)}$. Both lattice vectors can now be defined in terms of Bessel and layer line orders. f) The same Fourier transform with lattice vectors defined by their Bessel and layer line orders.

Then, the Bessel order of each layer line is estimated. For this, it was shown that intensity on the n^{th} layer line is proportional to ${J_n}^2(2\pi rR)$, where r is the radius of the helix in real space, R is the radius along the n^{th} layer line (ie the distance to the helical axis) in reciprocal space and J_n is a Bessel function of the n^{th} order [100]. In particular, the first maximum of the Bessel function follows the approximation:

$$2\pi rR_{\max} \sim J_n \sim 1.1.|n| + 0.9$$

In this case r and $R_{\max}$ can be determined empirically from a projection and its Fourier transform, respectively. In practice, $R_{(1,0)}$, $R_{(0,1)}$ are measured, then the Bessel orders $n_{(1,0)}$ and $n_{(0,1)}$ are estimated using the equation above. For the tube shown as an example, $n_{(1,0)}$ is estimated at -13 while $n_{(0,1)}$ is 8 (Figure 1.11.e). The corresponding layer line orders $l_{(1,0)}$ and $l_{(0,1)}$ are 1 and 6. Therefore, the two vectors that define the lattice are $(-13, 1)$ and $(8, 6)$ (Figure 1.11.f).

Then, the reciprocal of the repeat distance $(\frac{1}{c})$ is measured as the height that separates two consecutive layer lines in Fourier space (Figure 1.12.a). Subsequently, the number of subunits per repeat (u) and the number of turns (t) per repeat are calculated. Visualisation of the 2D lattice in real space is helpful to determine u and t (Figure 1.12.b). To do so, both lattice vectors undergo rotation by a 90° and mirror symmetry, *ie* vectors $(-13, 1)$ and $(8, 6)$ in Fourier space become $(1, -13)$ and $(6, 8)$ in real space, respectively. The real space lattice derived from these vectors shows variation of helical height for each subunit at different azimuthal angles (Φ).

To estimate the number of subunits per repeat, the closest subunit positioned at the same azimuthal angle as the first subunit (0,0) should satisfy $\Phi = 0$ from a combination of both lattice vectors (Figure 1.12.c). Thus, one can solve the following equation $a \times (6, 8) + b \times (1, -13) = (0, u)$, where $|a|$ and $|b|$ are co-prime integers. In this case, $(a, b) = (1, -6)$ is a solution, which gives a value of $u = 1 \times 8 + (-6) \times (-13) = 86$ subunits per repeat. Then, the rise can be derived from $\Delta z = \frac{c}{u}$.

From the real space lattice, one can also identify the vector that points from the origin to the first asymmetric unit (Figure 1.12.b). This subunit is indexed with the Bessel order $n = 1$. In particular, the helical height contribution from this vector represents the

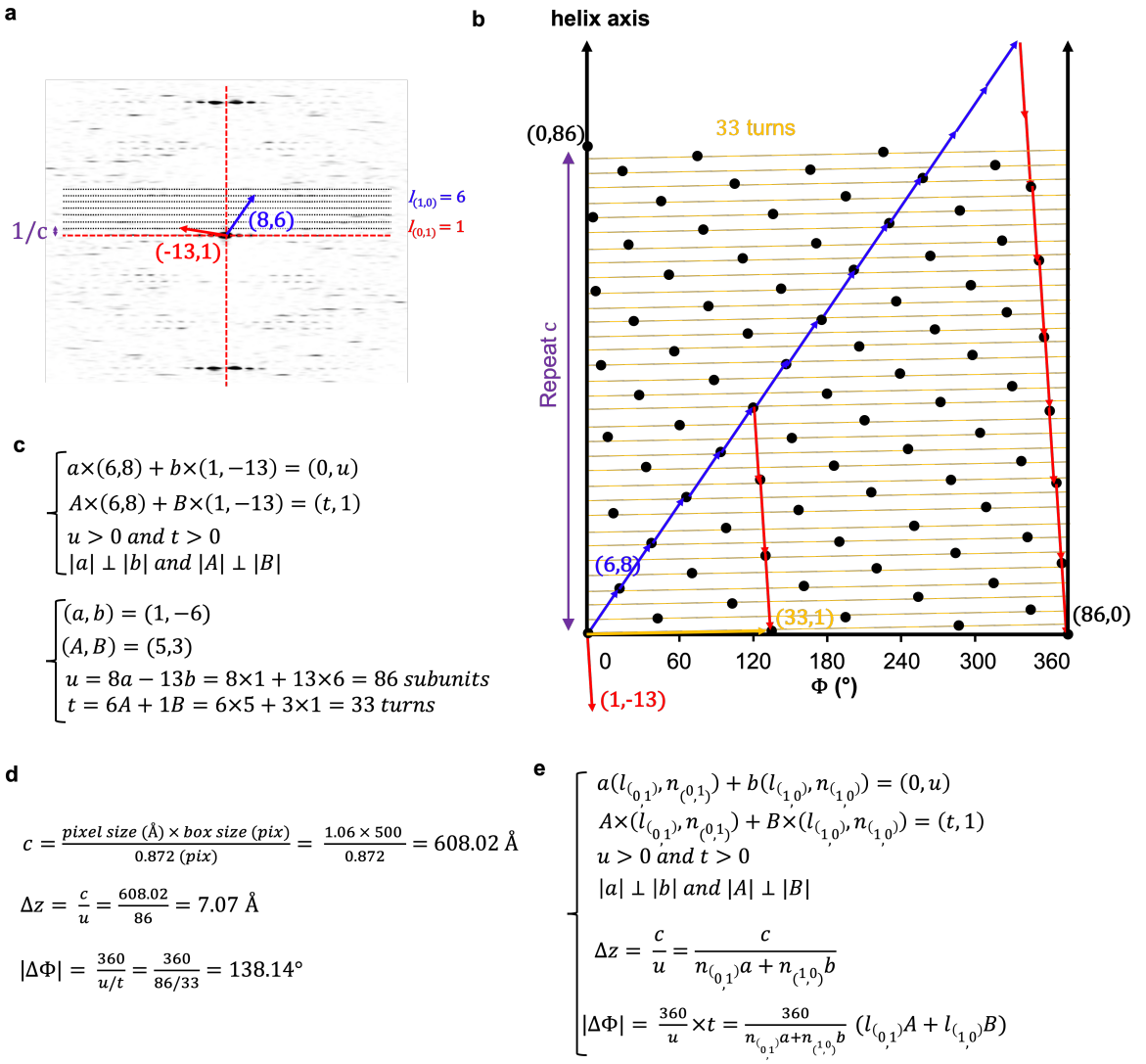


Figure 1.12: Determination of screw symmetry. a) Diffraction pattern obtained for HIV-1 CA example tube in complex with CypA. Lattice vectors and a series of layer lines are drawn. b) Real space representation of the 2D lattice obtained from the same tube. The orange vector links two consecutive asymmetric units, therefore describing rotation and translation by rise and twist, respectively. From this vector, a *1*-start helix (in orange) that pass through 86 subunits (u) in 33 turns (t) per repeat (c) is drawn. c) Calculations of the repeat distance, rise and twist. Attention should be paid as values measured on the Fourier transform are in pixels. d) Equations that characterise our case helix, allowing determination of u and t without the need of drawing the real space lattice. e) Equations to solve for determination of screw symmetry of any n-start helix.

axial distance between two asymmetric units, in other words, the rise of the helix. To get to the first Bessel order, one can solve the equation $A \times (6,8) + B \times (1,-13) = (1,t)$, where $|A|$ and $|B|$ are co-prime integers. Here, $(A,B) = (5,3)$ is a solution, which gives a value of $t = 6 \times 5 + 1 \times 3 = 33$. In Fourier space, the subunit at $(1,33)$ is on the 33^{rd} layer line at a height of $\dfrac{1}{\Delta z}$ $(m = 1)$ on the meridian. Therefore, the layer line order

of the Bessel order $n = 1$ gives the number of turns in the helix repeat, in this case 33 turns. Eventually, the helical twist can be calculated as $\Delta\Phi = 360 \times \frac{t}{u}$ (Figure 1.12.d).

Drawing the entire real space lattice is a lengthy process that can be tedious when helical symmetry is not simple. Based on the example presented here, I found that after calculation of the Bessel and layer lines orders of the two lattice vectors, the helical twist and rise of any n-start helix can systematically be derived from the measure of the repeat distance c in Fourier space and determination of the number of turns t and subunits per repeat u using the general equations shown in Figure 1.12.e.

1.3.2.2.5 Validation of screw symmetry

The rise and twist determined after indexing need to be validated. For this, both parameters and relevant tubes can be used as inputs in iterative helical real space reconstruction (IHRSR) [101].

To start with, picked tube projections are divided into many overlapping boxes along their tube length (Figure 1.13.a,b). For samples with heterogenous diameter distributions, diameters are estimated and image stacks of segments of similar diameter and Fourier transform are created.

During each cycle of IHRSR (Figure 1.13.c), segments of a stack are matched with projections generated from a 3D tube reference (often a beads model which satisfies the input screw symmetry, or a solid cylinder), then back-projected to generate a 3D asymmetric volume. Subsequently, helical symmetry is imposed with newly estimated rise and twist, leading to a 3D reconstruction with refined screw symmetry. The 3D map is used as a reference model in the next cycle, and iterations are performed until helical rise and twist converge to stable values. This procedure is relatively fast and allows validation of screw symmetry derived from Fourier Bessel indexing, as observed for our (-13,8) helix, which displays repeats of hexameric subunits (Figure 1.13.d).

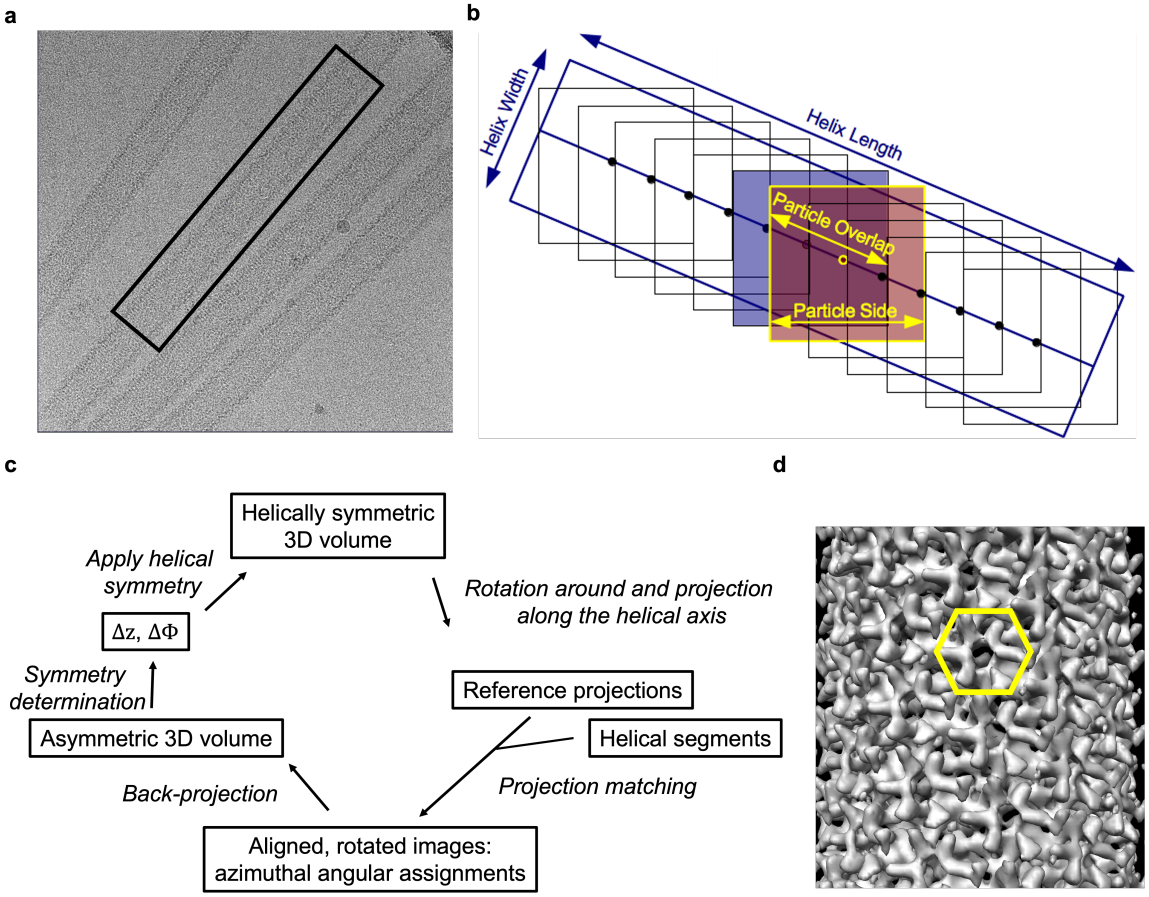


Figure 1.13: Validation of screw symmetry by IHRSR. a) Micrograph of a tube of HIV-1 CA in complex with CypA. The tube is boxed in black. Magnification is 31,000. b) Segmentation of a helical tube into overlapping boxes for projection matching. This can be done in EMAN2 with *e2helixboxer.py* (picture taken from EMAN2 website). The user defines the overlapping distance and the box size. c) Description of the IHRSR cycle. Scheme adapted from [102]. d) 3D reconstruction of HIV-1 CA (-13,8) tube complexed with CypA obtained after ten iterations of IHRSR. Hexameric shapes can be seen (delimited in yellow) and are representative of capsomers of the CA tube. This indicates that screw symmetry has been correctly assigned.

1.3.2.3 Classifications and refinement of tubular assemblies

After determination of helical symmetry, the rise and twist need to be refined to reach higher resolution. For this, IHRSR is implemented in RELION 2.1 and later versions. Exposures and groups of tubes sorted by diameter and symmetry were imported in RELION 2.1 [93]. Contrast transfer function (CTF) was estimated with GCTF [103]. Tubes were extracted with a particle box size of 720 × 720 pixels, with an inter-box distance, *ie* the distance between centres of consecutive extracted segments, of six asymmetrical units ($6 \times \Delta z$). After downsampling by two (box size of 360 × 360 pixels),

the rise and twist obtained by IHRSR were used as input parameters for reference-free 2D classification into 50 classes. 25 iterations were performed with a regularisation parameter (T) of 2, a mask diameter of 686 Å (90 % of the box size) and a tube diameter based on sorting of helical diameter in SPIDER.

For each symmetry, the best 2D classes were selected and used for 3D classification using a beads model initial 3D reference. Beads models of different symmetry were generated with *relion_helix_toolbox* using outer diameters of tubes, pixel size (2.12 Å), helical rises and twists and box dimension (360 × 360 pixels) in C_1 symmetry.

For each symmetry, 3D classification into three classes was performed over 25 iterations using an initial low-pass filter of 60 Å, a regularisation parameter (T) of 4, a mask diameter of 686 Å (90 % of the box size) and an angular sampling of 3.7 Å. Inner and outer tube diameters, number of asymmetric units, and estimated twist and rise were provided. Local search was performed for 1° around the estimated value of the twist, with a step value of 0.1°. For the rise, local search was done for 0.5 Å around the input value, and a step value of 0.1 Å. Then, the best 3D class was selected and used in 3D refinement, with an initial low-pass filter of 60 Å, a mask diameter of 686 Å, an angular sampling of 3.7 Å. Again, inner and outer tube diameters, number of asymmetric units, and estimated twist and rise were provided. Local search was performed for 0.5° around the estimated value of the twist, with a step value of 0.1°. For the rise, local search was done for 0.5 Å around the input value, and a step value of 0.1 Å. Final helical reconstructions were obtained without binning after re-extraction with a particle box size of 720 × 720 pixels and 3D refinement with finer angular sampling.

1.3.3 Localised reconstructions of subregions of interest

To improve resolution of hexamers and better resolve and classify the binding interfaces of CypA, symmetry expansion and localised reconstruction was performed in Scipion [104]. This method allows extraction of regions of interest and their subsequent processing as individual subparticles, rather than processing of helical segments, in which interesting features (flexibility, interfaces) of these subregions may not be observed because helical symmetry is applied.

First, helical particles refined in RELION were symmetry-expanded by 6, the number of asymmetric units, using *relion_particle_symmetry_expand* with the refined helical parameters. This means that each original particle is duplicated into six symmetry-equivalent copies and $\frac{n}{6} \times 360°$ is added to the value of the first Euler angle of each copy (n ranges from 1 to 6).

In Chimera [105], a single region of interest (one hexamer, four CA_{NTD} at the dimer interface between hexamers, or three CA_{NTD} at the trimer interface between hexamers) was subtracted from a reconstruction of symmetry expanded particles generated with *relion_reconstruct*. More particularly, regions of interest were subtracted at a height close to that of the helix centre of mass and at a position that matched that of a subunit of a beads model of same symmetry, generated with *relion_helix_toolbox*. A Chimera map marker (cmm) file was created to save coordinates of the centre of mass of the tube ($x = y = z = 720 \times \frac{1.06(pixelsize)}{2}$) and the centre of the cropped density. These coordinates defined the vector that points toward the cropped region of interest.

For localised reconstruction, extracted symmetry-expanded particles, the cmm file, the cropped reconstruction and the corresponding mask were imported in Scipion [106]. For each tube, subtraction was performed using *relion_subtract_projection*, leaving only the region of interest in each segment. Subparticle extraction using *localized_extraction* with a box size of 144 pixels was followed by reconstruction using *relion_reconstruct*.

Hexamer subparticles were further 3D refined in RELION 2.1 with local angular search, in-plane translational searches and enforced C_2 symmetry. For CA-CypA interfaces, subparticles were subject to 3D classification into 10 classes in RELION 2.1. 25 iterations were performed without alignment, with a regularisation parameter (T) of 10 and a resolution limit (E-step) of 8 Å. Out of 10 classes, eight were selected and their corresponding raw particles without particle subtraction were refined, highlighting two different binding modes at the dimer interface between adjacent hexamers. Same binding modes were aligned and summed to yield their corresponding final reconstructions. Post-processing and local resolution estimates were generated in RELION 2.1.

1.3.4 Model building, refinement and data availability

Model building was performed by Dr Tao Ni (University of Oxford). The structure of CA hexamer was solved using rigid-body-fitted structures of CA monomers (PDB codes 4XFX and 3H4E). Structure modifications were performed in Coot [107]. CypA structure was built using the rigid-body-fitted crystal structure of CypA-CA_{NTD} (PDB code 1AK4). Real-space refinement was done in PHENIX [108].

3D reconstructions of (-13,8) tubes and the localised reconstructions of hexamer and CypA-occupied dimer interface are deposited at the electron microscopy data bank (EMDB codes EMD-10229, EMD-10226, EMD-10739, respectively), with refined atomic models deposited at the protein data bank (PDB code 6SKN, 6SKK, 6Y9W).

1.3.5 Analysis of conservation of CypA residues

Cyclophilins with binding activity to HIV-1 CA were listed based on previously published data [73, 79]. Their corresponding sequences were aligned using Clustal Omega (EMBL-EBI). The sequence and structure of CypA from PDB code 1AK4 was used as an input in ConSurf [109] with automatic parameters.

1.4 Results

1.4.1 Helical reconstruction of HIV-1 CA tubes in complex with CypA

In high salt, HIV-1 CA assembles into tubes stabilised by addition of CypA. Imaging by cryo-EM shows that these tubes are heterogenous. First, they display a heterogenous diameter distribution (Figure 1.14.a,b). Indeed, three groups were obtained, with diameters ranging from 265 Å to 284 Å, 285 Å to 310 Å and 311 Å to 345 Å. Since the second group contained most tubes (49.6 %), these were selected for further sorting.

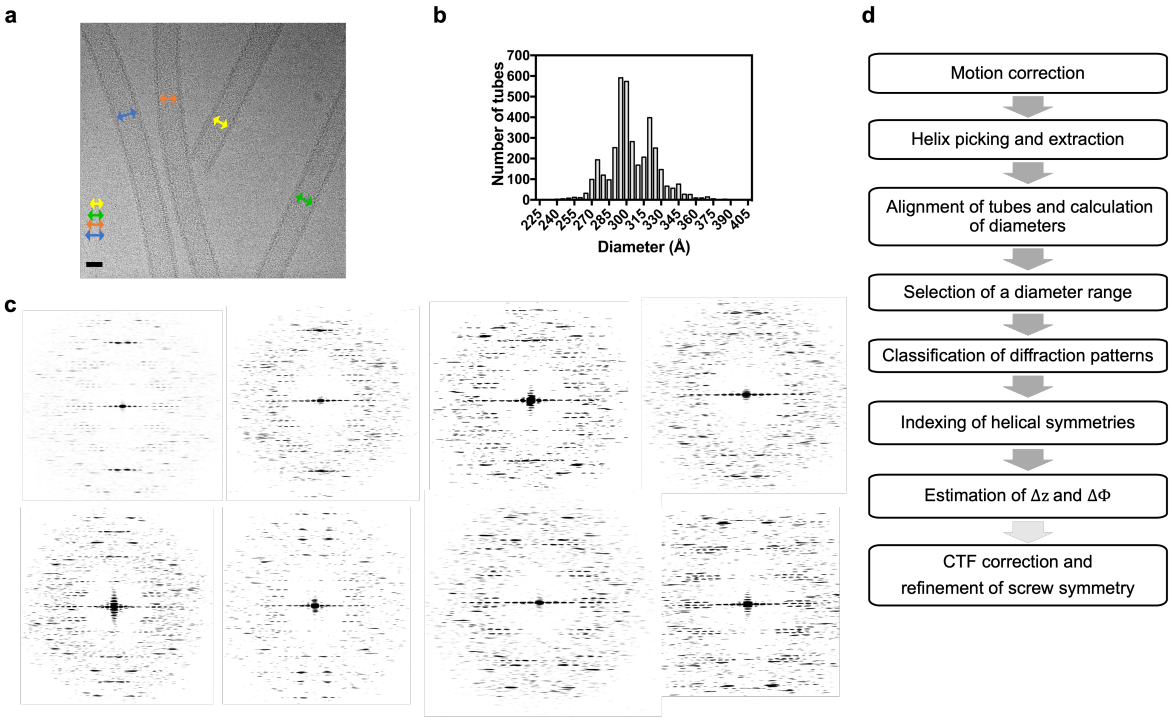


Figure 1.14: Sorting of CA tubes for 3D reconstruction. a) Micrograph showing CA tubes. Arrows are drawn to highlight differences in diameters. Scale bar is equal to 30 nm. b) Diameter distribution for 3,535 picked tubes. c) Examples of Fourier transforms of CA tubes. Diverse layer line profiles can be appreciated. d) Pipeline for preparation of tubes before helical reconstruction.

Notably, tubes also display distinct diffraction patterns (Figure 1.14.c). This means that CA arranges into helical assemblies of different symmetry. Given that this arrangement differs from one symmetry to another, tubes were sorted based on their Fourier transform. Tubes that shared identical layer line profiles were grouped together, whereas bent tubes that showed artefacts in their Fourier transform were discarded. In total, eight major diffraction patterns, therefore eight symmetries were found.

Before generation of helical 3D reconstructions, screw symmetries must be estimated (Figure 1.14.d). In fact, most efforts were put in the very first step of indexing for each group of tubes (method detailed in 1.4.2.2.4). For a given diffraction pattern, a 2D lattice was drawn, first Bessel orders $n_{(0,1)}$ and $n_{(1,0)}$ were estimated, the repeat distance (c) was measured and screw symmetry parameters (rise Δz and twist $\Delta\Phi$) were derived from calculations of numbers of subunits (u) and turns (t) per repeat (Figure 1.15).

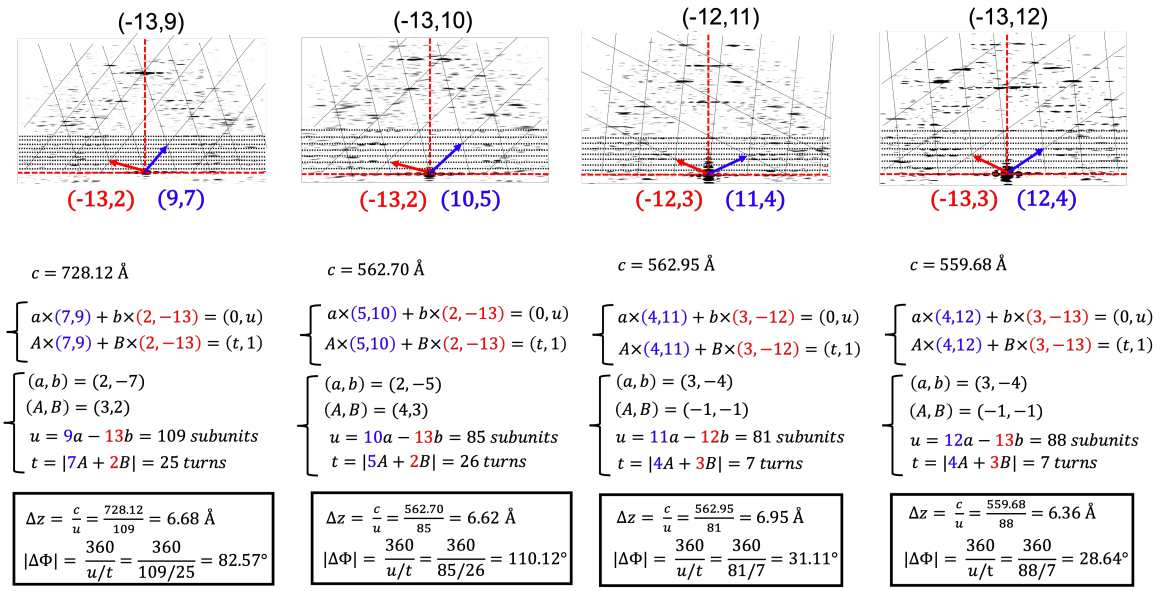


Figure 1.15: Examples of determination of screw symmetry. In a given Fourier transform, a 2D lattice is drawn from a set of two lattice vectors. These vectors are defined by their corresponding Bessel and layer line orders, which are used to determine u and t from which $\Delta\Phi$ and Δz are derived. These are sufficient to describe helical symmetry. The general equations and conditions given in 1.12.e are used here.

In total, seven major symmetries were assigned (Figure 1.16.a), among which the most abundant was $(-13, 8)$ (approximately 20 % of tubes). Other notable symmetries included $(-13, 9)$, $(-8, 13)$, $(-13, 10)$, $(-13, 12)$, $(-12, 11)$ and $(-13, 11)$ (approximately 10%, 6.5%, 6%, 3%, 2% and 1% of tubes, respectively). In each case, a preliminary 3D reconstruction was generated by IHRSR from selected tubes to validate screw symmetry (Figure 1.16.b). Hexameric density repeats along reconstructed tubes suggested that helical parameters were estimated correctly.

For each symmetry, helical parameters obtained by IHRSR were used to generate 2D classes after CTF correction (Figure 1.16.c). Selected 2D classes displayed repeats at the edges of helical segments and textures at the core. There, the contribution of CA_{NTD} as regularly interspaced small white triangles in the core of segments could already be observed. The corresponding particles were classified and refined to higher resolution (Figure 1.16.d).

In all refined tubes, densities for hexamers are clear and differences in their arrangement based on helical symmetry are observed (Figure 1.16.e). At the dimer interface between two adjacent hexamers lies the CypA binding site. For most symmetries,

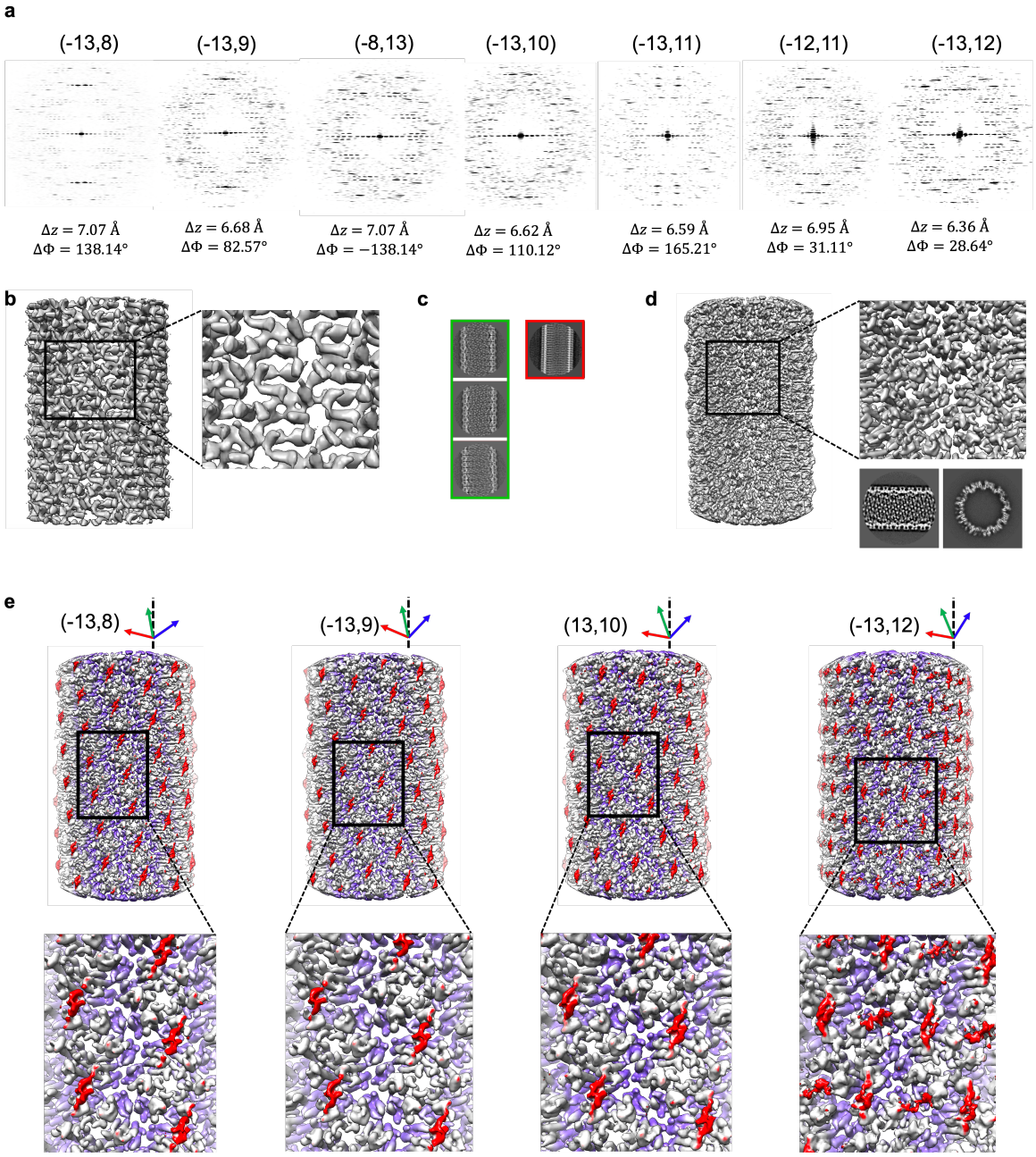


Figure 1.16: 3D reconstruction from helical parameters. a) Seven major Fourier transforms and their screw symmetry. The estimated rise and twist are shown below their corresponding Fourier transform. b) A typical 3D reconstruction from 10 iterations by IHRSR. An in-house script is used to quickly validate screw symmetry. Hexameric repeats can be observed. c) Examples of 2D classes of helical segments generated in RELION. Textured tubes (in green) are selected, whereas other tubes are discarded (in red). d) 3D refinement of screw symmetry after selection of 2D classes. The same region as in b) is magnified, showing well-resolved details there. On projections of such tube parallel and perpendicular to the helical axis, hexamers can be clearly distinguished (bottom right). e) Examples of 3D reconstructions of CA tubular assemblies after refinement in RELION (top). For each reconstruction, a region surrounding four hexamers is magnified (bottom). Electron density matching to CA_{CTD}, CA_{NTD} and CypA bridges are coloured in slate blue, light grey and red, respectively.

two dimer interfaces appear occupied per hexamer. In the case of $(-13, 12)$ tubes, four dimer interfaces per hexamer seem associated with CypA. This is likely caused by a mix of left- and right-handed symmetries, in which CypA binds similarly but along different directions. Notably, $(-13, 8)$ tubes were refined to 4.5 Å, the highest resolution among all symmetries. Therefore this symmetry was selected for localised reconstruction and classification of subregions of interest.

1.4.2 Localised reconstruction of CA hexamer

All previously reported structures of capsomers on curved assemblies have only intermediate resolutions that prevent precise building and description of their local geometry. In these cases, planar capsomers, for which C_6 or C_5 symmetry was imposed, were fitted in their corresponding cryo-EM densities. Here, symmetry expansion and localised reconstruction of hexamers were performed to improve resolution of a single capsomer on a tubular assembly (Figure 1.17.a,b). In particular, hexamers were extracted and treated as single particles, regardless of helical symmetry constraints.

After 3D classification and refinement, a hexamer reconstruction at an overall resolution of 3.6 Å was obtained from $(-13, 8)$ tubes originally solved at 4.5 Å (Figure 1.17.d). Local resolution was higher at the core of the hexamer, which consists of most of CA_{NTDs}, with resolution between 3.4 Å and 3.7 Å (Figure 1.17.c). At the edges of the hexamer, which comprises CA_{CTDs} and the CypA binding loops of CA_{NTDs}, resolution ranged from 3.7 Å to 4.1 Å. From this map, the entire structure of a hexamer on a tube of (-13,8) symmetry was built by Dr Tao Ni (University of Oxford).

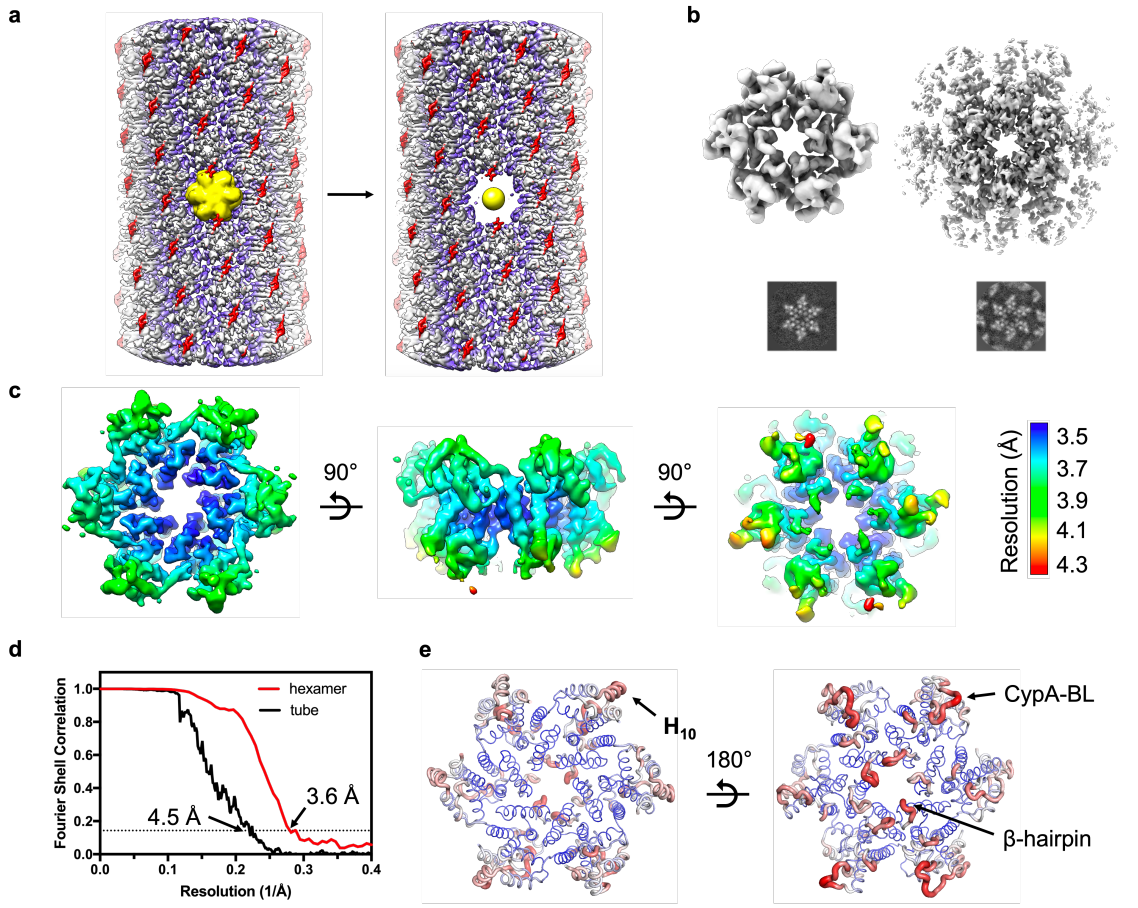


Figure 1.17: Localised reconstruction of a hexamer from (-13,8) tubes. a) Cropping of a hexamer in Chimera. The location of a bead at the centre of mass of the cropped hexamer is recorded for localised reconstruction. b) Localised reconstructions (top) of a hexamer subparticle after density subtraction (left), and its corresponding raw reconstruction on a tube (right), delimited by a box of 144×144 pixels. Corresponding projections are shown (bottom). c) Local resolution estimates of the reconstructed hexamer. d) FSC curves for reconstructed (-13,8) tubes (in black) and corresponding hexamers (in red) after localised reconstruction. Resolution is determined at the FSC value of 0.143. e) Sausage representation of B-factor distribution for a hexamer showing its most dynamic regions in red, while rigid domains are depicted in blue.

B-factor distribution across the structure reveals that the most dynamic regions are the CypA binding loop, the β-hairpin and the helix H_{10} of CA_{CTD} involved in the formation of the interface between three adjacent hexamers (Figure 1.17.e). As opposed to the rest of the capsomer, these regions can accommodate different host factors.

Most importantly, the overall structure presented here is curved and asymmetric. Indeed, in contrast to the previously reported structures of planar hexamers, a curvature of approximately 20° is observed (Figure 1.18.a,b). Since core helices of CA_{CTDs} and CA_{NTDs} are rather static, the plasticity of the hexamer seems to arise from flexibility

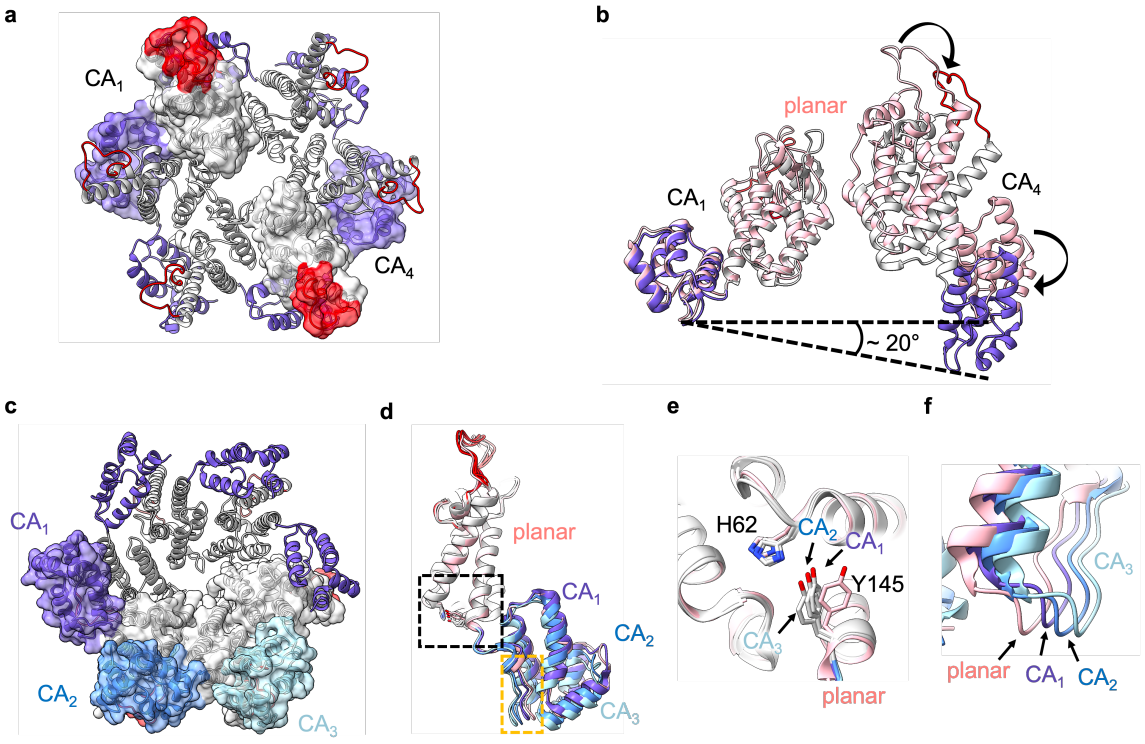


Figure 1.18: A curved hexamer. a) Structure of the hexamer of (-13,8) tubes, with surfaces displayed for two opposed CA protomers on the hexamer. b) Alignment of both protomers highlighted as surfaces in a) with corresponding protomers on the planar crystal structure of a hexamer (in pink, PDB code 4XFX). c) Structure of the hexamer of (-13,8) tubes, with surfaces displayed for three adjacent CA protomers on the hexamer. d) Structure alignment of each protomer highlighted in c) together with a protomer of the planar crystal structure of a hexamer (PDB code 4XFX). e) Close up on the region boxed in black in d). A newly described hydrogen bond is formed between tyrosine 145 and histidine 62. f) Close up on the region boxed in orange in d), showing flexibility at the linker and variations between protomers of a single hexamer.

induced by the linker between CA_{NTD} and CA_{CTD} (Figure 1.18.c-f). In fact, most residues of the flexible linker are essential for proper assembly of capsid cores [110]. In particular, on the most curved CA protomer, a newly described intramolecular hydrogen bond between tyrosine 145 of the flexible linker and histidine 62 of CA_{NTD} is observed (Figure 1.18.d,e). This bond is important for capsid assembly, since mutations of both residues that prevent such hydrogen bonding were reported to compromise CA assemblies *in vitro* and viral infection [23].

1.4.3 Localised reconstruction and classification of CypA bound to CA

In the 3D reconstruction of $(-13, 8)$ tubes, a density bridging two adjacent hexamers is apparent (Figure 1.19.a). Although this density can be attributed to CypA, precise fitting there was ambiguous. This suggested that CypA was wrongly averaged because of imposed helical symmetry and required classification of its binding modes at the dimer interface. Using a similar strategy as for hexamers, the subregion was reconstructed after subtraction of four CA_{NTD} molecules of the dimer interface and subtraction of the density bridging them (Figure 1.19.b).

After extraction, subparticles were classified without alignment (Figure 1.19.c). In all classes, the dimer interface was occupied by a globular density. Two clusters of classes were found: the first class displayed density located between one CA molecule of the bottom hexamer to two CA molecules of the top hexamer, while in the second class, this pose was inverted. This explains the averaged density observed after helical reconstruction. Two 3D reconstructions were produced after 3D refinement of each subset of particles. Given that interfaces were conserved between both classes, particles were aligned and merged together. A last round of refinement yielded a reconstruction with an overall resolution of 4.1 Å (Figure 1.19.d).

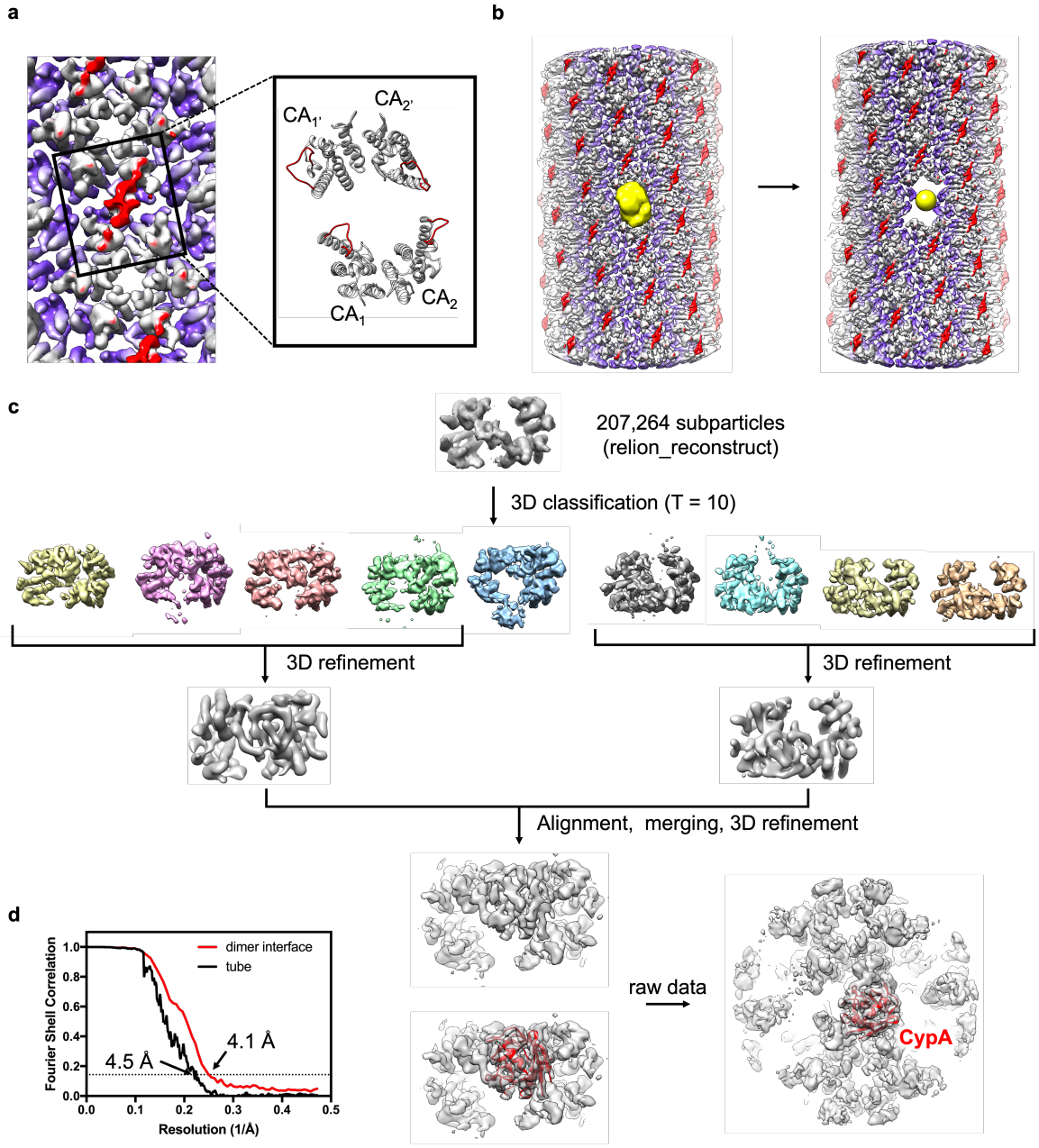


Figure 1.19: Density subtraction at the dimer interface. a) Surface representation centred on the dimer interface on $(-13, 8)$ tubes (left). Four CA_{NTD} molecules were fitted there (right). b) Cropping of the dimer interface in Chimera. The location of a bead at the centre of mass of the cropped region is recorded for localised reconstruction. c) Classification and refinement at the dimer interface. d) FSC curves for reconstructed $(-13, 8)$ tubes (in black) and the dimer interface (in red) after localised reconstruction. Resolution is determined at the FSC value of 0.143.

In the 3D reconstruction, resolution was highest for the core helices of CA, ranging from 3.7 Å to 4.1 Å, while resolution for the rest of the reconstruction was between 4.0 Å and 4.8 Å (Figure 1.20.a). The model derived from localised reconstruction of the curved hexamer was used and the bridging density was assigned to CypA. There, the

crystal structure of CypA could be fitted with confidence (Figure 1.20.b).

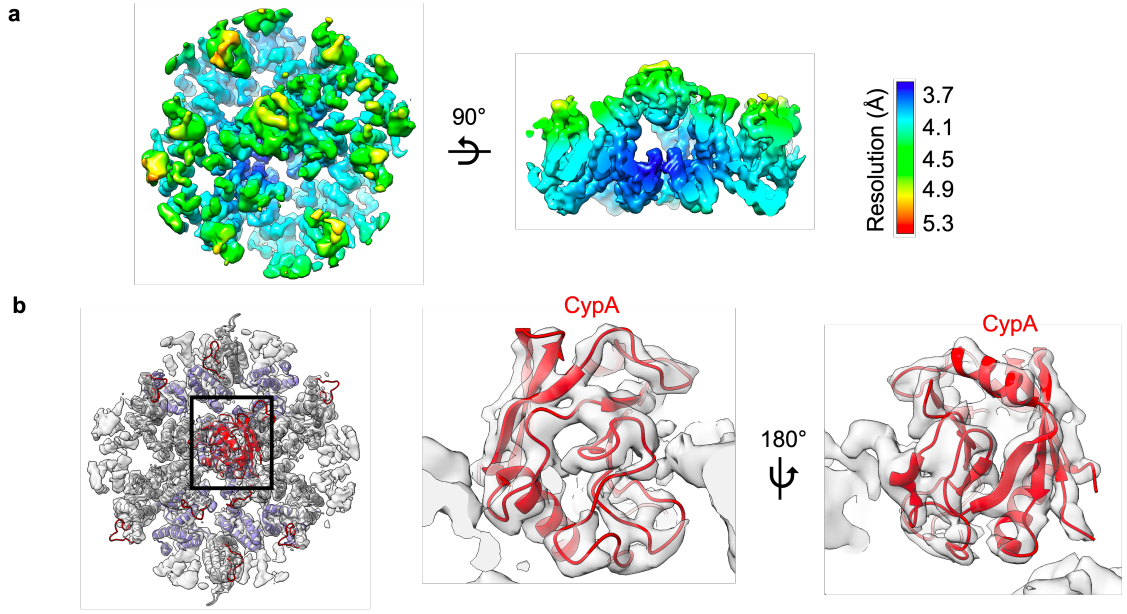


Figure 1.20: Resolution and binding mode of CypA at the dimer interface. a) Local resolution estimates at the dimer interface. b) Density representation of the reconstructed subregion on the tube (left). The boxed region is magnified (right) and shows fitting of CypA in the cryo-EM density.

Although at this resolution, side chains cannot be placed unambiguouly, the structure shows the presence of two CypA binding interfaces (Figure 1.21.a-d). On the one hand, a first binding occurs at the canonical interface, as previously observed by crystallography and cryo-EM studies (Figure 1.21.d). On the other hand, CypA binds to a newly observed non-canonical interface (Figure 1.21.b,c). In particular, CypA residues which appear in close contact with CA include a stretch of three residues (glutamine 43, lysine 44 and glycine 45) that are conserved between cyclophilins with reported binding activity to CA (Figure 1.21.e). Importantly, these residues were not conserved across all cyclophilins, as opposed to the well-conserved residues of the canonical interface (Figure 1.21.f). Therefore these three residues may be important for binding to CA assemblies.

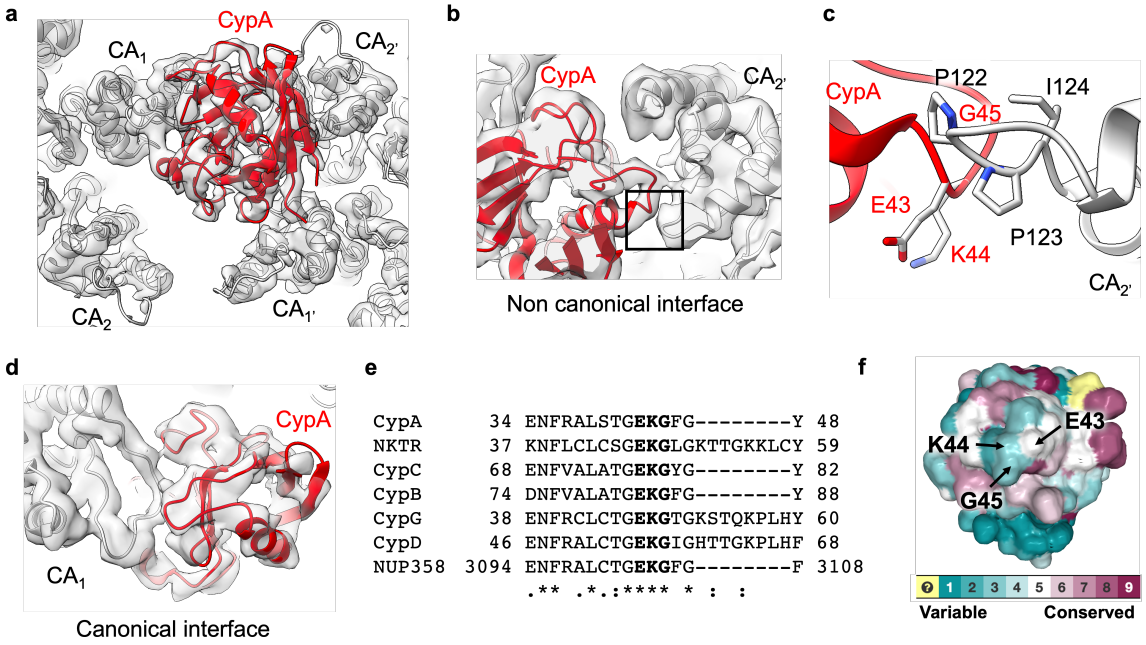


Figure 1.21: Binding of CypA at the dimer interface. a) Top view of the dimer interface, with models fitted into the cryo-EM density. CypA is coloured in red. b) Close up on the novel non-canonical interface, with models fitted in the cryo-EM density. c) Close up on the region delimited in b) (in black), showing contacting residues between CA and CypA. d) Close up on the canonical interface. e) Multi-sequence alignment of CypA with other cyclophilins with binding activity to HIV-1 CA. The three contacting residues of CypA are conserved. f) Conservation of residues across all cyclophilins.

Intringuingly, in about 30% of classes generated during 3D classification, the trimer interface, *ie* the interface between three adjacent hexamers appears to be occupied (Figure 1.22.a). This suggests that two CypA molecules could bind simultaneously at the dimer and trimer interfaces. A similar occupancy is obtained after localised reconstruction and 3D classification of the trimer interface (Figure 1.22.b-d). There, the interface is mostly empty, but also occupied in a minority of cases (Figure 1.22.d). More work is required to determine the structural details of this second binding pose. Nonetheless, these variations in occupancy suggest that the plasticity of hexamers and dynamics at their interface modulate binding of CypA.

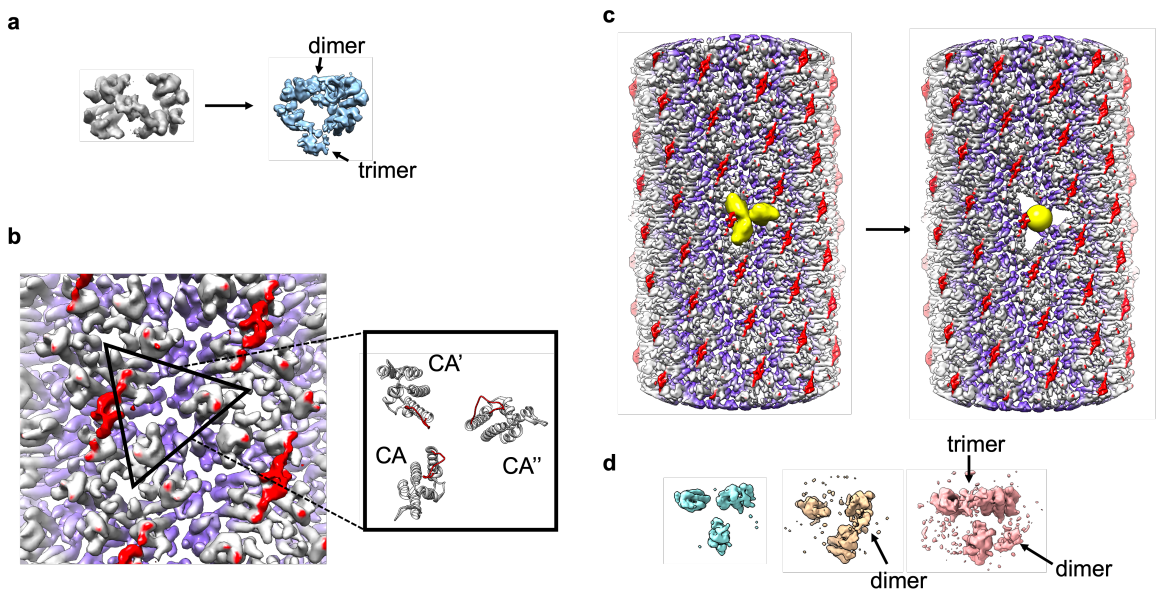


Figure 1.22: Localised reconstruction of the trimer interface. a) A 3D class that shows two binding modes of CypA during localised reconstruction of the dimer interface. The top pose is at the dimer interface, while at the bottom the density is loated on top of the adjacent trimer interface. b) Surface representation centred on the trimer interface on $(-13, 8)$ tubes (left). Three CA_{NTD} molecules were fitted there (right). c) Cropping of the trimer interface in Chimera. The location of a bead at the centre of mass of the cropped region is recorded for localised reconstruction. d) 3D classes obtained after classification of the trimer interface.

1.5 Discussion and future directions

This chapter provides a general framework for study of heterogenous tubular assemblies, with the case of HIV-1 CA tubes bound to the host factor CypA. In particular, the challenging process of symmetry indexing is detailed. This is a critical initial step for structure determination, since helical parameters must be refined to reach high resolution in the following steps of processing.

HIV-1 CA tubes represent a good case study, since they display a wide range of diameters and multiple symmetries that must be sorted before processing. Following indexing and validation of screw symmetry, refinement of tubes of different helical assemblies yielded different 3D reconstructions, in which CypA could be unambiguously observed above dimer interfaces between two adjacent hexamers.

Processing of particles of the most abundant symmetry $(-13, 8)$ is showcased here. In particular, symmetry expansion and localised reconstruction followed by 3D classification

and refinement were powerful to obtain a high resolution model of CA hexamer and an accurate model of CypA binding on the capsid.

Remarkably, the structure of a hexamer is asymmetric, since its constitutive protomers are not identical. Instead, their structure is modulated by flexibility of the linker between CA_{NTD} and CA_{CTD} of a protomer. This allows formation of newly observed curved hexamers, as opposed to all planar structures previously reported. This has implications for cone formation by the mature capsid. Indeed, the structure suggests that not only the interfaces between hexamers contribute to local curvature, but also plastic hexamers themselves.

Symmetry expansion and localised reconstruction were powerful to break helical symmetry and classify CypA binding at the dimer interface. From this, a novel non-canonical interface was found. This differs from the previously reported non-canonical interface obtained by molecular dynamics simulations after cryo-EM processing of tubes, but without localised reconstruction and classification [78].

Strikingly, CypA does not bind to all canonical interfaces at the same time. Instead, binding of CypA appears to be governed by local geometry and curvature that result from the arrangement of asymmetric and plastic hexamers. For instance, in the structure of CypA bound only at the dimer interface, alignment of the three protomers of the occupied dimer interface with that of an adjacent dimer interface shows that geometry is altered there and precludes CypA binding (Figure 1.23.a,b). In a minority of cases, changes in local geometry allows CypA to simultaneously bind on top of adjacent dimer and trimer interfaces.

The locations of CypA and that of some host factors are likely to be mutually exclusive in occupation. For instance, CypA binding sites overlap with that TRIM5α. This could explain how CypA modulates restriction by TRIM5α and a similar regulation may occur for restriction by other host factors, such as CPSF6. More structures of host factors bound to tubular or conical assemblies of CA are needed to confirm this regulatory role of CypA.

Following this work, the same processing strategy was used to visualise hexamers and CypA interfaces for other tubes of different symmetries [111]. From this, curvatures formed by hexamers on tubes of different symmetries were compared and differences of

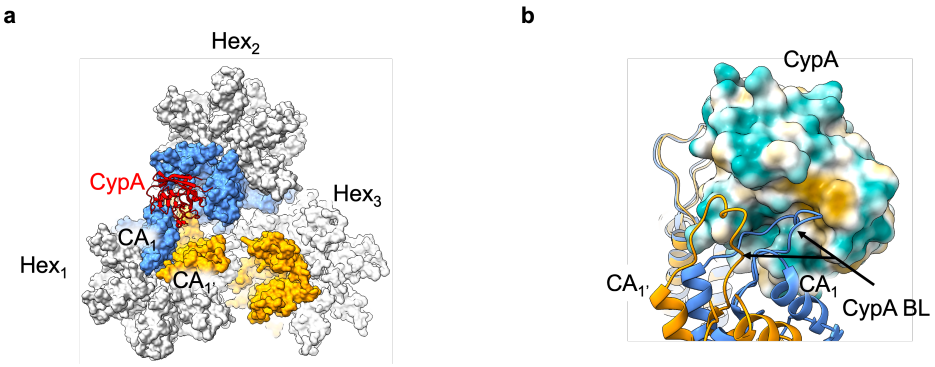


Figure 1.23: CypA binding is governed by local curvature. a) Surface representation of a trimer of hexamers derived from the structure of CypA (in red) bound at the dimer interface (three protomers coloured in blue). The adjacent dimer interface is empty (three protomers coloured in orange). b) Alignment of the three protomers from both dimer interfaces shown in b). In this conformation, the second interface (in orange) cannot accommodate CypA.

occupancy of CypA on CA tubes were analysed. This allowed mapping of CypA binding modes on a model of the mature capsid, based on changes in local curvature. Besides, in future experiments, mutational studies could be performed to further characterise binding of CypA at the non-canonical interface. In particular, the conserved triad of CypA residues there could be mutated and their effect on binding to and stability of the capsid could be investigated. In general, HIV-1 CA tubes stand out as relevant assemblies to predict binding on the mature conical capsid and the strategy presented here could be applied to any host factor.

References

[1] C. J. Pitcher et al. “HIV-1-specific CD4+ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged vital suppression”. In: *Nature Medicine* 5.5 (1999), pp. 518–525.

[2] A. M. Wensing et al. “2019 update of the drug resistance mutations in HIV-1”. In: *Topics in antiviral medicine* 27.3 (2019), pp. 111–121.

[3] P. Ramdas et al. “From Entry to Egress: Strategic Exploitation of the Cellular Processes by HIV-1”. In: *Frontiers in Microbiology* 11 (2020), p. 3021.

[4] E. O. Freed. “HIV-1 assembly, release and maturation”. In: *Nature Reviews Microbiology* 13.8 (2015), pp. 484–496.

[5] A. S. Reicin et al. “Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity”. In: *Journal of Virology* 69.2 (1995), pp. 642–650.

[6] A. Von Poblotzki et al. “Identification of a region in the pr55(gag)-polyprotein essential for HIV-1 particle formation”. In: *Virology* 193.2 (1993), pp. 981–985.

[7] C. T. Wang and E. Barklis. “Assembly, processing, and infectivity of human immunodeficiency virus type 1 gag mutants”. In: *Journal of Virology* 67.7 (1993), pp. 4264–4273.

[8] R. Welker et al. “Biochemical and Structural Analysis of Isolated Mature Cores of Human Immunodeficiency Virus Type 1”. In: *Journal of Virology* 74.3 (2000), pp. 1168–1177.

[9] C. Momany et al. “Crystal structure of dimeric HIV-1 capsid protein”. In: *Nature Structural Biology* 3.9 (1996), pp. 763–770.

[10] T. R. Gamble et al. “Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid”. In: *Cell* 87.7 (1996), pp. 1285–1294.

[11] C. Berthet-Colominas et al. “Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab”. In: *EMBO Journal* 18.5 (1999), pp. 1124–1136.

[12] R. K. Gitti et al. “Structure of the amino-terminal core domain of the HIV-1 capsid protein”. In: *Science* 273.5272 (1996), pp. 231–235.

[13] S. Abdurahman et al. “Characterization of the invariable residue 51 mutations of human immunodeficiency virus type 1 capsid protein on in vitro CA assembly and infectivity”. In: *Retrovirology* 4 (2007).

[14] U. K. Von Schwedler et al. “Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly”. In: *EMBO Journal* 17.6 (1998), pp. 1555–1568.

[15] S. Li et al. “Image reconstructions of helical assemblies of the HIV-1 CA protein”. In: *Nature* 407.6802 (2000), pp. 409–413.

[16] B. K. Ganser-Pornillos, A. Cheng, and M. Yeager. “Structure of Full-Length HIV-1 CA: A Model for the Mature Capsid Lattice”. In: *Cell* 131.1 (2007), pp. 70–79.

[17] O. Pornillos et al. “X-Ray Structures of the Hexameric Building Block of the HIV Capsid”. In: *Cell* 137.7 (2009), pp. 1282–1292.

[18] O. Pornillos, B. K. Ganser-Pornillos, and M. Yeager. “Atomic-level modelling of the HIV capsid”. In: *Nature* 469.7330 (2011), pp. 424–427.

[19] A. T. Gres et al. “X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability”. In: *Science* 349.6243 (2015), pp. 99–103.

[20] G. Zhao et al. “Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics”. In: *Nature* 497.7451 (2013), pp. 643–646.

[21] S. Mattei et al. “The structure and flexibility of conical HIV-1 capsids determined within intact virions”. In: *Science* 354.6318 (2016), pp. 1434–1437.

[22] O. Pornillos et al. “X-Ray Structures of the Hexameric Building Block of the HIV Capsid”. In: *Cell* 137.7 (2009), pp. 1282–1292.

[23] I. J. L. Byeon et al. “Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function”. In: *Cell* 139.4 (2009), pp. 780–790.

[24] Y. L. Li et al. “Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids”. In: *eLife* 5.JUNE2016 (2016), p. 582.

[25] S. S. Smaga et al. “MxB Restricts HIV-1 by Targeting the Tri-hexamer Interface of the Viral Capsid”. In: *Structure* 27.8 (2019), 1234–1245.e5.

[26] Z. Ambrose and C. Aiken. “HIV-1 uncoating: Connection to nuclear entry and regulation by host proteins”. In: *Virology* 454-455.1 (2014), pp. 371–379.

[27] D. A. Jacques et al. “HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis”. In: *Nature* 536.7616 (2016), pp. 349–353.

[28] A. E. Hulme, O. Perez, and T. J. Hope. “Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.24 (2011), pp. 9975–9980.

[29] Y. Yang, T. Fricke, and F. Diaz-Griffero. “Inhibition of Reverse Transcriptase Activity Increases Stability of the HIV-1 Core”. In: *Journal of Virology* 87.1 (2013), pp. 683–687.

[30] S. Rankovic et al. “Reverse Transcription Mechanically Initiates HIV-1 Capsid Disassembly”. In: *Journal of Virology* 91.12 (2017).

[31] K. Warren et al. “Eukaryotic elongation factor 1 complex subunits are critical HIV-1 reverse transcription cofactors”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.24 (2012), pp. 9587–9592.

[32] A. C. Francis et al. “HIV-1 uncoating and nuclear import precede the completion of reverse transcription in cell lines and in primary macrophages”. In: *Viruses* 12.11 (2020).

[33] V. Zila et al. “Cone-shaped HIV-1 capsids are transported through intact nuclear pores”. In: *Cell* 184.4 (2021), 1032–1046.e18.

[34] D. M. Sayah et al. “Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1”. In: *Nature* 430.6999 (2004), pp. 569–573.

[35] M. Stremlau et al. “The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys”. In: *Nature* 427.6977 (2004), pp. 848–853.

[36] S. L. Sawyer et al. “Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.8 (2005), pp. 2832–2837.

[37] C. A. Virgen et al. “Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.9 (2008), pp. 3563–3568.

[38] S. J. Wilson et al. “Independent evolution of an antiviral TRIMCyp in rhesus macaques”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.9 (2008), pp. 3557–3562.

[39] G. Brennan, Y. Kozyrev, and S. L. Hu. “TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.9 (2008), pp. 3569–3574.

[40] B. K. Ganser-Pornillos et al. “Hexagonal assembly of a restricting TRIM5α protein”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.2 (2011), pp. 546–550.

[41] K. A. Skorupka et al. “Hierarchical assembly governs TRIM5α recognition of HIV-1 and retroviral capsids”. In: *Science Advances* 5.11 (2019), eaaw3631.

[42] T. Pertel et al. “TRIM5 is an innate immune sensor for the retrovirus capsid lattice”. In: *Nature* 472.7343 (2011), pp. 361–365.

[43] J. M. Wagner et al. “Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5α”. In: *eLife* 5.JUNE2016 (2016).

[44] E. M. Campbell et al. “TRIM5α-Mediated Ubiquitin Chain Conjugation Is Required for Inhibition of HIV-1 Reverse Transcription and Capsid Destabilization”. In: *Journal of Virology* 90.4 (2016), pp. 1849–1857.

[45] M. Stremlau et al. “Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5α restriction factor”. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.14 (2006), pp. 5514–5519.

[46] C. R. Langelier et al. “Biochemical Characterization of a Recombinant TRIM5α Protein That Restricts Human Immunodeficiency Virus Type 1 Replication”. In: *Journal of Virology* 82.23 (2008), pp. 11682–11694.

[47] L. R. Black and C. Aiken. “TRIM5α Disrupts the Structure of Assembled HIV-1 Capsid Complexes In Vitro”. In: *Journal of Virology* 84.13 (2010), pp. 6564–6569.

[48] G. Zhao et al. “Rhesus TRIM5α disrupts the HIV-1 capsid at the inter-hexamer interfaces”. In: *PLoS Pathogens* 7.3 (2011).

[49] R. Lamichhane et al. “Dynamic conformational changes in the rhesus TRIM5α dimer dictate the potency of HIV-1 restriction”. In: *Virology* 500 (2017), pp. 161–168.

[50] S. Zhuang and B. E. Torbett. "Interactions of HIV-1 capsid with host factors and their implications for developing novel therapeutics". In: *Viruses* 13.3 (2021).

[51] R. A. Dick et al. "Inositol phosphates are assembly co-factors for HIV-1". In: *Nature* 560.7719 (2018), pp. 509–512.

[52] D. L. Mallery et al. "IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis". In: *eLife* 7 (2018), e1002762.

[53] D. L. Mallery et al. "Cellular IP6 Levels Limit HIV Production while Viruses that Cannot Efficiently Package IP6 Are Attenuated for Infection and Replication". In: *Cell Reports* 29.12 (2019), 3983–3996.e4.

[54] P. T. Huang et al. "FEZ1 Is Recruited to a Conserved Cofactor Site on Capsid to Promote HIV-1 Trafficking". In: *Cell Reports* 28.9 (2019), 2373–2385.e7.

[55] S. K. Carnes, J. Zhou, and C. Aiken. "HIV-1 Engages a Dynein-Dynactin-BICD2 Complex for Infection and Transport to the Nucleus". In: *Journal of Virology* 92.20 (2018).

[56] A. Dharan et al. "Bicaudal D2 facilitates the cytoplasmic trafficking and nuclear import of HIV-1 genomes during infection". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.50 (2017), E10707–E10716.

[57] V. Malikov et al. "HIV-1 capsids bind and exploit the kinesin-1 adaptor FEZ1 for inward movement to the nucleus". In: *Nature Communications* 6 (2015).

[58] Z. Lukic et al. "HIV-1 Uncoating Is Facilitated by Dynein and Kinesin 1". In: *Journal of Virology* 88.23 (2014), pp. 13613–13625.

[59] M. P. Dodding and M. Way. "Coupling viruses to dynein and kinesin-1". In: *EMBO Journal* 30.17 (2011), pp. 3527–3539.

[60] K. A. Matreyek et al. "Nucleoporin NUP153 Phenylalanine-Glycine Motifs Engage a Common Binding Pocket within the HIV-1 Capsid Protein to Mediate Lentiviral Infectivity". In: *PLoS Pathogens* 9.10 (2013).

[61] C. Buffone et al. "Nup153 Unlocks the Nuclear Pore Complex for HIV-1 Nuclear Translocation in Nondividing Cells". In: *Journal of Virology* 92.19 (2018).

[62] K. Bichel et al. "HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358". In: *Retrovirology* 10.1 (2013).

[63] R. C. Burdick et al. "Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes". In: *PLoS Pathogens* 13.8 (2017).

[64] A. De Iaco et al. "TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm." In: *Retrovirology* 10 (2013).

[65] G. N. Maertens et al. "Structural basis for nuclear import of splicing factors by human Transportin 3". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.7 (2014), pp. 2728–2733.

[66] A. Saito et al. "Capsid-CPSF6 Interaction Is Dispensable for HIV-1 Replication in Primary Cells but Is Selected during Virus Passage In Vivo". In: *Journal of Virology* 90.15 (2016), pp. 6918–6935.

[67] A. J. Price et al. "CPSF6 Defines a Conserved Capsid Interface that Modulates HIV-1 Replication". In: *PLoS Pathogens* 8.8 (2012).

[68] M. Kane et al. “MX2 is an interferon-induced inhibitor of HIV-1 infection”. In: *Nature* 502.7472 (2013), pp. 563–566.

[69] C. Goujon et al. “Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection”. In: *Nature* 502.7472 (2013), pp. 559–562.

[70] Z. Liu et al. “The interferon-inducible MxB protein inhibits HIV-1 infection”. In: *Cell Host and Microbe* 14.4 (2013), pp. 398–410.

[71] H. Ke et al. “Crystal structure of recombinant human T-cell cyclophilin A at 2.5 Å resolution”. In: *Proceedings of the National Academy of Sciences of the United States of America* 88.21 (1991), pp. 9483–9487.

[72] D. A. Bosco et al. “Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A”. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.8 (2002), pp. 5247–5252.

[73] E. K. Franke, H. E. H. Yuan, and J. Luban. “Specific incorporation of cyclophilin a into HIV-1 virions”. In: *Nature* 372.6504 (1994), pp. 359–362.

[74] J. Luban et al. “Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B”. In: *Cell* 73.6 (1993), pp. 1067–1078.

[75] T. Fricke et al. “Human Cytosolic Extracts Stabilize the HIV-1 Core”. In: *Journal of Virology* 87.19 (2013), pp. 10587–10597.

[76] S. Yoo et al. “Molecular recognition in the HIV-1 Capsid/Cyclophilin A complex”. In: *Journal of Molecular Biology* 269.5 (1997), pp. 780–795.

[77] W. Peng et al. “Functional analysis of the secondary HIV-1 capsid binding site in the host protein cyclophilin A”. In: *Retrovirology* 16.1 (2019).

[78] C. Liu et al. “Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site”. In: *Nature Communications* 7 (2016).

[79] D. Braaten and J. Luban. “Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells”. In: *EMBO Journal* 20.6 (2001), pp. 1300–1309.

[80] Z. Ambrose et al. “Human Immunodeficiency Virus Type 1 Capsid Mutation N74D Alters Cyclophilin A Dependence and Impairs Macrophage Infection”. In: *Journal of Virology* 86.8 (2012), pp. 4708–4714.

[81] Y. Li, A. K. Kar, and J. Sodroski. “Target Cell Type-Dependent Modulation of Human Immunodeficiency Virus Type 1 Capsid Disassembly by Cyclophilin A”. In: *Journal of Virology* 83.21 (2009), pp. 10951–10962.

[82] T. Hatziioannou et al. “Cyclophilin Interactions with Incoming Human Immunodeficiency Virus Type 1 Capsids with Opposing Effects on Infectivity in Human Cells”. In: *Journal of Virology* 79.1 (2005), pp. 176–183.

[83] A. De Iaco and J. Luban. “Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA”. In: *Retrovirology* 11.1 (2014).

[84] V. B. Shah et al. “The Host Proteins Transportin SR2/TNPO3 and Cyclophilin A Exert Opposing Effects on HIV-1 Uncoating”. In: *Journal of Virology* 87.1 (2013), pp. 422–432.

[85] Z. Zhong et al. “Cytoplasmic CPSF6 regulates HIV-1 capsid trafficking and infection in a cyclophilin A-dependent manner”. In: *mBio* 12.2 (2021).

[86] A. Selyutina et al. “Cyclophilin A Prevents HIV-1 Restriction in Lymphocytes by Blocking Human TRIM5α Binding to the Viral Core”. In: *Cell Reports* 30.11 (2020), 3766–3777.e6.

[87] T. Schaller et al. “HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency”. In: *PLoS Pathogens* 7.12 (2011), e1002439.

[88] K. A. Matreyek and A. Engelman. “The Requirement for Nucleoporin NUP153 during Human Immunodeficiency Virus Type 1 Infection Is Determined by the Viral Capsid”. In: *Journal of Virology* 85.15 (2011), pp. 7818–7827.

[89] S. Q. Zheng et al. “MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy”. In: *Nature Methods* 14.4 (2017), pp. 331–332.

[90] G. Tang et al. “EMAN2: An extensible image processing suite for electron microscopy”. In: *Journal of Structural Biology* 157.1 (2007), pp. 38–46.

[91] M. Van Heel et al. “A new generation of the IMAGIC image processing system”. In: *Journal of Structural Biology* 116.1 (1996), pp. 17–24.

[92] J. Frank et al. “SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields”. In: *Journal of Structural Biology* 116.1 (1996), pp. 190–199.

[93] S. He and S. H. W. Scheres. “Helical reconstruction in RELION”. In: *Journal of Structural Biology* 198.3 (2017), pp. 163–176.

[94] D. J. De Rosier and A. Klug. “Reconstruction of three dimensional structures from electron micrographs”. In: *Nature* 217.5124 (1968), pp. 130–134.

[95] C. Sachse. “Single-particle based helical reconstruction-how to make the most of real and Fourier space”. In: *AIMS Biophysics* 2.2 (2015), pp. 219–244.

[96] M. Stewart. “Computer image processing of electron micrographs of biological structures with helical symmetry”. In: *Journal of Electron Microscopy Technique* 9.4 (1988), pp. 325–358.

[97] E. D. T. Atkins. “Crystal Structure by X-Ray Diffraction”. In: *Comprehensive Polymer Science and Supplements*. Pergamon, 1989, pp. 613–650.

[98] J. M. Squire and C. Knupp. “X-ray diffraction studies of muscle and the crossbridge cycle”. In: *Advances in Protein Chemistry* 71 (2005), pp. 195–255.

[99] A. Klug, F. H. C. Crick, and H. W. Wyckoff. “Diffraction by helical structures”. In: *Acta Crystallographica* 11.3 (1958), pp. 199–213.

[100] W. Cochran, F. H. Crick, and V. Vand. “The structure of synthetic polypeptides. I. The transform of atoms on a helix”. In: *Acta Crystallographica* 5.5 (1952), pp. 581–586.

[101] E. H. Egelman. “A robust algorithm for the reconstruction of helical filaments using single-particle methods”. In: *Ultramicroscopy* 85.4 (2000), pp. 225–234.

[102] E. H. Egelman. “Reconstruction of helical filaments and tubes”. In: *Methods in Enzymology*. Vol. 482. C. 2010, pp. 167–183.

[103] K. Zhang. “Gctf: Real-time CTF determination and correction”. In: *Journal of Structural Biology* 193.1 (2016), pp. 1–12.

[104] S. L. Ilca et al. “Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes”. In: *Nature Communications* 6 (2015).

[105] E. F. Pettersen et al. “UCSF Chimera - A visualization system for exploratory research and analysis”. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612.

[106] J. M. de la Rosa-Trevín et al. “Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy”. In: *Journal of Structural Biology* 195.1 (2016), pp. 93–99.

[107] P. Emsley and K. Cowtan. “Coot: Model-building tools for molecular graphics”. In: *Acta Crystallographica Section D: Biological Crystallography* 60.12 I (2004), pp. 2126–2132.

[108] P. V. Afonine et al. “Real-space refinement in PHENIX for cryo-EM and crystallography”. In: *Acta Crystallographica Section D: Structural Biology* 74.6 (2018), pp. 531–544.

[109] H. Ashkenazy et al. “ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules”. In: *Nucleic Acids Research* 44.W1 (2016), W344–W350.

[110] J. Jiang et al. “The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability”. In: *Virology* 421.2 (2011), pp. 253–265.

[111] T. Ni et al. “Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A”. In: *Nature Structural and Molecular Biology* 27.9 (2020), pp. 855–862.

2

Structure of the inhibited state of the Sec translocon

2.1 Summary

Across all kingdoms of life, about a third of all synthesised proteins cross or integrate into a membrane. This process, named translocation, mainly occurs at the endoplasmic reticulum (ER) membrane in eukaryotes and at the plasma membrane in prokaryotes, and is mainly mediated by the Sec translocon. Remarkably, the pathogen *Mycobacterium ulcerans* secretes mycolactone, a toxin that blocks Sec-mediated translocation in human cells and responsible for the development of Buruli ulcer. This chapter describes the first structure of the inhibited state of the Sec translocon and provides the unexpected outcome that the inhibitor holds the translocon in a partly activated state, while blocking access of the signal peptide to the lateral gate. This explains how small molecule inhibitors, including mycolactone can block translocation through the Sec apparatus.

2.2 Introduction

2.2.1 The importance of translocation in protein biogenesis

2.2.1.1 Protein synthesis by the ribosome factory

In the cytosol, ribosomes assemble and mediate synthesis of all proteins [112, 113]. Ribosomes are big molecular machines, that consist of ribonucleic acids and ribosomal proteins, with their assembly defined by their coefficient sedimentation in Svedberg units (S). In eukaryotes, 60S large ribosomal subunits associate with 40S small subunits to form 80S ribosomes (Figure 2.1.a). This assembly provides a necessary scaffold which ensures efficient translation [114].

The 40S subunit consists of a head, a body and a neck (Figure 2.1.b). The mRNA binding cleft is located along the neck, between the body and the head [115–117]. mRNAs consist of a sequence of nucleotides, three of which are read at a time by ribosomes. These codons are translated into different amino-acids, depending on their arrangement in nucleic acids. During translation, the 40S neck presents series of codons, on which transfer ribonucleic acids (tRNAs) dock [113, 118]. These deliver corresponding amino-acids that form polypeptide chains [119, 120]. The 40S subunit binds and reads mRNA, while the 60S subunit catalyses peptide bond formation through the action of the peptidyl-transferase centre (PTC) [121].

The 60S subunit also forms a tunnel, through which a growing polypeptide exits (Figure 2.1.c). The tunnel can accommodate and protect a nascent chain of 30 to 40 amino-acids [122, 123]. The region surrounding the exit tunnel, where the nascent chain emerges, interacts with a variety of proteins responsible for targeting, folding and translocation through the ER [124]. In addition, the 60S subunit harbours a site responsible for activation of translational guanosine triphosphate hydrolases (GTPases) that mediate guanosine triphosphate (GTP) hydrolysis, required to ensure fast translation [125, 126].

Translation consists of four steps: initiation [124], elongation [127–129], termination and recycling [130] of the ribosome subunits (Figure 2.1.d). Through these, high fidelity and optimal speed are ensured by the action of various protein factors.

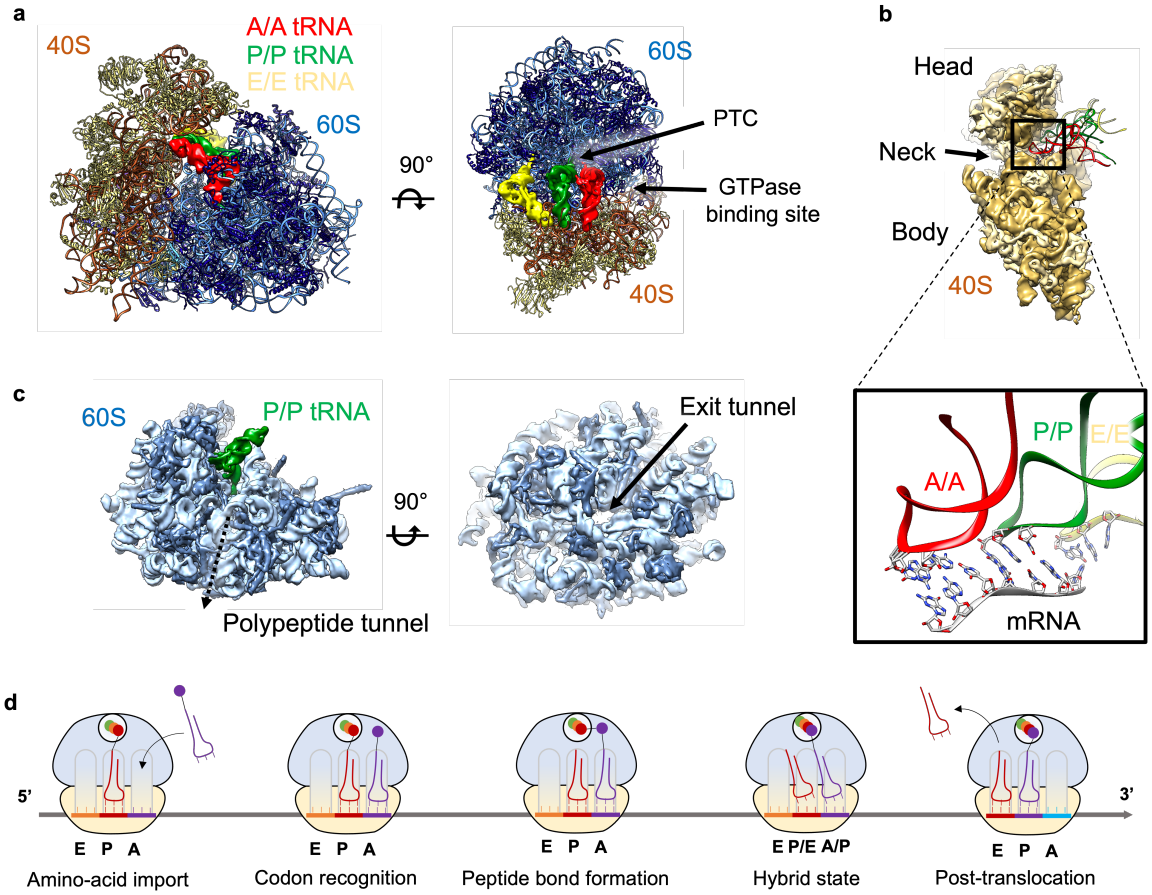


Figure 2.1: Eukaryotic protein synthesis by the ribosome. a) Ribbon representation of the 80S ribosome and occupancy of tRNAs at A/A-site, P/P-site and E/E-site (in red, green and yellow, respectively). The 40S and 60S subunits are depicted in gold and blue, respectively, with constitutive ribosomal RNA darker than proteins. b) Surface representation of the 40S subunit bound to P/P-site tRNA using the same colour code as in a). A close-up on the mRNA association with complementary tRNAs is shown. c) Surface representation of the 40S subunit bound to tRNAs using the same colour code as in a). d) Simplified schematic representation of the different states of tRNAs during polypeptide elongation. After import of tRNA at the A/A-site, a peptide bond is formed and the polypeptide transferred to the A/A-site. Then, a hybrid state is formed at P/E and A/P sites, followed by translocation of hybrid tRNAs to P/P- and E/E-sites. This gives space for a new tRNA at the A/A-site (figure adapted from [127]).

The ribosome harbours three tRNA binding sites, which span the space between both 60S and 40S subunits. These are the aminoacyl site (A-site), the peptidyl site (P-site) and the exit site (E-site). The state of tRNA during protein synthesis is defined by both sites it occupies on the 40S and 60S subunits, respectively. For instance, tRNA located at the A-site of both ribosome subunits occupies the A/A-site.

Translation is initiated by binding of mRNA to the 40S subunit, finding of the start codon (AUG), binding of the 60S subunit and association with an initiator methionyl

tRNA. The latter binds at the ribosomal P/P-site, responsible for binding tRNAs attached to the growing polypeptide chain.

During elongation, a polypeptide chain is synthesised. First, an aminoacyl-tRNA is delivered by the elongation factor eEF1 to the A/A-site, adjacent to the P/P-site. The A/A site binds to aminoacyl-tRNA for subsequent formation of a peptide bond. The tRNA covalently linked to the nascent polypeptide (peptidyl-tRNA) at the P/P-site is cleaved, and the nascent polypeptide forms a new peptide bond with the aminoacyl-tRNA. During peptide bond formation, the 40S and 60S subunits adopt a rotated conformation, in which tRNAs occupy the A- and P-sites of the 40S subunit, while they are positioned at the P- and E-sites of the 60S subunit, respectively. These tRNAs, originally in P/P and A/A states, adopt 'hybrid' P/E and A/P states. Subsequently, the elongation factor eEF2 catalyses translocation of tRNAs from A/P- and P/E-sites to P/P- and E/E-sites, respectively.

By the end of translocation, mRNA moves over by one codon and the A/A-site is vacant. A new cycle starts with the addition of the next aminoacyl-tRNA at the A/A-site, and repeats until the ribosome reaches a stop codon.

2.2.1.2 The endoplasmic reticulum: a central organelle for biogenesis of secretory and membrane proteins

After synthesis by ribosomes, proteins eventually adopt specific folds, linked to particular tasks, fulfilled in unique locations inside or outside the cell. In eukaryotes, this is even more true, as cells are composed of many organelles. Protein segregation from the cytosol to their final location is therefore essential.

Most eukaryotic membrane and secretory proteins are targeted to the ER [131]. Here, cytosolic polypeptides are translocated through the ER-embedded translocation channel [132]. Soluble proteins destined for secretion are entirely transferred to the ER lumen, while membrane proteins are inserted into the ER membrane, with some parts still exposed to the cytosol and others to the ER lumen.

During and after translocation, various protein complexes in the ER mediate maturation of polypeptides into functional proteins [133]. Different systems exclusively ensure proper folding, formation of disulfide bonds, multimeric assembly, proteolytic

cleavage, glycosylation or addition of glycosylphosphatidylinositol (GPI) anchor. Then, proteins are sorted and exported to other constituents of the endomembrane system [134, 135], such as the Golgi apparatus, lysosomes, endosomes and secretory vesicles shuttled to the plasma membrane and the extracellular environment (Figure 2.2). Immature proteins, with defects that cause misfolding or incorrect multimeric forms, are retained or transported back to the ER, where they undergo ER-associated degradation [136]. There, protein defects are recognised by chaperones, targeted to a retro-translocon, retrotranslocated from the ER back to the cytosol, ubiquitinated and delivered to the proteasome.

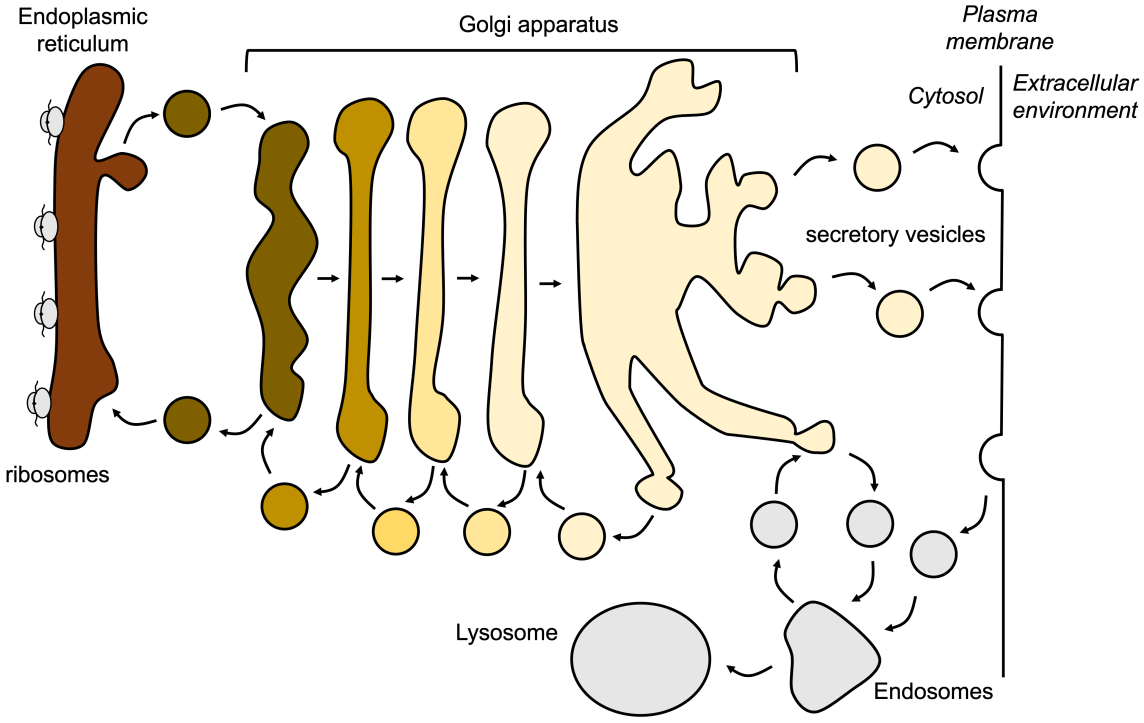


Figure 2.2: Conventional protein secretory pathway in eukaryotes. After synthesis by ribosomes in the cytosol, polypeptides are translocated into the ER lumen, where they undergo maturation. Then, mature proteins are sorted at the Golgi apparatus. They can reach the extracellular environment through secretory vesicles, retrograde to earlier sites of maturation or be degraded by lysosomes.

Research over the past decades has provided insights into how proteins are translocated across the ER, modified in the lumen, transported to their destination via vesicular trafficking or sent back to the cytosol for degradation. These have depicted the ER as a key organelle in the regulation of protein biogenesis, maturation and quality control.

2.2.1.3 Protein targeting to the ER: segregation of secretory and membrane proteins

In eukaryotes, mRNA sequences of most secretory and some membrane proteins start with an N-terminal signal peptide that contains a core of around 7 to 15 hydrophobic residues [131, 137]. After initiation of translation, the protrusion of a signal peptide from the tunnel exit of a ribosome-nascent chain complex (RNC) attracts a signal recognition particle (SRP) [138]. Binding of the SRP to the hydrophobic region of the signal peptide triggers pausing of translation, which resumes once the SRP docks on its ER membrane-associated receptor (SR) [138–140].

Subsequently, release of SRP from the polypeptide is followed by transfer of the RNC to the translocation channel [132], where translation resumes. Eventually, polypeptides are translocated through the ER membrane by the translocon, while they are translated. This process, called co-translational translocation [131], is energetically driven by the protein synthesis reaction (Figure 2.3.a).

A minority of proteins cannot be targeted to the ER through co-translational translocation. Instead, they are guided there after termination of translation in the cytosol, through a post-translational route [141] (Figure 2.3.b). This is the case of small secretory proteins, for which translation is complete before recognition by SRP can occur. In this case, chaperones such as Heat shock protein 70, stabilise polypeptides in the cytosol and deliver them to the translocon in an unfolded state that is prone to translocation. There, translocation is energetically driven by adenosine triphosphate hydrolysis performed by accessory components of the translocon.

Likewise, tail-anchored proteins, which are membrane proteins with a single transmembrane domain at the C-terminal end of their sequence, are also translocated post-translationally. However in this case, polypeptides interact with very different cytosolic factors and subsequent insertion into the ER membrane occurs through receptors distinct from the translocon [142].

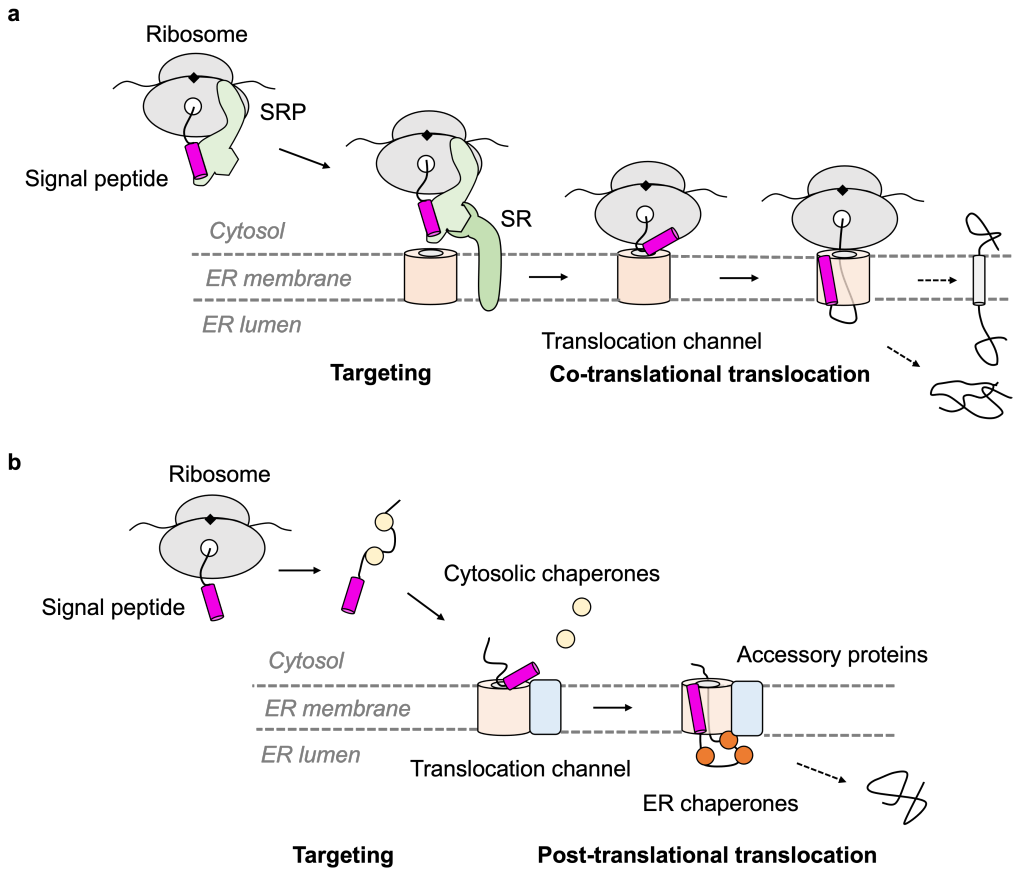


Figure 2.3: Co-translational and post-translational translocations. a) Simplified schematic representation of the proteins involved at the ER during co-translational translocation. A signal peptide (in magenta), which protrudes from the ribosome (in grey) tunnel exit is recognised by a SRP (in light green). The complex is targeted to the ER-associated SR (in green), followed by transfer to the protein conducting channel (in salmon), where translocation of secretory and membrane proteins occurs. b) Simplified schematic representation of the molecules involved at the ER during post-translational translocation. A small secretory protein exits from the ribosome tunnel exit and is stabilised by cytosolic chaperones (in yellow). The complex is targeted to the translocon, where post-translational translocation occurs through the help of ER chaperone (in orange) and accessory proteins (in blue). Colour codes are the same as in a) for identical components.

2.2.2 The central role of the protein translocation channel

2.2.2.1 Identification of the protein conducting channel

The discovery of signal sequences and their importance in the secretory pathway were followed by identification of components involved in protein translocation.

In *E. coli*, genetic mutations responsible for efficient translocation of proteins with defects in their signal sequences were identified. These protein localisation (*prl*) mutants were found in three *sec* genes, *secY*, *secA* and *secE*, important for protein secretion [143–151].

Later, the respective yeast and mammalian homologues were found. Around the same time, electrophysiological experiments demonstrated that the translocon provides an aqueous environment that facilitates polypeptide translocation through the ER [132, 152]. Eventually, complexes of the three proteins encoded by *E. coli* or mammalian *sec* genes were able to translocate unfolded polypeptides into reconstituted transport systems, such as proteoliposomes [153–157]. Therefore, these proteins were identified as minimal components of the 'Sec translocon'.

2.2.2.2 An evolutionary conserved complex

The protein composition of Sec is evolutionary conserved. In eukaryotes, the protein complex Sec61 mediates translocation of secretory and membrane proteins across the ER. In bacteria and archea, the respective homologues SecYEG and SecYEβ translocate proteins from the cytosol across the cytoplasmic membrane. Structural snapshots of Sec homologues were obtained by X-ray crystallography and cryo-EM [158–180]. Together, they highlight universally conserved structural features.

Sec61 is a heterotrimer, which consists of the channel-forming subunit Sec61α (Y in bacteria), associated with smaller subunits Sec61β (SecG) and Sec61γ (SecE). Sec61α constitutes the central 10-transmembrane (TM) core, which forms an hourglass-shaped pore, through which polypeptides are translocated. Two halves formed by TM helices 1-5 and 6-10 (H_1-H_5, H_6-H_{10}) are arranged in a pseudo-symmetric manner (Figure 2.4.a,b).

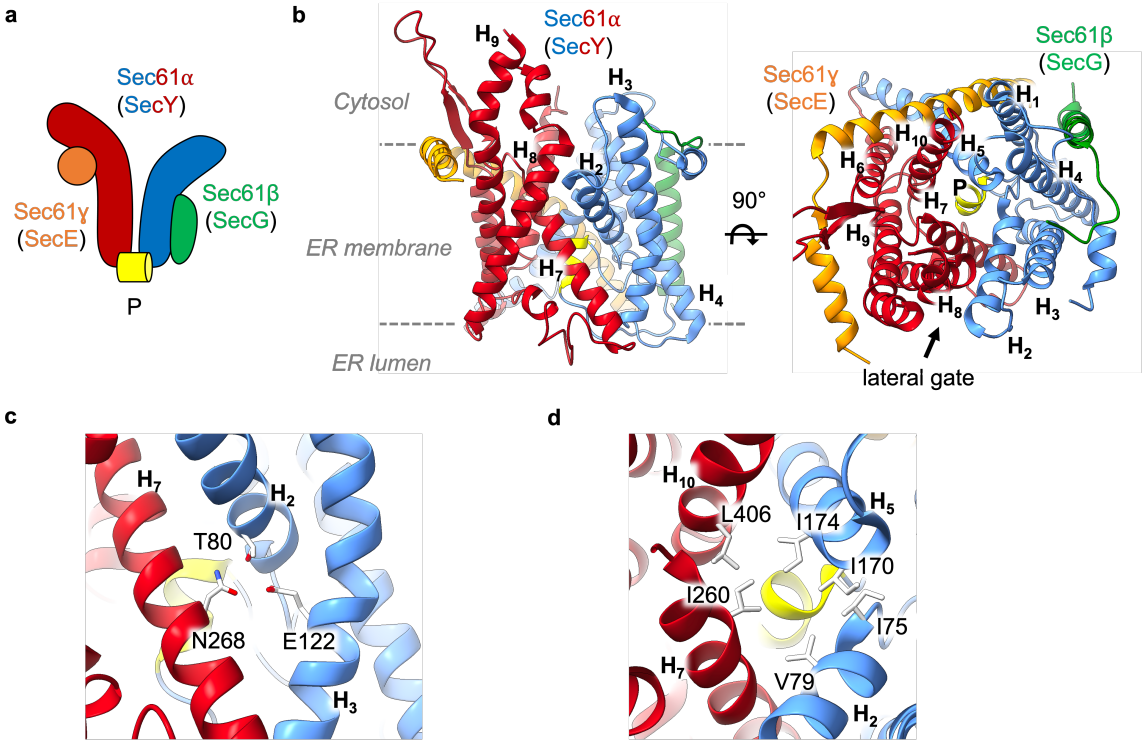


Figure 2.4: Structural features of the Sec translocon. a) Schematic representation of the Sec translocon. Sec61α (SecY) N-terminal TM helices H_1 to H_5 are coloured in blue, while C-terminal TM helices H_6 to H_{10} are coloured in red. The plug helix (P) is depicted in yellow. Sec61γ (SecE) and Sec61β (SecG) subunits are coloured in orange and green, respectively. b) Ribbon representation of the idle Sec translocon, with same colour code as in a). The plug (P) and TM helices of the pore-forming protein are labeled. The structure depicted is that of the archeal idle Sec complex (PDB code 1RH5) [158]. c) Ribbon and sticks close-up on the polar cluster of residues that maintain the lateral gate in a closed state. Residues T80, E122 and N268 in archea correspond to T87, Q127 and N300 in human Sec61. d) Ribbon and sticks close-up on the pore-lining hydrophobic residues. Residues I75, V79, I170, I174, I260 and L406 in archea correspond to I81, V85, I179, I183, I292 and L449 in human Sec61.

Sec61α has two opening sites, one for signal peptide insertion into the ER membrane, and one for translocation of growing polypeptides through the ER lumen [158]. Signal peptide insertion occurs at the lateral gate, a seam along the side of Sec61α formed by the TM helix pairs 2,3 and 7,8 (H_2, H_3 and H_7, H_8). In addition, the lumenal side of the central pore of Sec61α is gated by a small helix, which acts as a 'plug' [158] during translocation of polypeptides.

In unbound idle translocons, the lateral gate is closed, and the lumenal side of the central pore is gated by the plug [158]. In particular, a cluster of three conserved polar residues at the lateral gate and a set of conserved hydrophobic pore-ring residues at the cytosolic side of the pore maintain Sec61 closed conformation (Figure 2.4.c,d). Residues

of the plug region are the least conserved between different species. Plug domains have distinct structures and in some cases, their deletion is followed by formation of new plugs from surrounding residues [181]. Importantly, sealing of the channel also prevents leakage of ions and other molecules into or out of the ER. Besides, the loops that connect TM helices 6-7 (H_6-H_7) and 8-9 (H_8-H_9) extend into the cytosol to provide a platform for binding partners during translocation.

Sec61γ acts as a clamp, which consists of two helices connected by a small hinge. Together, they wrap around the two halves of Sec61α. The N-terminal helix binds to the C-terminal half of Sec61α at its cytosolic side, horizontally. The C-terminal helix joins both halves at the back of Sec61α, diagonally. This wrapping stabilises Sec61α, while allowing breathing at the side where the lateral gate is exposed to the lipid phase. Last, Sec61β is a tail-anchored protein, in proximity to TM helices 1 and 4 (H_1 and H_4) of Sec61α. In structures, only the TM helix is resolved, while the cytosolic domain is disordered. The role of Sec61β remains unclear, although it might be involved in binding to ribosomes and SR during transfer of RNC after release of SRP.

2.2.2.3 A dynamic channel

The Sec translocon mediates selective transfer of polypeptides across membranes, while preventing other molecules to pass through. To do so, dynamics of translocation must be tightly regulated.

2.2.2.3.1 Sec61 during co-translational translocation

During co-translational translocation, Sec61 associates with a RNC, close to the exit site of its polypeptide tunnel, through the C-terminal cytosolic loops of Sec61α linking H_6-H_7 and H_8-H_9 and the N-terminal helix of Sec61γ. Binding of the macromolecule triggers cracking of the lateral gate at its cytoplasmic side, while the central pore remains sealed (Figure 2.5). The translocon adopts a primed state, in which it is ready to receive a nascent chain [170]. Notably, three residues (L89, I179, L293) form a hydrophobic patch deep inside the central cavity of Sec61α, which may attract hydrophobic signal peptides.

In structures of signal peptide-engaged Sec61, this hydrophobic patch is broken, the N-terminal half of Sec61 is rotated and the lateral gate opens at its lumenal side, allowing

intercalation of a signal peptide on its outer surface [167, 175]. This movement is coupled with opening of the pore through displacement of the plug away from the central cavity. In this conformation, Sec61 can translocate polypeptides across the ER membrane through the pore and insert membrane proteins into the lipid phase through the lateral gate.

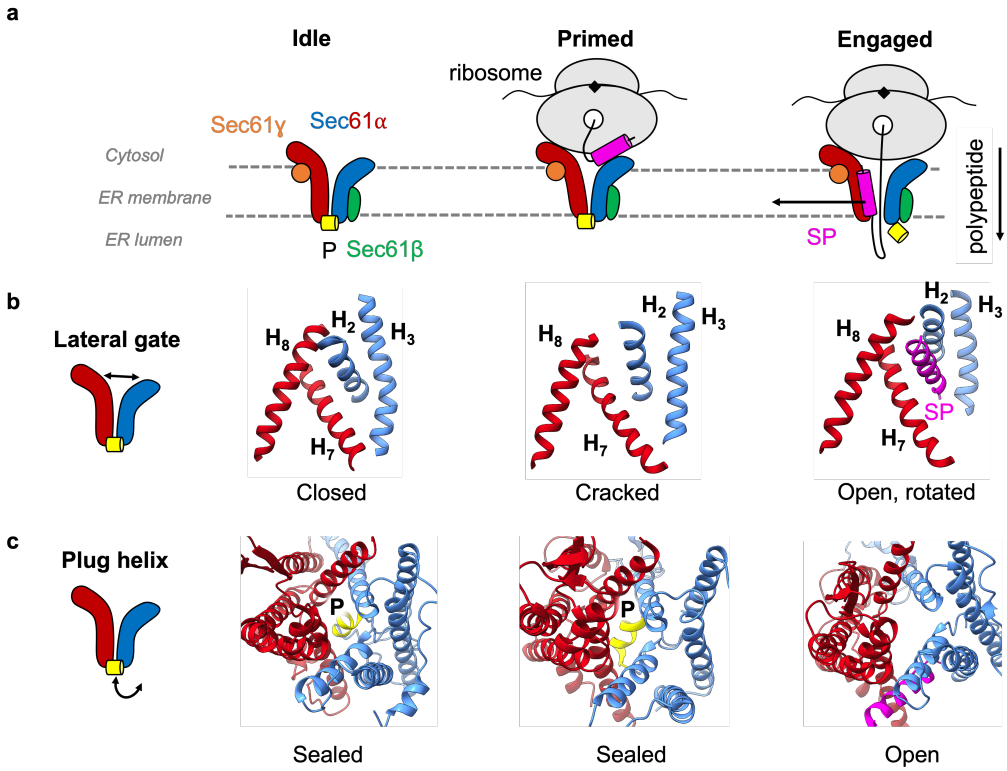


Figure 2.5: Dynamics of Sec61 during co-translational translocation. a) Schematic representation of Sec61 states during co-translational translocation. Sec61α N-terminal TM helices H_1 to H_5 are coloured in blue, while Sec61α C-terminal TM helices H_6 to H_{10} are coloured in red. The plug helix (P) is depicted in yellow. Sec61γ and Sec61β subunits are coloured in orange and green, respectively. The signal peptide (SP) is coloured in magenta. b) Schematic and ribbon representations of Sec61 movement at the lateral gate in states matching those shown above in a), with the same colour code. Helices H_7-H_8 were aligned for each state of Sec61α. A first opening is provided by movement of pairs of helices H_2-H_3 and H_7-H_8. Cracking at the cytosolic side first occurs during binding of a RNC, followed by movement of helices H_2-H_3 to accommodate a signal peptide in a position similar to that of H_2 in the idle state. c) Schematic and ribbon representations of Sec61 movement of the plug helix in states matching those shown above in a), with the same colour code. Opening at the lateral gate is coupled with displacement of the plug domain to allow passage of translocating polypeptides. PDB codes for idle, primed and engaged states are 1RH5, 3J7Q and 3JC2, respectively [158, 167, 170].

2.2.2.3.2 Sec61 during post-translational translocation

During eukaryotic post-translational translocation, Sec61 undergoes different conformational changes, as it binds to different partners [182].

In yeasts, Sec61 interacts with the Sec62/Sec63/Sec71/Sec72 complex, which facilitates substrate import together with the transiently-associated lumenal chaperone BiP [159, 160, 163, 166]. Sec63 interacts with Sec61 at the cytosolic side through the loops linking H_6-H_7 and H_8-H_9, similarly to ribosomes during co-translational translocation. Sec71 and Sec72, which act as receptors for post-translational substrates, associate with Sec63 and lock it in place. Association with Sec63 causes significant opening of Sec61 at the cytosolic and lumenal sides of the lateral gate, while the plug seals the pore (Figure 2.6). Then, Sec62 interacts with the lateral gate of Sec61, which further widens at its lumenal side through rotation of the N-terminal half of Sec61α. In this state, Sec61 is ready to accept substrates [159]. Upon engagement with a signal peptide, the conformation remains similar to that observed with Sec62 prising open the translocon. A signal peptide is accommodated along the seam [166] and, as in co-translational translocation, occupies the position of H_6 observed in the inactive state.

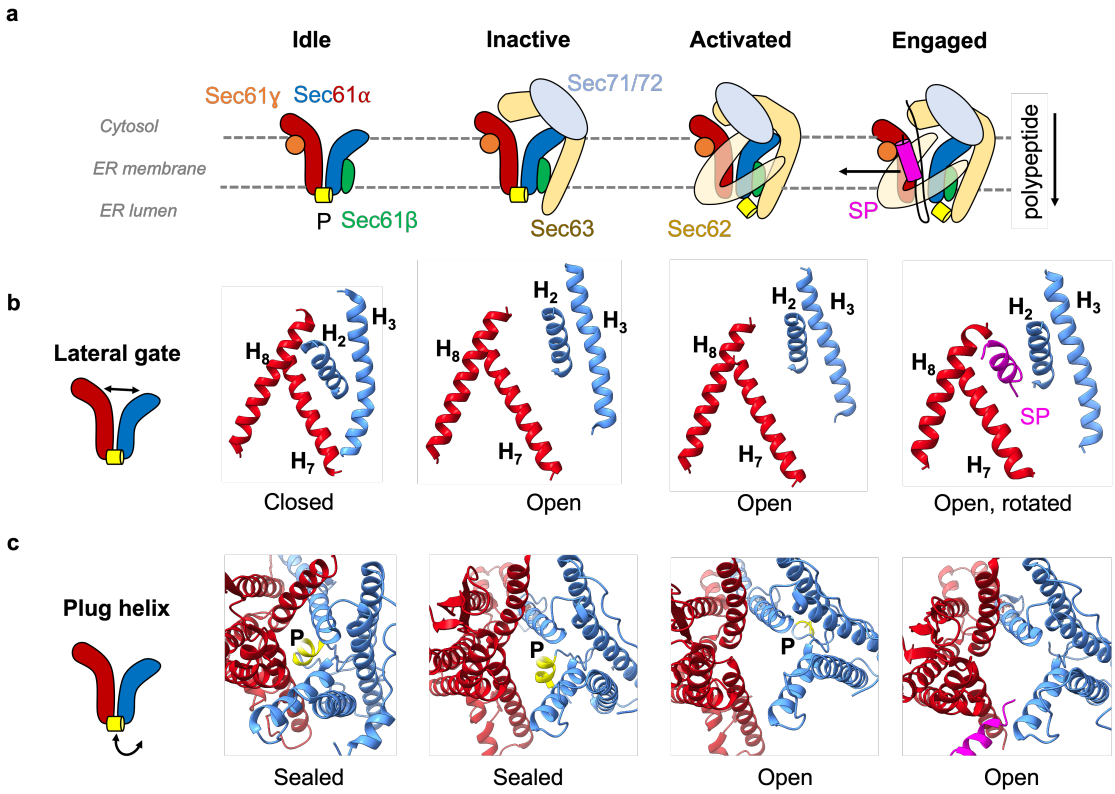


Figure 2.6: Dynamics of Sec61 during post-translational translocation. a) Schematic representation of Sec61 states during post-translational translocation. Sec61α N-terminal TM helices H_1 to H_5 are coloured in blue, while Sec61α C-terminal TM helices H_6 to H_{10} are coloured in red. Sec61γ and Sec61β subunits are coloured in orange and green, respectively. Sec71 and Sec72 are coloured in light blue, Sec63 is in gold and Sec62 in transparent light gold. The signal peptide (SP) is coloured in magenta. b) Schematic and ribbon representations of Sec61 movement at the lateral gate in states matching those shown above in a), with the same colour code. Helices H_7-H_8 were aligned for each state of Sec61α. Opening of the lateral gate occurs upon binding of Sec63, followed by a bigger opening upon binding of Sec62. During engagement, a signal peptide inserts a similar channel, in a position that coincides with H_2 in the idle state. c) Schematic and ribbon representations of Sec61 movement of the plug helix in states matching those shown above in a), with the same colour code. Opening at the lateral gate is coupled with displacement of the plug domain to activate the channel and allow passage of translocating polypeptides. In the activated state, the plug helix is displaced near the C-terminus of Sec61γ. PDB codes for idle, inactive, activated and engaged states are 1RH5, 7KAH, 7KAI and 7AFT, respectively [158, 159, 166].

2.2.2.3.3 The path of substrates during engagement with Sec61

Through polypeptide translocation, the lateral gate provides the site for insertion of hydrophobic stretches, while pore opening allows vertical transfer of hydrophilic segments. The structures of signal-peptide engaged translocons suggest that nascent chains are inserted as hairpins in a tail-first conformation rather than a head-first conformation [167, 175]. Nonetheless, the dynamics of insertion of hydrophobic segments into the ER

membrane remains uncertain. On one hand, insertion could be initiated by interaction of substrates with Sec61α central core, triggering opening of the lateral gate, allowing subsequent lateral movement through the lateral gate into the lipid phase. On the other hand, upon docking of a RNC on Sec61, the lateral gate may oscillate between open and closed states, allowing partitioning of substrates according to their hydrophobicity.

Opening of the lateral gate at the cytoplasmic side provides an entry site for hydrophobic substrates, followed by their subsequent sliding along the outside of the lateral seam through the lipid phase, and subsequent displacement of the plug helix for translocation of hydrophilic polypeptides.

2.2.3 Study of Sec61 protein inhibition

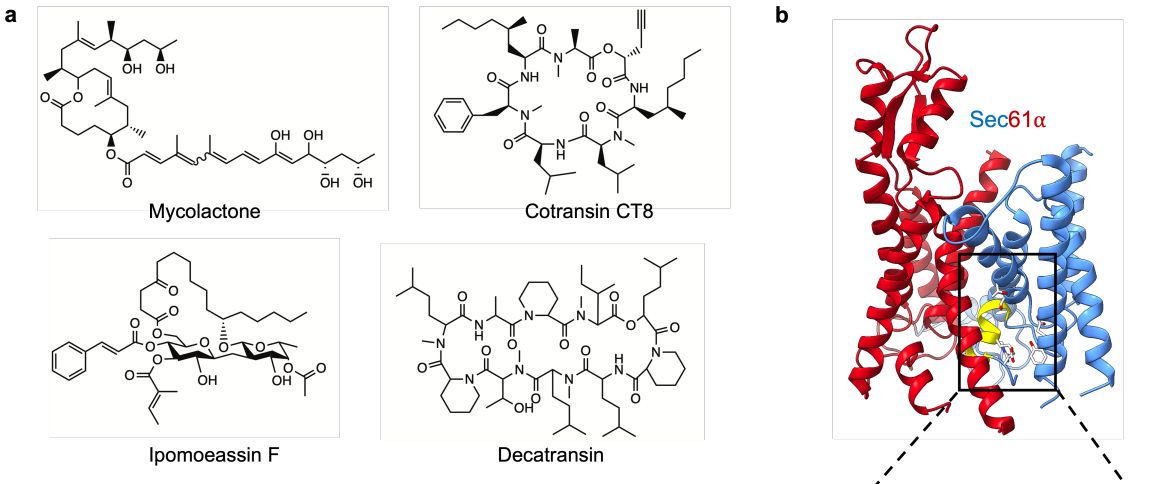
Because protein biogenesis is essential for cells to be functional, some microbes and plants have evolved to produce a wide variety of inhibitors of protein translocation [183–186]. These act at different stages of protein biogenesis and differentially on both types of translocation.

A number of inhibitors specifically target the Sec61 machinery. Most of them are gating inhibitors thought to affect the dynamics of translocation and to share a common mechanism of action, though they have different chemical structures. Several gating inhibitors display anti-cancer, anti-inflammatory or anti-viral effects. Therefore, understanding their action on Sec61 is essential for potential drug therapy. In addition, molecules which trap translocation intermediates represent valuable tools that can be used to decipher the dynamics of this process.

2.2.3.1 A shared mechanism of action?

Inhibitors of protein translocation are produced by different organisms and have different chemical structures (Figure 2.7.a). Among them, cotransins are cyclodepsipeptide derivatives of a molecule originally isolated from a fungal broth in a screen for inhibitors of adhesion molecule expression. Similarly, the fungus *Chaetosphaeria tulasneorum* produces the decadepsipeptide decatransin, which was found through a cancer screen [187]. Other screens for anti-tumor drugs identified bacterial compounds

such as apratoxins [188] and coibamide A [189], respectively produced by the *Moorea bouillomii* and *Leptolyngbya* species of marine cyanobacteria, and plant ipomoeassins [190] produced by *Ipomoea squamosa*. Other inhibitors are the effectors of bacterial virulence, such as *Pseudomonas aeruginosa* exotoxin A [191] and mycolactone produced by *Mycobacterium ulcerans*.



c

	Residue	Inhibitors
H_1	I41	decatransin
	D60	decatransin
Plug	M65	decatransin
	R66	cotransin, decatransin, mycolactone, ipomoeassin F, coibamide A
	S71	decatransin, coibamide A
H_2	G80	cotransin, decatransin
	S82	cotransin, decatransin, mycolactone, ipomoeassin F
	T86	apratoxin A
H_3	Y131	apratoxin A
	M136	cotransin, decatransin

Figure 2.7: Gating inhibitors of Sec61α. a) Chemical structures of four different gating inhibitors. b) Ribbon representation of Sec61α and mutations around the plug domain that confer resistance to inhibition of protein translocation. Sec61α N-terminal and C-terminal halves are depicted in blue and red, respectively, with the plug helix in yellow. Residues that, when mutated, confer resistance to inhibition of protein translocation are displayed as sticks. The structure displayed is that of the primed translocation channel (PDB code 3J7Q) [170]. c) Summary of Sec61α residues that, when mutated, confer resistance to inhibition of protein translocation, and their corresponding inhibitor(s) (adapted from [192]).

Some inhibitors, such as mycolactone and cotransin analogues, inhibit protein translocation differently based on the nature of their signal peptide or transmembrane domains [193]. Conversely, decatransin and ipomoeassin F act on translocation in a substrate-non-specific fashion [187, 190].

Most of these inhibitors directly bind to Sec61α, as evidenced by identification of Sec61α mutants resistant to inhibitors, co-immunoprecipitations and competitive photo-crosslinking experiments [183]. Intriguingly, resistance-conferring mutations in Sec61α are often shared between inhibitors and are located at the lumenal side of helices H_2 and H_3 and the plug (Figure 2.7.b,c). In addition, most inhibitors compete when binding to Sec61α, probably through partially overlapping binding sites. Therefore, gating inhibitors are likely to share a common mechanism of action, and their differential effects on protein translocation might be explained by their differences in chemical structures.

2.2.3.2 Mycolactone: a unique inhibitor

2.2.3.2.1 Roles in the virulence of *Mycobacterium ulcerans*

Infection by *Mycobacterium ulcerans* is associated with virulence in humans, fish and frogs. In humans, *Mycobacterium ulcerans* is responsible for Buruli ulcer, a neglected infectious skin disease mainly endemic in Central West Africa [194] (Figure 2.8.b). The pathogen was first found to cause tissue necrosis and a lack of inflammation beyond its site of infection [195]. This suggested that *Mycobacterium ulcerans* secreted a diffusible toxin. Later, *Mycobacterium ulcerans* was shown to be cytotoxic to eukaryotic cells and to cause tissue necrosis in animal models [196]. In addition, infection with *Mycobacterium ulcerans* was associated with immunosuppression [197, 198].

Eventually, a unique toxin secreted by *Mycobacterium ulcerans* was isolated and identified as mycolactone A/B, a mixture of *cis* and *trans* isomers [199, 200] (Figure 2.8.a). More mycolactone molecules were identified from closely related aquatic mycobacteria [201], all sharing a conserved 12-membered lactone core, extended by a northern side chain and an unsaturated acyl southern side chain. While the northern chain is conserved, the southern chain varies in length and oxidation state. Mycolactone-producing mycobacteria appear to have evolutionarily diverged from *Mycobacterium marinum* through acquisition of a megaplasmid, which encodes for a set of polyketide synthases involved in the biosynthesis of mycolactones [202]. Given that mycolactone A/B is the best characterised mycolactone molecule, it will be referred as mycolactone in the following sections.

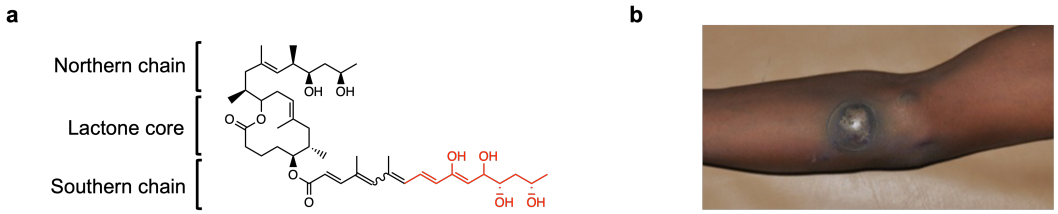


Figure 2.8: Mycolactone and Buruli ulcer. a) Chemical structure of mycolactone. Mycolactone is a mixture of *cis* and *trans* isomers (A/B), which consists of a lactone core attached to a small northern chain and a longer unsaturated acyl side southern chain. Part of the southern chain coloured in red shows where differences occur between all mycolactones, while the rest of the molecule is conserved. b) Photograph of the arm of a patient with Buruli ulcer, with initial formation of a characteristic nodule. If untreated, infection spreads and progresses to skin ulceration (picture taken from [203]).

Mycolactone is central to virulence in Buruli ulcer. First, for cell entry, the toxin is able to passively diffuse through the plasma membrane, as evidenced by tracking of fluorescently labeled mycolactone incubated with fibroblasts [204]. *In vitro*, mycolactone treatment altered the actin cytoskeleton of HeLa cells, and triggered cell death in skin and neuronal cells [205, 206]. Nerve damage induced by mycolactone suggested that the molecule is responsible for the characteristic lack of pain observed in patients. *In vivo*, intradermal injection of mycolactone into mouse ears changed the architecture of the surface epithelium of the skin [207]. Most importantly, numerous studies of various immune cells showed that mycolactone blocks secretion of cytokines and immune receptors, therefore hampering immune response at the site of infection [208–210].

2.2.3.2.2 Sec61: the molecular target of mycolactone

Mycolactone-mediated cell death and inhibition of synthesis of immune receptors and cytokines are essential to sustain development of Buruli ulcer. Interestingly, inhibition of protein secretion by mycolactone was recapitulated in a cell-free system, in which protein synthesis and translocation occured *in vitro* through incubation of ER-derived microsomes with mRNA and ribosomes purified from rabbit reticulocytes [211]. This suggested that mycolactone affects co-translational translocation at the ER. Using this system, the molecular target of mycolactone was investigated [212]. Binding of SRP to RNC, RNC delivery to SR and RNC delivery to the translocation channel were not affected by mycolactone. However, mycolactone prevented nascent chains emerging from

ribosomes to proceed through Sec61-mediated insertion into the ER. This highlighted Sec61 as the potential target of mycolactone.

Later, in genetic screens of cells resistant to mycolactone treatment, resistance-conferring mutations were identified in Sec61 [213]. In particular, mutant T cells recovered their ability to produce cytokines and HEK293 mutant cells were protected against mycolactone-mediated cytotoxicity. Together, these results linked Sec61 inhibition by mycolactone to underlying immunosuppresive effects and cell death.

Sec61 resistance mutations were located at its central pore-forming core, the Sec61α subunit. Limited proteolysis experiments, in which rough microsomes were treated with trypsin, showed a reduced protease activity on Sec61α in the presence of mycolactone, while no change was observed for other Sec61 subunits [212]. Hence, mycolactone was proposed to induce a conformational change of Sec61α that is less prone to substrate engagement. In addition, mycolactone competed with and displaced other Sec61 inhibitors [210] and, to date, mycolactone stands out as the most potent inhibitor of Sec61 activity.

Interestingly, mycolactone was less effective at preventing post-translational translocation of short secretory proteins [212]. Post-translational translocation involves accessory proteins that stabilise Sec61 in distinct conformations than those observed during co-translational translocation, most particularly during priming of the channel. In these states, mycolactone binding may be disfavoured.

In addition, mycolactone blocked integration of most type I and type II membrane proteins, but type III membrane proteins were resistant to inhibition [214]. Since the mechanisms of engagement are different between membrane proteins of different topologies, mycolactone-mediated inhibition appears to be regulated by the way nascent polypeptides engage with the translocon.

2.2.4 Project aims

The aim of this project is to structurally characterise how mycolactone, the main virulence factor of *Mycobacterium ulcerans*, binds to the mammalian Sec61 channel and inhibits co-translational translocation. This could explain how immunosuppression and cell death occurs at the site of bacterial infection. In addition, given the existence

of various gating inhibitors with similar effects and resistance mutations, a general model of inhibition could be proposed. Last, comparison of the inhibited conformation of the translocon with other states of Sec may deepen our understanding of the dynamics of translocation.

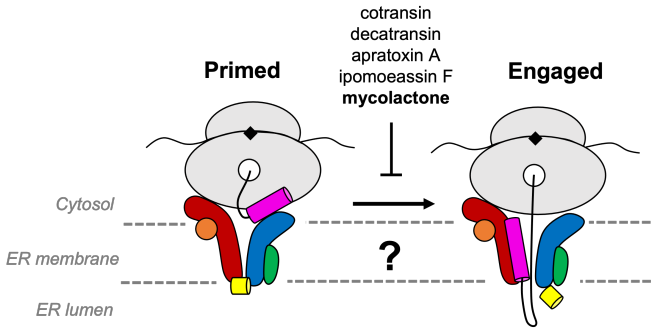


Figure 2.9: How does inhibition of protein translocation work? Schematic representation of the Sec translocon, with inhibition occuring between priming and signal peptide engagement. Notable gating inhibitors are listed.

2.3 Materials and methods

2.3.1 Synthetic mycolactone

Synthetic mycolactone A/B [215], a gift from Prof Yoshito Kishi (Harvard university), was used for all experiments described in this thesis. Mycolactone was provided in ethyl-acetate, dried down under nitrogen and resuspended in dimethylsulfoxide (DMSO) at 0.5 mg/mL. Aliquots were protected from light and stored at -80°C. Generation of mycolactone-resistant cell lines required greater amounts of mycolactone. Therefore, a mixture of mycolactone epimers at C12' was used (kind gift from Dr Nicolas Blanchard, CNRS France [216, 217]) for selection.

2.3.2 Purification of ribosome-translocon complexes (RTCs) in detergent

RTCs were prepared largely as described [170]. First, rough microsomal membranes were prepared by Dr Belinda Hall (University of Surrey) from dog pancreas as previously described [218] and stored at -80°C in 50 mM triethanolamine pH 7.5, 250 mM sucrose, 1 mM dithiothreitol (DTT). Aliquots were incubated with 1 mM $CaCl_2$ and 150 U/mL

micrococcal nuclease for 7 min at room temperature. The reaction was stopped by addition of 2 mM egtazic acid and microsomes were flash-frozen in 50 µL aliquots. Mycolactone (stock at 0.5 mg/mL in DMSO) was diluted to 20 µg/mL in 0.1 % (w/v) bovine serum albumin (BSA). In control samples, the same amount of DMSO was diluted in 0.1 % (v/v) BSA. Microsomes were thawed on ice and mycolactone was added to the membranes at a final concentration of 200 ng/mL (269.2 nM). The membranes were gently mixed, incubated on ice for 20 min then solubilised by addition of an equal volume of 100 mM HEPES pH 7.5, 800 mM potassium acetate, 20 mM magnesium acetate, 3.5 % (w/v) digitonin, 2 mM DTT for 10 min on ice. Samples were centrifuged at 20,000 *g* for 15 min at 4°C. 5 µL of supernatant was saved for liquid chromatography mass spectrometry (LC-MS) analysis and the rest was applied to a 1-mL Sephacryl-300 size-exclusion chromatography (SEC) column pre-equilibrated with ice-cold column buffer (50 mM HEPES pH 7.5, 200 mM potassium acetate, 10 mM magnesium acetate, 0.25 % (w/v) digitonin, 1 mM DTT). 2 mL of ice-cold column buffer was applied and fractions of around 100 µL were collected. Absorbance at 260 nm was calculated using a Nanodrop (ThermoFisher Scientific). Molar concentrations were estimated using the extinction coefficient of 80S ribosomes (5×10^7 $cm^{-1}M^{-1}$) [219]. Aliquots of each fraction were taken for Western blot analysis. Fractions with highest absorbance were saved for cryo-EM sample preparation. Samples were centrifuged at 20,000 *g* for 15 min at 4°C before subsequent application on cryo-EM grids.

2.3.3 Extraction of RTCs in nanodiscs and diisobutylene/maleic acid lipid particles

Extraction of intact RTCs in diisobutylene/maleic acid (DIBMA) lipid particles and nanodiscs were tested. For extraction using DIBMA, the same protocol as with detergent (see section 2.4.2) was followed with key modifications. Digitonin was substituted with 2.5 % DIBMA (gift from Tina Bolstehdt, University of Oxford) in solubilisation buffer and the column buffer was free of any detergent. Solubilisation was tested for 45 min and overnight. After elution from a 1-mL Sephacryl-300 column pre-equilibrated with

ice-cold column buffer, the fraction with the highest absorbance at 260 nm was saved and applied on cryo-EM grids. An aliquot was kept for analysis by Western blot.

For extraction in nanodiscs, the same protocol was used as with detergent, but digitonin was replaced with 2 % 3-cholamidopropyl dimethylammonio 1-propanesulfonate (CHAPS) in solubilisation buffer and 0.65 % CHAPS in column buffer. After elution from a 1-mL Sephacryl-300 column pre-equilibrated with ice-cold column buffer, fractions with high absorbance at 260 nm were pooled together. A 15:85 molar ratio of bovine brain lipids (type I, Folch fraction I, #B1502, Sigma Aldrich) and phosphatidylcholine (#850457P, Avanti Polar Lipids), previously dried under a stream of nitrogen, was resuspended at 20 mg/mL in 2 % CHAPS. RTCs, membrane scaffold protein 1D1 (MSP1D1, #M6574, Sigma Aldrich) and lipids were mixed in a 1:4:100 molar ratio and incubated for 30 min. An excess of Bio-Beads SM-2 (#1523920, Bio-Rad) was added and incubated for 1 h at 4°C. After removal of beads, the mixture was applied on nickel-nitriloacetic acid (Ni-NTA) beads, washed with ice-cold 50 mM HEPES pH 7.5, 1 mM DTT, 10 mM magnesium acetate, 200 mM potassium acetate, 10 mM imidazole and eluted with the same buffer, supplemented with 250 mM imidazole. Absorbance at 260 nm was determined for each fraction and samples were analysed by Western blot.

2.3.4 SDS-PAGE and immunoblotting

Proteins from aliquots of fractions purified by SEC on Sephacryl-300 beads or Ni-NTA chromatography were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on 4–20 % acrylamide gradient gels (Bio-Rad) together with a protein marker (#26623, ThermoFisher Scientific). Proteins were blotted onto polyvinylidene difluoride membranes, followed by blocking with 5 % milk dissolved in phosphate-buffered saline (PBS) with 0.1 % Tween-20 (PBS-T) for 1 h at room temperature. After incubation with primary antibodies diluted in PBS-T for 1 h at room temperature, membranes were washed with PBS-T and incubated secondary antibodies diluted in PBS-T for 1 h at room temperature. Membranes were incubated with Immobilion Chemiluminescent HRP substrate (#WBKLS0050, Merck Millipore) or Amersham ECL detection reagents (#RPN3004, Cytiva) and imaged on using a

printer or a Vilber Fusion FX imaging system. Primary antibodies used were: mouse monoclonal anti-Sec61α; rabbit polyclonal anti-Sec61β (gift from the Dobberstein lab) [220]; rabbit polyclonal anti-ribosomal protein S6 and mouse monoclonal anti-ribophorin 2 (#sc-166421, Santa Cruz Biotechnology). Secondary antibodies were ECL Mouse IgG, HRP-linked (#NXA931, GE Healthcare) and ECL Rabbit IgG, HRP-linked (#NA934, GE Healthcare). For detection of polyhistidine from nanodiscs, a penta-His HRP conjugate was used (#34460, Qiagen).

2.3.5 Grid preparation and data acquisition

Initial screening was performed on carbon-coated holey grids (R1.2/1.3, 200 mesh, #Q210AR1.3, Quantfoil), or in the presence of a film of ultrathin (2-3 nm) amorphous carbon. Holey grids were plasma-cleaned using a Quorum device (#Q150R ES, Quorum Technologies) at 50 mA for 60 s, while grids coated with carbon film were glow-discharged for 10 s at 15 mA. 4 μL of freshly purified RTCs in the presence and absence of mycolactone were applied on conventional holey grids at absorbances ranging from 20 to 60 at 260 nm. Sample vitrification and plunge freezing were in liquid ethane were performed using a Vitrobot (FEI), with an incubation time of 15 s, a blotting time of 3 s and a blotting force of -5 at 4°C. On carbon film, 4 μL of sample with absorbance of 10 to 15 at 260 nm were similarly plunge-frozen, with an incubation time of 15 s, a blotting time of 5 s and a blotting force of -15 at 4°C.

Later, carbon-coated holey grids (R1.2/1.3, 200 mesh, #Q210AR1.3, Quantfoil) were used for the preparation of graphene-oxide grids. The grids were plasma-cleaned using a Quorum device at 50 mA for 60 s. For each grid, 3 μL of 0.2 mg/mL graphene oxide dispersion (Graphenea) was applied on the carbon surface, manually blotted, and allowed to dry overnight. 4 μL of sample with an absorbance of 10 at 260 nm was applied on the surface of each grid and similarly vitrified by plunge-freezing with an incubation time of 15 s, a blotting time of 5 s and a blotting force of -15 at 4°C.

Grids were screened on a 200 kV Arctica transmission electron microscope (FEI) using the automated single particle acquisition software EPU (FEI) and transferred to a Titan Krios transmission electron microscope (FEI) operating at 300 kV for data

collection. Data were collected manually using EPU software, at a defocus range between -1 and -2.5 μm. All data were recorded on a K2 detector under low dose conditions with a nominal pixel size of 0.822 Å/pixel on the object scale. Small data collections of around 300 movies were initially performed to assess sample quality and behaviour on ice. For graphene oxide grids, a total of 5062 and 3833 micrographs were collected for RTCs with and without mycolactone, respectively. Each micrograph experienced a total exposure of 49 electrons per $Å^2$ fractionated into 32 frames. Foil holes containing the thinnest regions of ice and darkest edges were manually selected as they mostly contained single layer graphene oxide, as observed by Fourier transform.

2.3.6 Image processing

Original movie frames were motion corrected and combined using MotionCor2 [221]. Contrast transfer function (CTF) parameters were estimated using CTFFIND4.1 [222]. Manual inspection of graphene oxide and ice quality was performed and particles were picked automatically with SIMPLE 2.1 [223]. Coordinates were imported in RELION 3[224, 225], in which 2D classifications, 3D classifications and 3D refinements were performed. Following reference-free 2D classification, particles (76,650 for RTCs with mycolactone, 35,366 for the control dataset) were subjected to 3D classification into 10 classes after an initial round of refinement. The best 3D classes differed by absence or presence of the elongation factor eEF2, while most ribosomes contained tRNA in P/E-state or E-state. Two main classes were selected, which included ribosomes with eEF2 (32,456 particles in the presence of mycolactone, 8,607 in the control dataset) and without eEF2 (18,365 particles with mycolactone, 16,245 without). These ribosome states were compared with those previously reported by Voorhees et al. [170] using a similar protocol for purification of RTCs. The group observed 13 % active ribosomes (with A/P and P/E-site tRNAs), 65 % ribosomes without tRNA but with eEF2 and 22 % empty ribosomes [226]. Likewise, we observed a majority of idle 80S ribosomes, most of which were bound to eEF2. No ribosome was bound to a tRNA at the A-site. Our ribosome states differed by the presence of E or P/E site tRNAs in our RTCs, while Voorhees et al. observed predominantly tRNA-free ribosomes. These differences may be

due to different sources of microsomes (porcine versus canine) or subtle differences in the preparation procedure of microsomes or time taken for purification.

For each state, 3D refinement was performed with a mask around Sec61 and the 60S subunit. This was followed by particle subtraction of the ribosome core and focused 3D classification without alignment into 15 classes using a regularisation parameter (T value) of 100 and a mask around Sec61. The best classes were selected and subjected to masked 3D refinement around Sec61 and the 60S subunit after particle polishing and CTF refinement. No differences were observed in Sec61 conformation. Therefore, particles from both reconstructions were combined and refined using a mask around Sec61 and the 60S subunit. In the case of the control dataset, the same pipeline was used, without focused classification. Final reconstructions were subjected to post processing using a wide soft edge mask. This resulted in final resolutions of 2.63 Å and 2.85 Å, according to the FSC 0.143 criterion following the Relion gold-standard refinement [227]. Local resolution at the translocon was estimated using ResMap [228]. For better visualisation and interpretation of the Sec density, and due to the heterogenous resolution across the translocon, the map used for model building was lowpass-filtered to 4.0 Å.

2.3.7 Homology modelling and refinement of atomic models

The structure of Sec61 was built using the structures of the primed state (PDB code 3J7Q) and that of the signal peptide-engaged state (PDB code 3JC2) as reference models [167, 170]. Mycolactone structure was generated the electronic Ligand Builder and Optimization Workbench (eLBOW)[229] in the Phenix suite [230] using its SMILES string [231]. The model was built in COOT [232] and refined using Phenix [230] using Ramachandran and secondary structure restraints. Data are deposited for the inhibited state of Sec61 (EMDB 11064, PDB code 6Z3T) and the primed state of Sec61 (additional map in EMDB 11064).

	Ribosome:Sec61 + mycolactone	Ribosome:Sec61
Data collection and processing		
Magnification	165,000	165,000
Voltage (kV)	300	300
Electron exposure ($e^-/Å^2$)	49	49
Defocus range (μm)	-1.0 to -2.5	-1.0 to -2.5
Pixel size (Å)	0.822	0.822
Initial particle images (no.)	108,129	35,366
Final particle images (no.)	45,733	24,852
Map resolution (Å)	2.63	2.85
FSC threshold	0.143	0.143
Refinement		
Model resolution (Å)	4	
FSC threshold	0.5	
Map sharpening B factor ($Å^2$)	5.7	
Model composition		
Non-hydrogen atoms	3608	
Protein residues	441	
Ligands	MYC: 1	
Average B factors ($Å^2$)		
Protein	65.98	
Ligand	71.85	
RMSD		
Bond lengths (Å)	0.006	
Bond angles (°)	1.047	
Validation		
MolProbity score	1.79	
Clashscore	5.40	
Poor rotamers (%)	0.53	
Ramachandran plot		
Favoured (%)	91.61	
Allowed (%)	8.39	
Disallowed (%)	0.00	

Table 2.1: Refinement and model statistics.

2.3.8 *In vitro* translation/translocation assays

In vitro translocation assays were performed by Dr Belinda Hall (University of Surrey), as detailed in [233] (publication in Appendix B).

2.3.9 Liquid chromatography mass spectrometry

LC-MS experiments were performed by Dr Belinda Hall (University of Surrey) to quantify mycolactone binding to microsomes and RTC fractions from purification on

Sephacryl-300 beads. Details are found in [233] (publication in Appendix B).

2.3.10 Forward genetic screen, viability assays and purification of microsomes from mycolactone-resistant cell lines

Forward genetic screens, viability assays and purification of microsomes from mycolactone-resistant cell lines were carried out by Dr Belinda Hall (University of Surrey), as detailed in [233] (publication in Appendix B).

2.3.11 Molecular dynamics simulations

Molecular dynamics simulations were performed to assess stability of mycolactone conformations in the translocon with and without introduction of resistance mutations. The model of mycolactone-inhibited Sec61 derived from cryo-EM experiments was used as a starting point. First, two poses of mycolactone were fitted in the electron density. The original conformation was that of mycolactone with its northern chain toward the inside of the translocon and its southern chain toward the putative lipid bilayer. In the reversed conformation both chains were swapped. In each case, the lactone core remained at the lateral gate. The $\Delta 4',5'$-*trans* isomer, was selected for simulations since it is the presumptive active form of mycolactone [234].

Simulations were carried out by Dr Afroditi Maria Zaki (University of Oxford), as thoroughly detailed in [233] (publication in Appendix B). The mycolactone-bound translocon was embedded in a pre-equilibrated phosphatidylcholine lipid bilayer using the General Amber force field for mycolactone [235], the Amber 99SB-ILDN force field for Sec61 [236] and the Slipids force field for POPC [237]. Position restraints were applied to maintain Sec61 in its conformation in the absence of ribosome and three repeats of 100 ns-long molecular dynamics simulations were performed in each case. To investigate the effect of mutations on mycolactone stability and dynamics at the lateral gate, three mutant systems were generated (S71F, G80W and S82F) in PyMOL [238]. Molecular dynamics runs were performed using the same parameters and procedure as described above, except for position restraints, which were applied only to Sec61γ subunit and the transmembrane helices H_6-H_{10} of Sec61α. This allows

the rest of Sec61α to adapt to conformational changes caused by mutations. For each mutant and wild-type Sec61, three 100 ns-long runs were performed.

2.4 Results

2.4.1 Preparation of mammalian RTCs for cryo-EM

Several cryo-EM studies have provided structural insights of RTCs in different states or associated to different binding partners [159, 160, 162, 163, 166, 167, 169, 170, 176]. Therefore, a similar strategy was adopted to capture and compare the states of mycolactone-inhibited and non-inhibited RTCs.

Canine microsomes, which contain a heterogenous set of ER vesicles isolated from dog pancreas, were used since these are rich in RTCs and therefore optimal for extraction in quantities suitable for cryo-EM studies. First, microsomal membranes prepared by Dr Belinda Hall (University of Surrey) were incubated in the presence or absence of saturating amount of mycolactone, long enough to fully block translocation [239]. Mycolactone saturating concentration (125 ng/mL, 168.2 nM) was determined by Dr Belinda Hall (University of Surrey) based on *in vitro* translation and translocation assay of yeast prepro-α-factor by canine microsomes. Then, microsomal membranes were solubilised and RTCs were extracted with the mild non-ionic detergent digitonin, followed by small scale SEC. High absorbance values at 260 nm suggested fractions enriched in nucleic acids, indicative of the presence of ribosomes (Figure 2.10.a). 40S, 60S ribosomal components and the three subunits of Sec61 were detected by Western blot and LC-MS confirmed the presence or absence of mycolactone in the inhibited and control samples, respectively (Figure 2.10.b-d).

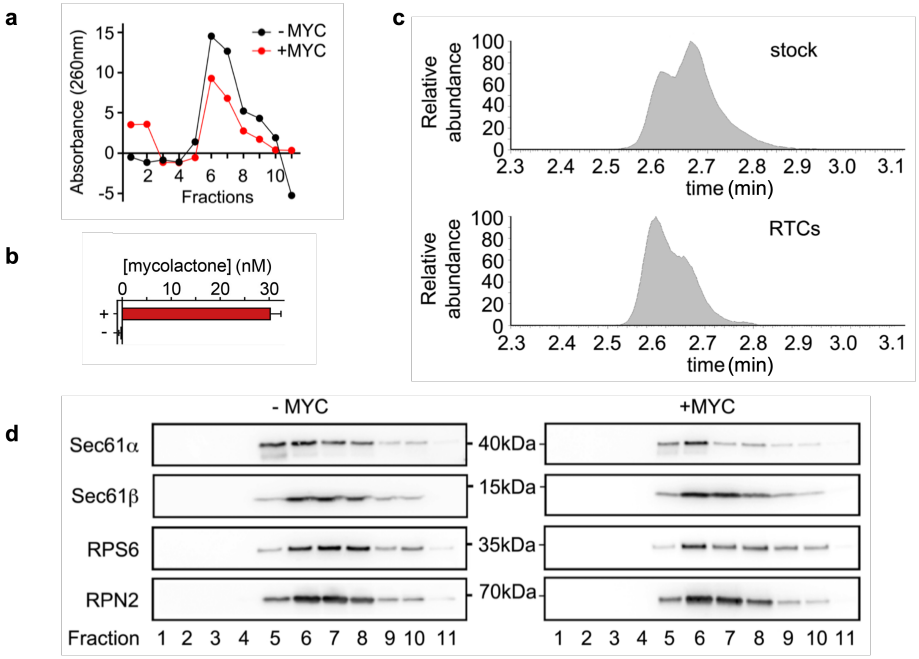


Figure 2.10: Purification of RTCs from canine microsomes in the presence and absence of mycolactone. a) Plot of absorbance at 260 nm for each fraction collected after application on Sephacryl-300 beads. b) Mycolactone concentrations in inhibited (+) and non-inhibited (-) RTCs, estimated by LC-MS. c) LC-MS identification of mycolactone in inhibited RTCs. The extracted ion chromatograms depicted are for the initial stock of mycolactone diluted in 0.1 % BSA before addition to canine microsomes (stock) and for inhibited RTCs pooled fractions after elution by SEC. The region extracted is for the $[M+Na]^+$ (m/z 765.4721 - 765.5103) ion eluting at 2.38 - 2.98 min. Two peaks are observed and may correspond to both *cis* and *trans* isomers of mycolactone. The first fraction most likely corresponds to $\Delta 4',5'$-*trans* isomer, which may be enriched during sample purification. d) Detection by Western blot of RTC components for each SEC fraction, in the presence and absence of mycolactone (+MYC and -MYC, respectively). Sec61α and Sec61β detection indicates presence of translocation channels. Signal for ribosomal protein S6 (RPS6), a component of the 40S ribosomal subunit, indicates presence of ribosomes. Ribophorin II (RPN2), a component of the ER-associated oligosaccharyltransferase complex that interacts with ribosomes during protein translocation, is also detected.

2.4.2 Optimisation of grid support

Freshly purified RTCs were applied on different types of cryo-EM grids (Figure 2.11.a,b). On conventional holey carbon film, ribosome particles were observed at a very low density and clustered at the edges of the foil holes. However, increasing concentration only led to aggregation of particles there.

Because particles favourably adsorbed on carbon support, samples were applied on grids coated with a 2-3 nm layer of amorphous carbon [167, 170]. Drastic improvements of particle density and behaviour were observed, and preliminary data collection and 2D classification highlighted the presence of only intact 80S associated with

translocons. However, strong preferential particle orientations were evidenced by 2D and 3D classifications (Figure 2.11.c). Notably, the refined 3D reconstruction showed streaky electron density artefacts (Figure 2.11.d).

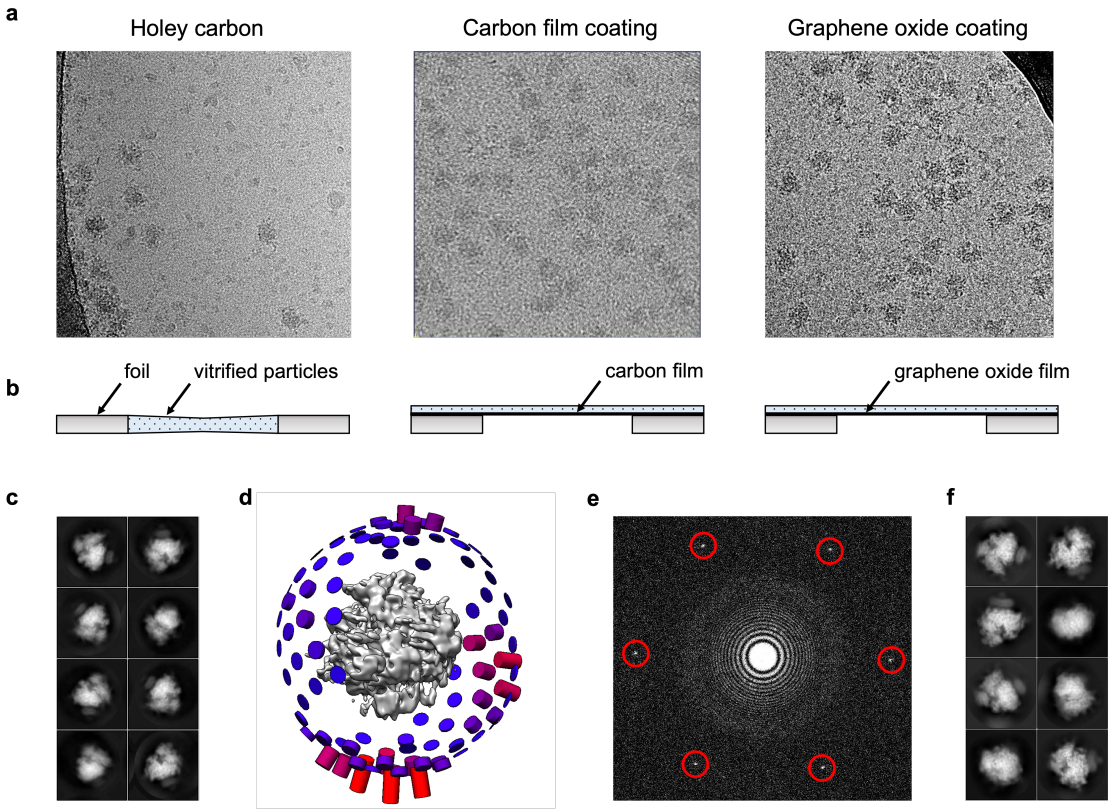


Figure 2.11: Optimisation of grid support for cryo-EM data collection. a) Micrograph exposures for RTCs applied on a holey carbon grid, a grid coated with an amorphous layer of carbon and a grid coated with graphene oxide. RTCs were applied at an absorbance at 260 nm of 60, 15 and 10, respectively. On holey carbon, clusters of ribosomes are adsorbed on carbon at the edges of foilholes. Exposures were recorded with a magnification of 165,000. b) Schematic representation of a grid foilhole in each case, matching with exposures above in a). c) 2D classes of RTCs on carbon film, with a box size of 600×600 pixels. d) 3D refined reconstruction of RTCs on carbon film and its angular distribution. e) Fourier transform of a micrograph collected for an acquisition area with graphene oxide. A hexagonal reciprocal lattice (6 dots circled in red) around Thon rings is characteristic of single layer graphene oxide. f) 2D classes of RTCs on graphene oxide film, with a box size of 528×528 pixels.

Alternatively, grids covered with graphene oxide were prepared (Figure 2.11.e). These have not only been successfully used to reduce preferential orientation and denaturation at the air-water interface but also to increase particle density by adsorption, therefore requiring less sample [240–242]. In addition, the background noise contribution from graphene oxide film is negligible compared to that of an amorphous carbon film [243]. RTCs applied on graphene oxide grids displayed a satisfactory particle density and a set

of multiple views, as evidenced by 2D classification (Figure 2.11.f). Graphene oxide was therefore selected as the support of choice to collect data of inhibited and control RTCs.

2.4.3 Cryo-EM processing of RTCs in the presence and absence of mycolactone

For both cryo-EM datasets in the presence and absence of mycolactone, detailed 2D classes of only intact RTCs were obtained, indicating that each 80S ribosome was bound to a translocon. These particles were refined and further subjected to 3D classification with fine angular sampling, to distinguish between different states. 3D classes only differed by the presence or absence of the elongation factor eEF2, and most ribosomes contained tRNA at P/E- or E-sites (Figures 2.12.a, 2.13.a). No active ribosomes were found.

Classes were refined separately, using a mask around the 60S:translocon for better assignment of angles, ignoring heterogeneity brought by the 40S subunit. Subtraction of the ribosome contribution followed by focused classification without alignment allowed us to select classes with well-defined density for transmembrane helices of the translocation channel and increase its signal contribution at a given threshold. 3D refinements of the resulting particles yielded translocons in identical conformations, regardless of the ribosome state, allowing us to merge data with and without eEF2.

Processing of the control dataset yielded a map with an overall resolution of 2.80 Å (Figure 2.12.b). This map was used to dock the previously solved structure of the primed state of Sec61 [170]. As expected, transmembrane helices were found to fit in the density with no major shift (Figure 2.12.c).

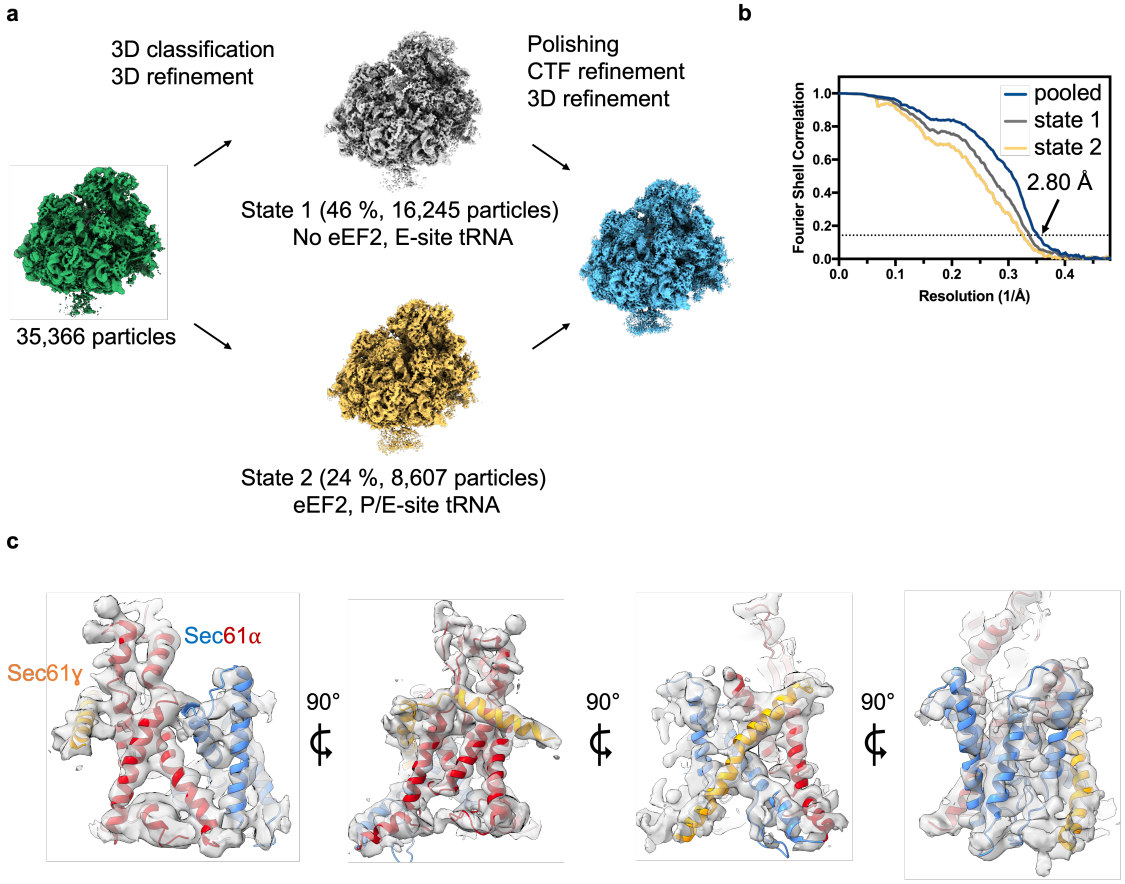


Figure 2.12: Cryo-EM processing of RTCs in the absence of mycolactone. a) Pipeline for cryo-EM processing of RTCs in the absence of mycolactone. After reference-free 2D classification and 3D refinement, 3D classification with low angular sampling yielded two major states, in which tRNAs either occupied E-site of ribosomes unbound to eEF2 or P/E site of ribosomes attached to eEF2 (states 1 and 2, respectively). These were refined and combined to yield the final 3D reconstruction. b) FSC curves obtained after 3D refinement for each state and combined particles. c) Ribbon representation of the primed conformation of Sec61 (PDB code 3J7Q) fitted in the corresponding electron density. The map was filtered to 5 Å for clarity.

In the presence of mycolactone, the map obtained reached an overall resolution of 2.63 Å, with local resolution around Sec61 ranging from 2.4 to 8.7 Å (Figure 2.13.b,c). Notably, the rigid C-terminal half of Sec61α (H_6-H_{10}) directly in contact with the ribosome was better resolved than the more dynamic N-terminal half (H_1-H_5).

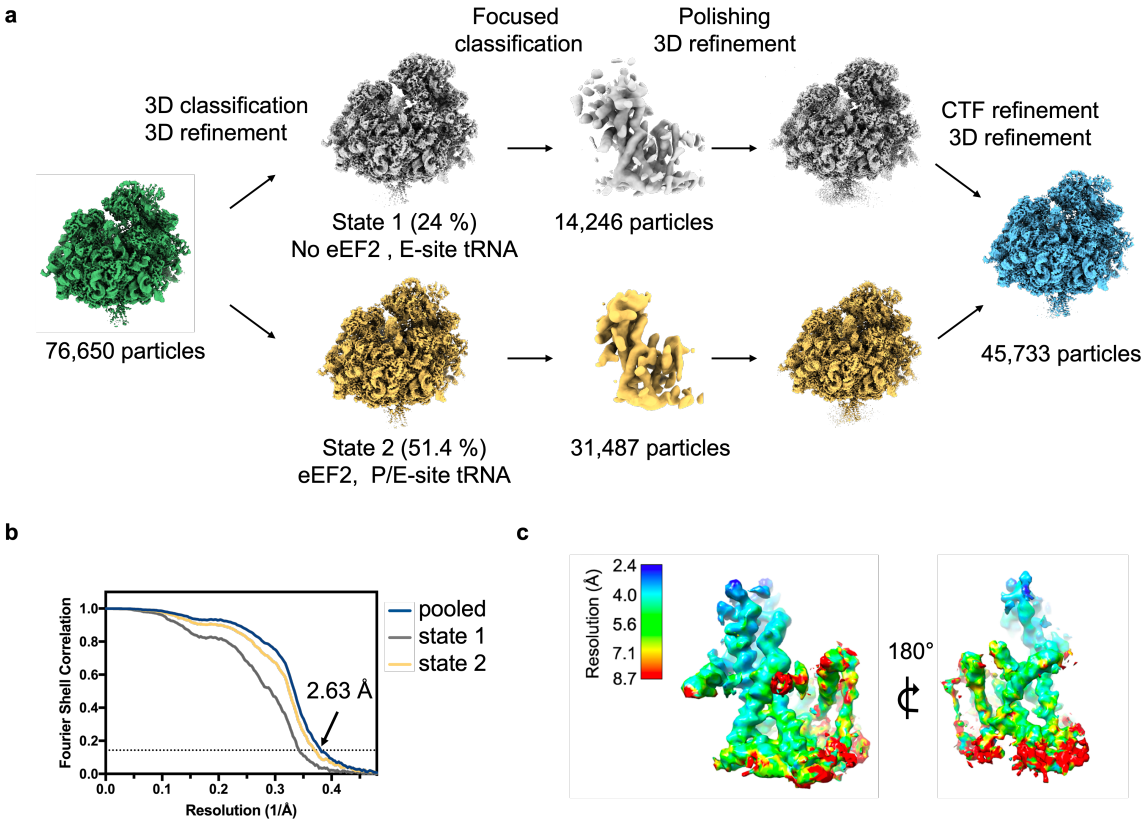


Figure 2.13: Cryo-EM processing of RTCs in the presence of mycolactone. a) Pipeline for cryo-EM processing of RTCs in the presence of mycolactone. After reference-free 2D classification and 3D refinement, 3D classification with low angular sampling yielded two major states, in which tRNAs either occupied E-site of ribosomes unbound to eEF2 or P/E site of ribosomes attached to eEF2 (states 1 and 2, respectively). These were refined after focused classification on Sec61 and combined to yield the final reconstruction used to build the inhibited state of Sec61. b) FSC curve obtained after 3D refinement for each state and combined particles. c) Local resolution estimates calculated for Sec61 density. The C-terminal half of Sec61, which is associated to the 60S subunit, is more rigid and better resolved than the rest of the protein complex.

2.4.4 Structure of the inhibited state of Sec61

Using a composite of both models of signal-peptide engaged and primed Sec61, the structure of the inhibited channel was built (Figure 2.14). Sec61α and Sec61γ subunits were built with confidence, while a polyalanine helix was modelled for Sec61β.

The structure of Sec61α C-terminal helices H_6-H_{10} is similar to previously reported structures of Sec61α. Notably, unique features are observed for the rest of the channel. For instance, at the cytoplasmic side of the lateral gate, a wide opening occurs between H_2 and H_8. Intriguingly, although the lateral gate is open, the plug helix seals the pore. This suggests that the channel is ready to accept a substrate from the cytosolic side but

is not active. Indeed, in this conformation, no polypeptide can pass through the pore.

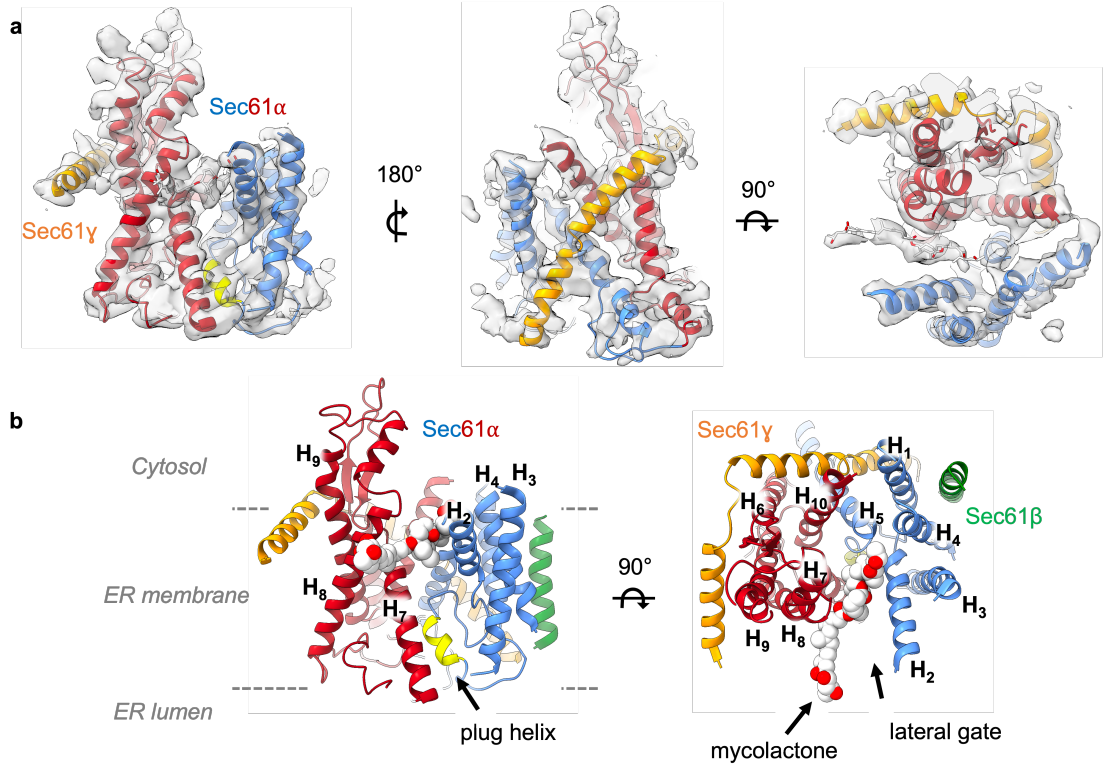


Figure 2.14: Structural features of the inhibited state of Sec61. a) Ribbon representation of the mycolactone-inhibited conformation of Sec61 fitted in the corresponding electron density. The map was filtered to 5 Å for clarity. Mycolactone was fitted in the elongated density observed at the cytoplasmic side of the lateral gate. b) Ribbon representation of the structure of mycolactone-inhibited Sec61. Main features include opening of the lateral gate defined by H_2-H_3 and H_7-H_8, sealing of the pore by the plug helix (P) and wedging of the lateral seam at the cytosolic side by mycolactone (shown as spheres).

In the 3D reconstruction, no density for a signal peptide could be observed. However an elongated density was relatively well defined at the cytosolic side of the lateral gate, between H_2 and H_8 (Figure 2.14.a). Since this conformation was induced by the addition of inhibitor, and because no other extra density could be found anywhere else around the translocon, this density was assigned to mycolactone.

2.4.5 Differences between the inhibited Sec translocon and other states

The inhibited Sec61 was compared with different conformations taken by the channel during translocation (Figure 2.15).

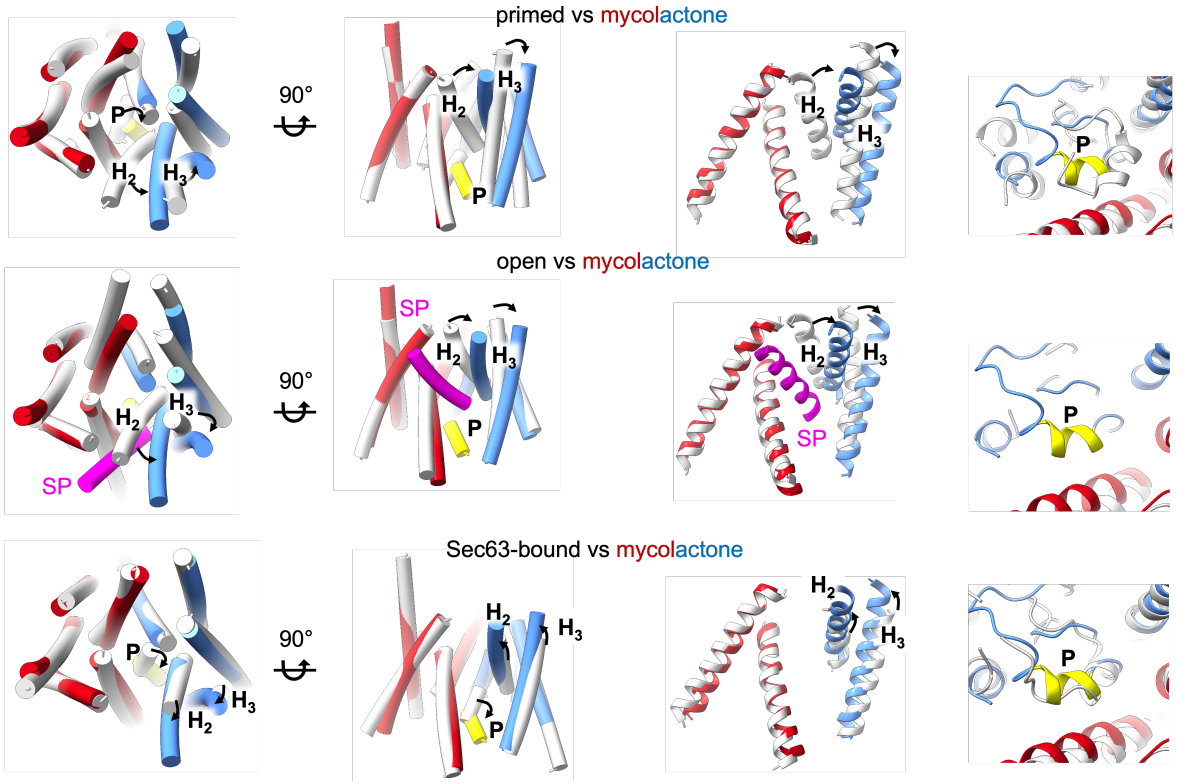


Figure 2.15: Comparison of Sec61 inhibited conformation with other states adopted during translocation. The structure of the inhibited state of Sec61 (in colours) was aligned with structures of the primed (PDB code 3J7Q) and open states (PDB 3JC2) observed during co-translational translocation and the Sec63-bound state (PDB code 6ND1, all in light grey). The C-terminal halves of the Sec61α were aligned in ChimeraX. Transmembrane helices are depicted as cylinders for clarity (left half) and viewed from the cytoplasmic side (left) and the putative lipid bilayer facing the lateral gate (right). Helices of the lateral gate and plug helix are depicted as ribbons (right half). The signal peptide of the signal-peptide engaged state is depicted in magenta. Movements are highlighted for the plug helix (P, in yellow) and the pair of dynamic helices H_2-H_3 at the lateral gate.

In the absence of mycolactone and before engagement with a signal peptide, Sec61α primed by ribosome binding displays a loosened lateral gate formed by H_2-H_3 and H_7-H_8, and the plug helix seals the pore [226].

In the presence of mycolactone, the N-terminal helices H_2 and H_3 moved laterally by approximately 9 Å at their cytoplasmic side, while helix H_4 is tilted. A much wider opening of the translocon lateral gate open is observed at its cytoplasmic side. Though the plug helix occupies the pore and prevents passage of translocating polypeptides, it is shifted by approximately 7 Å compared to that of a primed translocon.

Furthermore, the movement observed at the lateral gate is wider and occurs more horizontally than that of the active translocon, for which the N-terminal helices undergo

rotation as a rigid body to give space for intercalation of a signal peptide [167].

Remarkably, the inhibited conformation is most similar to that of Sec61 bound to Sec63, with comparable positions of plug helices and similar movements of H_2-H_3 at the lateral gate. In this state, the channel is primed by Sec63 binding to Sec61α C-terminal cytosolic loops and N-terminal transmembrane helices [160, 163]. Further recruitment of Sec62 activates the channel through rotation of the N-terminal half of helices and displacement of the plug helix to receive translocating polypeptides [159].

2.4.6 Mode of binding of mycolactone to Sec61

In the model of inhibited Sec61, assignment of the position of mycolactone at the cytoplasmic side of the lateral gate was not unambiguous, for several reasons. First, the resolution in this region of the map was not sufficient to place side chains confidently. Second, mycolactone exists as a mix of two isomers, and although the Δ4′,5′-*trans* isomer is thought to be the active form of mycolactone [244], both molecules may occupy Sec61 binding site. Furthermore mycolactone is flexible and elongated. Therefore, conformation flexibility will blur the density during cryo-EM processing. For these reasons, collecting more data is unlikely to allow building of mycolactone side chains. Last, in reconstructions of Sec63-bound translocons, a density was observed at the lateral gate, albeit poorly resolved when compared to that of mycolactone, and attributed to digitonin or glyco-diosgenin, a substitute for digitonin [160, 163] (Figure 2.16.a).

Extraction of RTCs was attempted in lipids rather than detergents. First, insertion into lipid nanodiscs was investigated. Since digitonin is hard to remove due to its low critical micellar concentration (CMC), it was substituted with CHAPS, a non-denaturing zwitterionic detergent with a high CMC that can be used to extract intact RTCs [245]. In addition, MSP1D1 was selected since the bacterial homologue SecYEG was successfully inserted into this scaffold [171]. Detergent removal and insertion into a mixture of bovine brain lipids and POPC yielded some translocons inserted into nanodiscs after elution during immobilised metal-ion affinity chromatography (IMAC), as evidenced by immunoblotting (Figure 2.16.b). However, most ribosomes were collected in the flow-through and the amount of material obtained after elution was negligible, although three

times more microsomal membranes than with digitonin were used, therefore preventing further purification of RTCs from empty nanodiscs by SEC and cryo-EM studies.

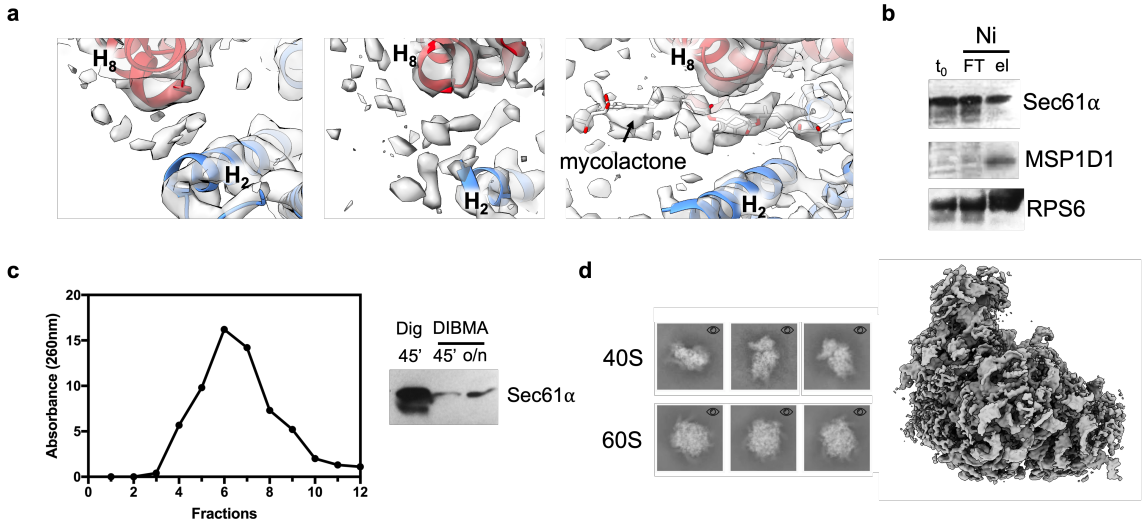


Figure 2.16: Insertion trials of RTCs into lipids. a) Ribbon and stick representations of structures of the lateral gate in Sec63-bound translocons (left and middle, PDB codes 6N3Q and 6ND1, respectively) compared with mycolactone-associated Sec61 (right), all fitted into their corresponding electron density. At a threshold in which helical features are observed, glyco-diosgenin (left) and digitonin (middle) contribute to noisy densities, whereas density is better defined in the mycolactone-bound state. b) Western blot detection of Sec61α, MSP1D1 and 40S ribosomal protein S6 (RPS6) during insertion into nanodiscs. Samples were collected after SEC (t_0), and in the flow-through (FT) and elution (el) during IMAC (Ni). The yield obtained was too low to carry on with further purification steps and grid preparation (absorbance at 260 nm of 0.15 in 60 µL). c) DIBMA solubilisation and extraction of RTCs. Absorbance at 260 nm is plotted for each fraction collected after application on Sephacryl-300 beads (left). Sec61α was detected in the void peak using digitonin (Dig) as a positive control, and DIBMA over 45 min and overnight incubation with microsomes. Extraction efficiency with DIBMA was much lower. d) Cryo-EM analysis of DIBMA-solubilised ribosomal particles applied on graphene oxide grids. 2D classes (left) are typical of 40S and 60S subunits projections. The 3D reconstruction of the 60S subunit (right) shows the absence of 40S and translocon.

Alternatively, insertion of RTCs into DIBMA lipid particles was investigated. DIBMA is a synthetic polymer of diisobutylene and maleic acid, which can solubilise proteins and withstand higher concentrations of divalent cations than polymers of styrene maleic acid. Since Sec61 was successfully inserted into lipid bilayers of bovine brain lipids and POPC, the same mixture was used for extraction with DIBMA. Notably, extraction efficiency was decreased by eight fold when using DIBMA rather than digitonin. Using five times more microsomes, ribosome particles and Sec61α co-eluted in the void peak obtained by SEC (Figure 2.16.c). However, a small cryo-EM data collection of this sample showed the absence of intact RTCs (Figure 2.16.d). This indicates that RTCs

collapsed during sample freezing, or that DIBMA lipid particles were unstable in the buffer and aggregated, possibly because of magnesium.

In parallel, molecular dynamics simulations were performed by Dr Afroditi Maria Zaki (University of Oxford) to assess stability of mycolactone in the translocon. The $\Delta 4',5'$-*trans* isomer, which is the presumptive active form of mycolactone [244], was modelled.

At the lateral gate, two possible orientations of mycolactone could be fitted in the cryo-EM density. In the original orientation, the lactone core was positioned at the cytosolic side of the lateral gate, the northern tail pointed toward the centre of Sec61 cytosolic side and the southern chain tail extended toward the lipid bilayer. In the second pose, the location of southern and northern tails were reversed, while the lactone core remained at the cytosolic side of the lateral gate. Molecular dynamics simulations were run over 100 ns for each orientation to assess the stability of mycolactone. In both cases, mycolactone remained associated with Sec61, but the molecule in the original orientation was more stable, especially at its southern chain (Figure 2.17.a). Moreover, it showed a better fit in the electron density than the molecule in the reversed orientation (Figure 2.17.b). This indicates that the original orientation is the most likely binding mode of mycolactone to Sec61.

In this orientation, interactions with Sec61 are mainly formed by the lactone core and the northern chain, while the southern chain is flexible and disordered in the lipid bilayer. Notably, during molecular dynamics simulations, mycolactone stayed in close contact with mostly hydrophobic residues (V85, T86, L89, I289, P290, L293, Q294, L297, W379 and V382) of H_2, H_8 and the core of the translocon over time (Figure 2.17.c-e). This suggests that opening of the lateral gate exposes hydrophobic residues, which form a binding pocket for mycolactone.

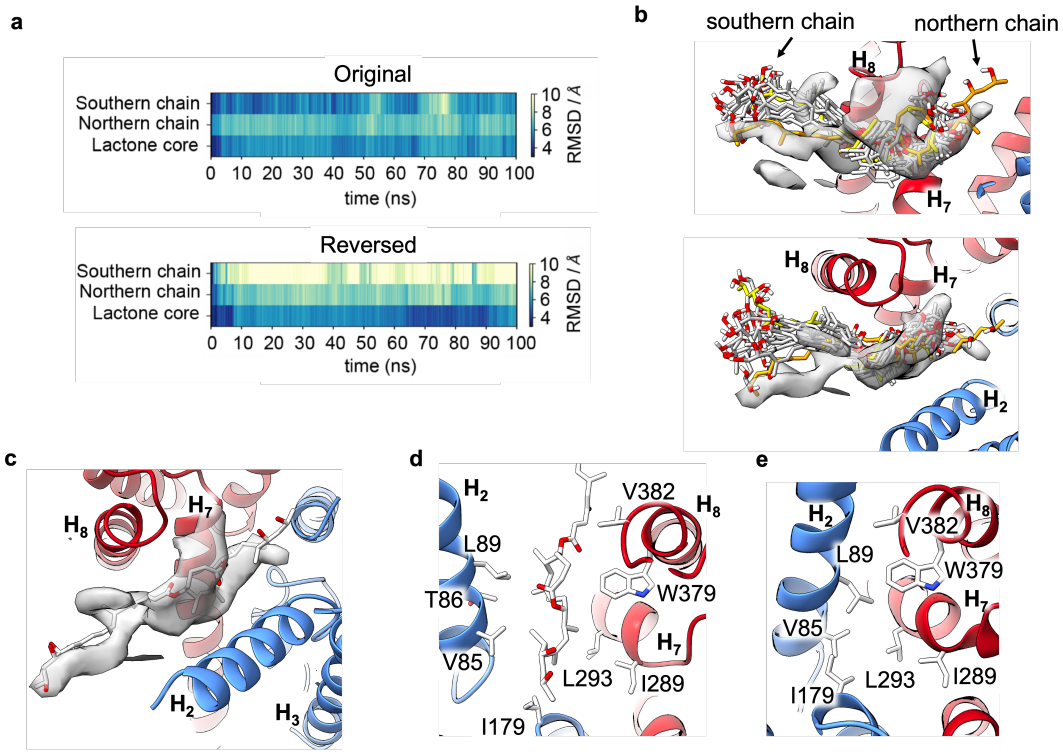


Figure 2.17: Mycolactone binding site. a) Molecular dynamics simulations of mycolactone stability in original and reversed orientations. RMSDs for the lactone core, the southern chain and the northern chain are averaged from three independent runs and shown over 100 ns. In the reversed orientation, the southern chain is unstable. b) Ribbon and sticks representation of mycolactone movement over 100 ns in the original pose, observed from cytosolic and translocon sides (bottom and top, respectively). The initial pose of mycolactone is shown in orange, eight intermediates at 10 ns intervals are shown in white, and the final position after 100 ns is depicted in yellow. c) Ribbon and sticks representation of the binding site of mycolactone, with its corresponding electron density. d) Ribbon and sticks representation of mycolactone binding site, with hydrophobic residues displayed in the pocket. c) Ribbon and sticks representation of the primed state of the translocon (PDB code 3J7Q), viewed from the same direction as in b), with the same hydrophobic residues displayed.

2.4.7 Mutational studies of Sec61 resistance to mycolactone

The inhibited structure of the translocon and the position of mycolactone were used to study how mutations modulate inhibition of protein translocation.

To address this, a forward genetic screen was performed by Dr Belinda Hall (University of Surrey) and ten substitutions in six codons (D60G, D60V, R66G, R66K, S71F, G80W, S82F, S82Y, Q127K, Q127R) triggered resistance to mycolactone, in addition to the previously reported mutation of T86 [190, 233]. Among these, mutations of D60, R66, S71, G80, S82 and T86 cause resistance to some other inhibitors [192]. Most mutated residues are located in the plug helix or at the lumenal side of Sec61α in close proximity to

the plug helix (Figure 2.18.a-c). Intriguingly, these are far from the mycolactone binding site, which is located at the cytoplasmic side of Sec61α. Nonetheless, no electron density was observed close to the plug which could be attributed to mycolactone. Therefore, these mutations do not affect binding to mycolactone directly.

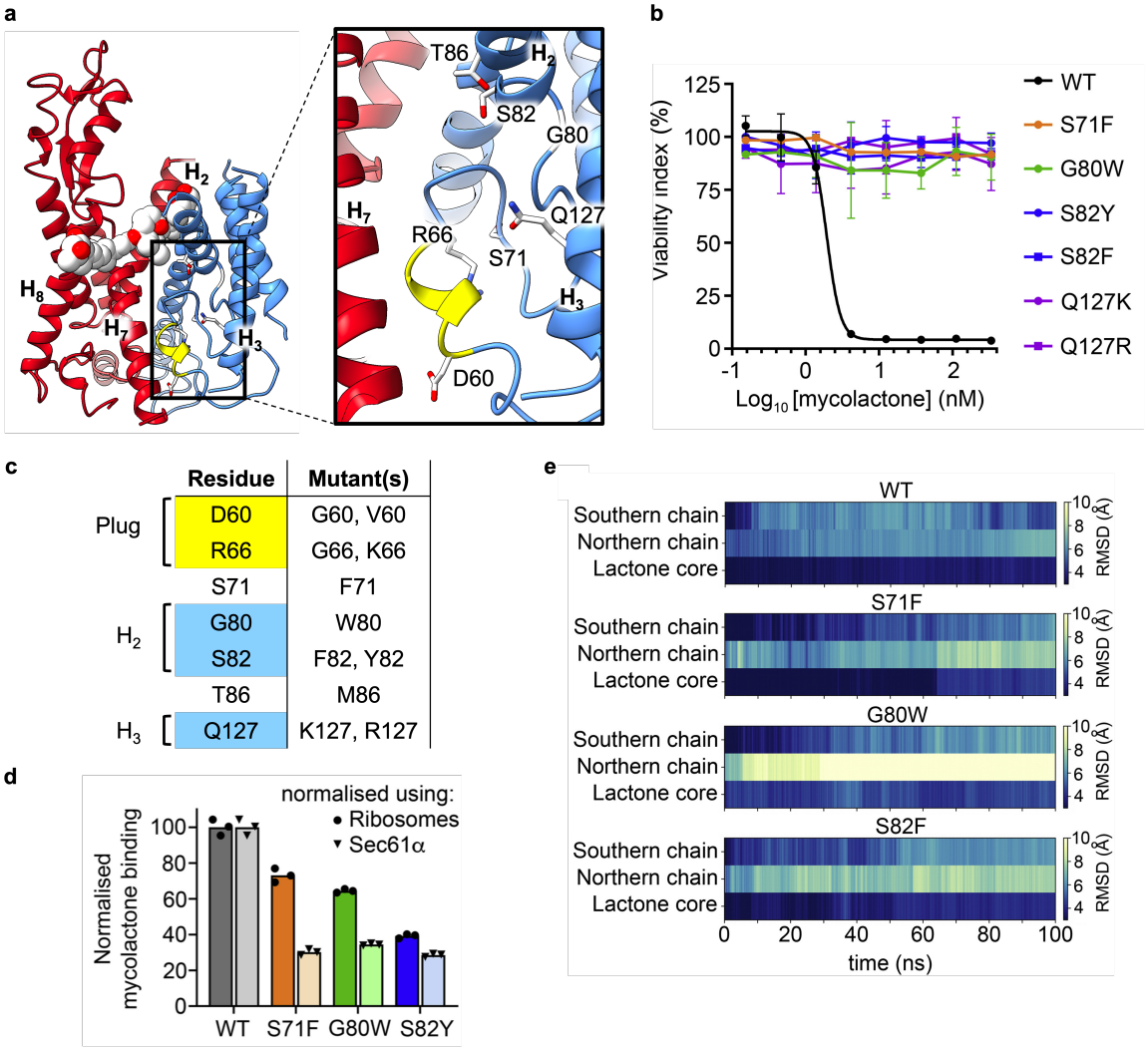


Figure 2.18: Sec61α resistance mutations to mycolactone and their effect on mycolactone binding. a) Ribbon, stick and sphere representation of the inhibited state of Sec61, with focus on resistance mutations around the plug helix (in yellow). b) Viability assay showing sensitivity of mutants to increasing concentrations of mycolactone. c) Table summary of mycolactone resistance mutations and their corresponding localisation on Sec61α. d) Relative mycolactone binding to RTC by LC/MS. Decrease in binding to mycolactone is observed for each mutant, although wild type Sec61α remains endogenously expressed in mutant heterozygous cell lines. e) Molecular dynamics simulation of mycolactone stability upon introduction of three resistance point mutations. RMSD from the starting position is shown at each nanosecond for the southern chain, northern chain and lactone core. Most variations are observed at the northern chain.

Intriguingly, mutations at residues of the mycolactone binding pocket were not

identified in the screens. This may be because these residues are critical for Sec61 function. Indeed, some of these residues are part of the functionally important pore ring (V85 and I179), hydrophobic patch (L89 and I179) and polar cluster (T86) [158]. For instance, V85D was reported to prevent co-translational translocation [246]. Such mutations leading to non-viable translocons would therefore not appear in forward genetic screens.

The effect of resistance mutations on mycolactone binding to Sec61 was further studied by LC/MS performed by Dr Belinda Hall (University of Surrey). Microsomes from wild type TRex-293 cells and stable cell lines overexpressing Sec61α mutated at S71, G80 and S82 were isolated. For each cell line, mycolactone-treated RTCs were extracted and the presence of inhibitor was quantified by LC/MS. A 70-80 % reduction in mycolactone association was estimated for the three mutants, when compared with binding to the wild-type translocon (Figure 2.18.d). Given that Sec61α remained endogenously expressed in these cell lines, these mutations appear to drastically decrease binding to mycolactone.

The indirect effect of S71F, G80W and S82F on mycolactone binding to Sec61 was also studied by molecular dynamics. The system was set so that the N-terminal half (H_1-H_5) of the translocation channel was allowed to move freely while the C-terminal half (H_6-H_{10}), which adopted equivalent conformation in the primed translocon, was fixed. For each mutant, mycolactone was found to be less stable at its binding site than for the wild type translocon, with major deviations observed for the northern chain (Figure 2.18.e).

Altogether, these results suggest that mutations do not affect binding to mycolactone directly. Instead, resistance mutations may affect the dynamics of Sec61α to disfavour of the formation of the mycolactone binding site. Furthermore, we looked at the dynamics around the lactone core binding site and assessed how mutations affect the opening of the lateral gate there. Distance variations between pairs of residues of H_2-H_8, V85-W379, L89-V382 and L89-W379, which contact mycolactone, were calculated over time [233]. In these simulations, introduction of S82F mutation increased the distance for each pair of residues, therefore favouring opening of the lateral gate. Conversely, G80W and S71F tended to favour closing of the lateral gate. This suggests that mutations into bulky residues favour opening or closing at the lateral gate, therefore affecting the conformational dynamics of the channel to prevent binding of mycolactone.

2.5 Discussion and perspectives

2.5.1 An unexpected mode of binding

The structural study of mycolactone-mediated inhibition of protein translocation reveals that the molecule associates with Sec61α in a unique conformation when compared to previously solved structures of translocons.

In the inhibited state, mycolactone wedges open the cytosolic side of the lateral gate of Sec61α, far from its resistance mutations. The location of mycolactone is supported by cryo-EM complemented with molecular dynamics simulations and shows that interactions with Sec61 are mostly hydrophobic and mediated by the lactone core and the northern chain.

Although the lateral gate is prised partly open by mycolactone, the plug helix remains in position, occluding the central pore of the channel. Similarly, translocons associated with the allosteric activator Sec63 also display a partly open lateral gate and a pore sealed by the plug helix. In this intermediate state, the translocon is primed for later engagement with weak signal peptides. Therefore, mycolactone also appears to act as an allosteric inhibitor, binding away from its resistance mutations and trapping Sec61 in an intermediate state that is partly activated.

2.5.2 Mycolactone prevents engagement of signal peptides

The location and effect of mycolactone on protein translocation has implications for our understanding of signal peptide engagement. Indeed, the path taken by signal peptides during translocation remains partly understood.

In trapped intermediates of active Sec61, the signal peptide is intercalated at the lateral gate. There, its position overlaps with that of mycolactone in our structure. In fact, mycolactone occupies the hydrophobic patch that potentially attracts signal peptides to their initial site of engagement. This prevents their intercalation in the open lateral gate, and subsequent opening of the plug helix.

2.5.3 Implications for the dynamics of translocation

It has been hypothesised that gating inhibitors either bind to closed idle translocons or stabilise an early state of channel opening [187]. Our data support this second hypothesis and provide new insights into the dynamics of translocation.

In ribosome-bound primed translocons, the space provided by cracking of helices at the cytosolic side of the lateral gate is not sufficient to accommodate mycolactone in its binding pocket. Furthermore, the inhibitor alone is unlikely to trigger opening of these helices. In addition, given that positions of signal peptides and translocating polypeptides in engaged translocons overlap with that of mycolactone in the inhibited structure of Sec61, mycolactone is unlikely to bind to active Sec61. In support of this, when mycolactone was added after initiation of translocation, it failed to prevent crosslinking of a substrate to Sec61α, suggesting that it is unable to compete with signal peptides for binding to active translocons. Instead, crosslinking of a substrate was only prevented when mycolactone was added before initiation of translocation, suggesting that mycolactone binds to Sec61 before engagement with substrates and prevent their interaction with Sec61α [212].

Rather than binding to closed or active translocons, mycolactone must associate with Sec61 when its lateral gate provides space to accommodate the molecule. To achieve this, ribosome-bound Sec61 must be in a dynamic equilibrium, with the mycolactone binding site transiently available due to dynamic opening and closing of the lateral gate.

This equilibrium is advantageous for Sec61, since lateral gate opening facilitates recruitment and gating of signal peptides and allosteric effectors such as Sec62/Sec63, while the plug helix seals the pore to prevent ion leak. However, opening at the lateral gate also provides space for inhibitors such as mycolactone, which inserts there, trapping translocons in the inhibited conformation, and lying in the path of incoming signal peptides, which cannot then engage Sec61 (Figure 2.19).

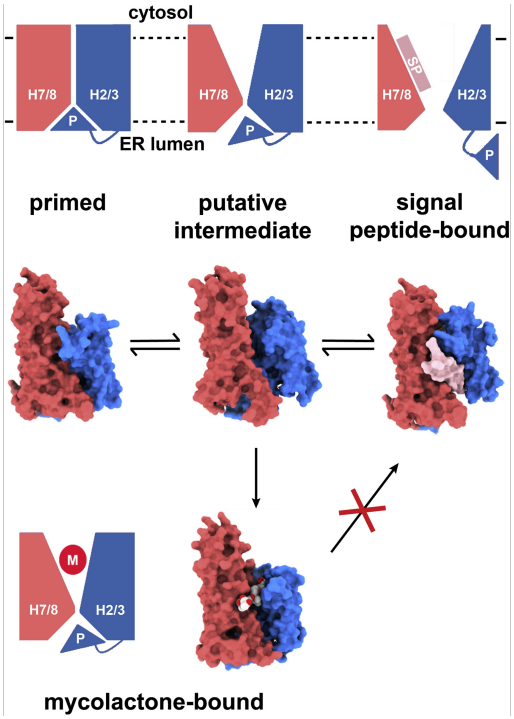


Figure 2.19: Model of inhibition of co-translational translocation. Sec61α is in a dynamic equilibrium between the primed state (PDB code 3J7Q) and a putative intermediate state, in which the lateral gate is more open, as in Sec63-bound translocons (PDB code 6ND1). There, space is provided for intercalation of and stabilisation by allosteric effectors such as mycolactone. Since mycolactone lies in the way of incoming signal peptides, these cannot access their binding site (PDB code 3JC2). Surface and schematic representations of Sec61α are depicted. In schematic representations, helices of the lateral gate (H2/3 and H7/8), the plug helix (P), the signal peptide (SP) and mycolactone (M) are highlighted. For clarity, only Sec61α is depicted.

2.5.4 Mode of action of other inhibitors

The inhibited structure provides the molecular basis for the mode of action of mycolactone and explains how it prevents protein secretion and allows the formation of Buruli ulcer during *Mycobacterium ulcerans* infection.

Other chemically diverse inhibitors of protein translocation compete with mycolactone when binding to Sec61. The inhibited conformation may also accommodate these molecules at the lateral gate. Moreover, most inhibitors share Sec61 resistance-conferring mutations, which appear to disfavour formation of the inhibited state. In fact, a similar allosteric mechanism was proposed for *prl* phenotypes, a set of Sec61 mutants which allow secretion of proteins with defective signal peptides. Remarkably, most resistance mutations identified here are shared with *prl* phenotypes (D60, R66, S71, G80, Q127)[187, 247–249].

Overall, these data suggest a mechanism of allosteric inhibition shared between inhibitors. Understanding of this process is important, since inhibitors have been shown to display anti-viral, anti-inflammatory and anti-tumoral activities. Therefore, our structure lays a foundation for future development of modulators of Sec61α gating.

2.5.5 Perspectives

Following this work, a number of directions could be followed. For instance, elucidation of structures of Sec61 mutants resistant to inhibition would be of interest to decipher how they disfavour the inhibited conformation. At the same time, such structures should explain how *prl* phenotypes can translocate defective signal peptides. In addition, similar structural studies of other inhibitors bound to Sec61α would be valuable to assess if their binding site is conserved, in support of our model of inhibition. Last, the differential effects of inhibitors on post-translational and co-translational translocations could be investigated structurally.

References

[112] G. E. Palade. "A small particulate component of the cytoplasm." In: *The Journal of biophysical and biochemical cytology* 1.1 (1955), pp. 59–68.

[113] F. H. Crick. "On protein synthesis." In: *Symposia of the Society for Experimental Biology* 12 (1958), pp. 138–163.

[114] T. M. Schmeing and V. Ramakrishnan. "What recent ribosome structures have revealed about the mechanism of translation". In: *Nature* 461.7268 (2009), pp. 1234–1242.

[115] G. Z. Yusupova et al. "The path of messenger RNA through the ribosome". In: *Cell* 106.2 (2001), pp. 233–241.

[116] M. Selmer et al. "Structure of the 70S ribosome complexed with mRNA and tRNA". In: *Science* 313.5795 (2006), pp. 1935–1942.

[117] G. Yusupova et al. "Structural basis for messenger RNA movement on the ribosome". In: *Nature* 444.7117 (2006), pp. 391–394.

[118] M. B. Hoagland et al. "A soluble ribonucleic acid intermediate in protein synthesis." In: *The Journal of biological chemistry* 231.1 (1958), pp. 241–257.

[119] D. Grunberger, I. B. Weinstein, and K. B. Jacobson. "Codon recognition by enzymatically mischarged valine transfer ribonucleic acid". In: *Science* 166.3913 (1969), pp. 1635–1637.

[120] F. Chapeville et al. "On the role of soluble ribonucleic acid in coding for amino acids." In: *Proceedings of the National Academy of Sciences of the United States of America* 48.6 (1962), pp. 1086–1092.

[121] M. Simonović and T. A. Steitz. "A structural view on the mechanism of the ribosome-catalyzed peptide bond formation". In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1789.9-10 (2009), pp. 612–623.

[122] N. R. Voss et al. "The Geometry of the Ribosomal Polypeptide Exit Tunnel". In: *Journal of Molecular Biology* 360.4 (2006), pp. 893–906.

[123] S. Jenni and N. Ban. "The chemistry of protein synthesis and voyage through the ribosomal tunnel". In: *Current Opinion in Structural Biology* 13.2 (2003), pp. 212–219.

[124] L. D. Cabrita, C. M. Dobson, and J. Christodoulou. "Protein folding on the ribosome". In: *Current Opinion in Structural Biology* 20.1 (2010), pp. 33–45.

[125] C. Maracci and M. V. Rodnina. "Review: Translational GTPases". In: *Biopolymers* 105.8 (2016), pp. 463–475.

[126] D. D. Leipe et al. "Classification and evolution of P-loop GTPases and related ATPases". In: *Journal of Molecular Biology* 317.1 (2002), pp. 41–72.

[127] A. P. Schuller and R. Green. "Roadblocks and resolutions in eukaryotic translation". In: *Nature Reviews Molecular Cell Biology* 19.8 (2018), pp. 526–541.

[128] R. M. Voorhees and V. Ramakrishnan. "Structural basis of the translational elongation cycle". In: *Annual Review of Biochemistry* 82 (2013), pp. 203–236.

[129] S. Melnikov et al. "One core, two shells: Bacterial and eukaryotic ribosomes". In: *Nature Structural and Molecular Biology* 19.6 (2012), pp. 560–567.

[130] C. U. T. Hellen. "Translation termination and ribosome recycling in eukaryotes". In: *Cold Spring Harbor Perspectives in Biology* 10.10 (2018).

[131] G. Blobel and B. Dobberstein. "Transfer of proteins across membranes: I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma". In: *Journal of Cell Biology* 67.3 (1975), pp. 835–851.

[132] S. M. Simon and G. Blobel. "A protein-conducting channel in the endoplasmic reticulum". In: *Cell* 65.3 (1991), pp. 371–380.

[133] R. S. Hegde and H. L. Ploegh. "Quality and quantity control at the endoplasmic reticulum". In: *Current Opinion in Cell Biology* 22.4 (2010), pp. 437–446.

[134] Z. Sun and J. L. Brodsky. "Protein quality control in the secretory pathway". In: *Journal of Cell Biology* 218.10 (2019), pp. 3171–3187.

[135] C. Hammond and A. Helenius. "Quality control in the secretory pathway". In: *Current Opinion in Cell Biology* 7.4 (1995), pp. 523–529.

[136] S. S. Vembar and J. L. Brodsky. "One step at a time: Endoplasmic reticulum-associated degradation". In: *Nature Reviews Molecular Cell Biology* 9.12 (2008), pp. 944–957.

[137] L. Käll, A. Krogh, and E. L. L. Sonnhammer. "A combined transmembrane topology and signal peptide prediction method". In: *Journal of Molecular Biology* 338.5 (2004), pp. 1027–1036.

[138] R. J. Keenan et al. "The signal recognition particle". In: *Annual Review of Biochemistry* 70 (2001), pp. 755–775.

[139] N. Halic and R. Beckmann. "The signal recognition particle and its interactions during protein targeting". In: *Current Opinion in Structural Biology* 15.1 SPEC. ISS. (2005), pp. 116–125.

[140] D. Akopian et al. "Signal recognition particle: An essential protein-targeting machine". In: *Annual Review of Biochemistry* 82 (2013), pp. 693–721.

[141] N. Johnson, K. Powis, and S. High. "Post-translational translocation into the endoplasmic reticulum". In: *Biochimica et Biophysica Acta - Molecular Cell Research* 1833.11 (2013), pp. 2403–2409.

[142] R. S. Hegde and R. J. Keenan. "Tail-anchored membrane protein insertion into the endoplasmic reticulum". In: *Nature Reviews Molecular Cell Biology* 12.12 (2011), pp. 787–798.

[143] S. D. Emr, S. Hanley-Way, and T. J. Silhavy. "Suppressor mutations that restore export of a protein with a defective signal sequence". In: *Cell* 23.1 (1981), pp. 79–88.

[144] A. I. Derman et al. "A signal sequence is not required for protein export in prlA mutants of Escherichia coli". In: *EMBO Journal* 12.3 (1993), pp. 879–888.

[145] A. M. Flower, R. C. Doebele, and T. J. Silhavy. "PrlA and PrlG suppressors reduce the requirement for signal sequence recognition". In: *Journal of Bacteriology* 176.18 (1994), pp. 5607–5614.

[146] M. K. Olsen, E. L. Rosey, and C. S. C. Tomich. "Isolation and analysis of novel mutants of Escherichia coli prlA (secY)". In: *Journal of Bacteriology* 175.21 (1993), pp. 7092–7096.

[147] R. S. Osborne and T. J. Silhavy. "PrlA suppressor mutations cluster in regions corresponding to three distinct topological domains". In: *EMBO Journal* 12.9 (1993), pp. 3391–3398.

[148] V. A. Bankaitis and P. J. Bassford. "Proper interaction between at least two components is required for efficient export of proteins to the Escherichia coli cell envelope". In: *Journal of Bacteriology* 161.1 (1985), pp. 169–178.

[149] J. D. Fikes and P. J. Bassford. "Novel secA alleles improve export of maltose-binding protein synthesized with a defective signal peptide." In: *Journal of bacteriology* 171.1 (1989), pp. 402–409.

[150] J. L. Huie and T. J. Silhavy. "Suppression of signal sequence defects and azide resistance in Escherichia coli commonly result from the same mutations in secA". In: *Journal of Bacteriology* 177.12 (1995), pp. 3518–3526.

[151] J. Stader, L. J. Gansheroff, and T. J. Silhavy. "New suppressors of signal-sequence mutations, prlG, are linked tightly to the secE gene of Escherichia coli." In: *Genes & development* 3.7 (1989), pp. 1045–1052.

[152] K. S. Crowley et al. "Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore". In: *Cell* 78.3 (1994), pp. 461–471.

[153] D. Görlich and T. A. Rapoport. "Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane". In: *Cell* 75.4 (1993), pp. 615–630.

[154] L. Brundage et al. "The purified E. coli integral membrane protein SecY E is sufficient for reconstitution of SecA-dependent precursor protein translocation". In: *Cell* 62.4 (1990), pp. 649–657.

[155] J. Akimaru et al. "Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from Escherichia coli". In: *Proceedings of the National Academy of Sciences of the United States of America* 88.15 (1991), pp. 6545–6549.

[156] M. Watanabe, C. V. Nicchitta, and G. Blobel. "Reconstitution of protein translocation from detergent-solubilized Escherichia coli inverted vesicles: PrlA protein-deficient vesicles efficiently translocate precursor proteins". In: *Proceedings of the National Academy of Sciences of the United States of America* 87.5 (1990), pp. 1960–1964.

[157] M. Watanabe and G. Blobel. "SecA protein is required for translocation of a model precursor protein into inverted vesicles of Escherichia coli plasma membrane". In: *Proceedings of the National Academy of Sciences of the United States of America* 90.19 (1993), pp. 9011–9015.

[158] B. Van den Berg et al. "X-ray structure of a protein-conducting channel". In: *Nature* 427 (2004), pp. 36–44.

[159] S. Itskanov et al. "Stepwise gating of the Sec61 protein-conducting channel by Sec63 and Sec62". In: *Nature Structural and Molecular Biology* 28.2 (2021), pp. 162–172.

[160] X. Wu, C. Cabanos, and T. A. Rapoport. "Structure of the post-translational protein translocation machinery of the ER membrane". In: *Nature* 566.7742 (2019), pp. 136–139.

[161] M. Gogala et al. "Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion". In: *Nature* 506.7486 (2014), pp. 107–110.

[162] S. Pfeffer et al. "Structure of the native Sec61 protein-conducting channel". In: *Nature Communications* 6 (2015).

[163] S. Itskanov and E. Park. "Structure of the posttranslational Sec protein-translocation channel complex from yeast". In: *Science* 363.6422 (2019), pp. 84–87.

[164] T. Becker et al. "Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome". In: *Science* 326.5958 (2009), pp. 1369–1373.

[165] J. F. Ménétret et al. "Single Copies of Sec61 and TRAP Associate with a Nontranslating Mammalian Ribosome". In: *Structure* 16.7 (2008), pp. 1126–1137.

[166] T. Weng et al. "Architecture of the active post-translational Sec translocon". In: *The EMBO Journal* 40.3 (2021), e105643.

[167] R. M. Voorhees and R. S. Hegde. "Structure of the Sec61 channel opened by a signal sequence". In: *Science* 351.6268 (2016), pp. 88–89.

[168] W. Li et al. "The Plug Domain of the SecY Protein Stabilizes the Closed State of the Translocation Channel and Maintains a Membrane Seal". In: *Molecular Cell* 26.4 (2007), pp. 511–521.

[169] K. Braunger et al. "Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum". In: *Science* 360.6385 (2018), pp. 215–219.

[170] R. M. Voorhees et al. "Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution". In: *Cell* 157.7 (2014), pp. 1632–1643.

[171] P. F. Egea and R. M. Stroud. "Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.40 (2010), pp. 17182–17187.

[172] E. Park et al. "Structure of the SecY channel during initiation of protein translocation". In: *Nature* 506.7486 (2014), pp. 102–106.

[173] V. Shanmuganathan et al. "Structural and mutational analysis of the ribosome-arresting human XBP1u". In: *eLife* 8 (2019).

[174] T. Tsukazaki et al. "Conformational transition of Sec machinery inferred from bacterial SecYE structures". In: *Nature* 455.7215 (2008), pp. 988–991.

[175] L. Li et al. "Crystal structure of a substrate-engaged SecY protein-translocation channel". In: *Nature* 531.7594 (2016), pp. 395–399.

[176] L. Kater et al. "Partially inserted nascent chain unzips the lateral gate of the Sec translocon". In: *EMBO reports* 20.10 (2019).

[177] Y. Tanaka et al. "Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State". In: *Cell Reports* 13.8 (2015), pp. 1561–1568.

[178] C. Ma et al. "Structure of the substrate-engaged SecA-SecY protein translocation machine". In: *Nature Communications* 10.1 (2019).

[179] L. Bischoff et al. “Visualization of a polytopic membrane protein during SecY-mediated membrane insertion”. In: *Nature Communications* 5 (2014).

[180] J. Zimmer, Y. Nam, and T. A. Rapoport. “Structure of a complex of the ATPase SecA and the protein-translocation channel”. In: *Nature* 455.7215 (2008), pp. 936–943.

[181] W. Li et al. “The Plug Domain of the SecY Protein Stabilizes the Closed State of the Translocation Channel and Maintains a Membrane Seal”. In: *Molecular Cell* 26.4 (2007), pp. 511–521.

[182] T. A. Rapoport, L. Li, and E. Park. “Structural and mechanistic insights into protein translocation”. In: *Annual Review of Cell and Developmental Biology* 33 (2017), pp. 369–390.

[183] V. Van Puyenbroeck and K. Vermeire. “Inhibitors of protein translocation across membranes of the secretory pathway: novel antimicrobial and anticancer agents”. In: *Cellular and Molecular Life Sciences* 75.9 (2018), pp. 1541–1558.

[184] E. Pauwels, R. Schülein, and K. Vermeire. “Inhibitors of the Sec61 complex and novel high throughput screening strategies to target the protein translocation pathway”. In: *International Journal of Molecular Sciences* 22.21 (2021), p. 22.

[185] K. U. Kalies and K. Römisch. “Inhibitors of Protein Translocation Across the ER Membrane”. In: *Traffic* 16.10 (2015), pp. 1027–1038.

[186] H. Luesch and V. O. Paavilainen. “Natural products as modulators of eukaryotic protein secretion”. In: *Natural Product Reports* 37.5 (2020), pp. 717–736.

[187] T. Junne et al. “Decatransin, a new natural product inhibiting protein translocation at the Sec61/SecYEG translocon”. In: *Journal of Cell Science* 128.6 (2015), pp. 1217–1229.

[188] A. O. Paatero et al. “Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel”. In: *Cell Chemical Biology* 23.5 (2016), pp. 561–566.

[189] D. Tranter et al. “Coibamide A Targets Sec61 to Prevent Biogenesis of Secretory and Membrane Proteins”. In: *ACS Chemical Biology* 15.8 (2020), pp. 2125–2136.

[190] G. Zong et al. “Ipomoeassin F Binds Sec61α to Inhibit Protein Translocation”. In: *Journal of the American Chemical Society* 141.21 (2019), pp. 8450–8461.

[191] M. Pollack. “The role of exotoxin A in pseudomonas disease and immunity.” In: *Reviews of infectious diseases* 5 Suppl 5 (1983).

[192] H. Luesch and V. O. Paavilainen. “Natural products as modulators of eukaryotic protein secretion”. In: *Natural Product Reports* 37.5 (2020), pp. 717–736.

[193] J. L. Garrison et al. “A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum”. In: *Nature* 436.7048 (2005), pp. 285–289.

[194] R. R. Yotsu et al. “Buruli Ulcer: a Review of the Current Knowledge”. In: *Current Tropical Medicine Reports* 5.4 (2018), pp. 247–256.

[195] D. H. Connor and H. F. Lunn. “Mycobacterium ulcerans infection (with comments on pathogenesis).” In: *International journal of Leprosy* 33.3 (1965).

[196] J. K. Read et al. “Cytotoxic activity of Mycobacterium ulcerans”. In: *Infection and Immunity* 9.6 (1974), pp. 1114–1122.

[197] A. A. Pahlevan et al. “The inhibitory action of Mycobacterium ulcerans soluble factor on monocyte/T cell cytokine production and NF-kappa B function.” In: *Journal of immunology (Baltimore, Md. : 1950)* 163.7 (1999), pp. 3928–35.

[198] M. Pimsler et al. “Immunosuppressive properties of the soluble toxin from mycobacterium ulcerans”. In: *Journal of Infectious Diseases* 157.3 (1988), pp. 577–580.

[199] K. M. George et al. “Mycolactone: A polyketide toxin from mycobacterium ulcerans required for virulence”. In: *Science* 283.5403 (1999), pp. 854–857.

[200] K. M. George et al. “Partial purification and characterization of biological effects of a lipid toxin produced by Mycobacterium ulcerans”. In: *Infection and Immunity* 66.2 (1998), pp. 587–593.

[201] H. Hong et al. “Mycolactones: Immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria”. In: *Natural Product Reports* 25.3 (2008), pp. 447–454.

[202] M. J. Yip et al. “Evolution of Mycobacterium ulcerans and other mycolactone-producing mycobacteria from a common Mycobacterium marinum progenitor”. In: *Journal of Bacteriology* 189.5 (2007), pp. 2021–2029.

[203] R. R. Yotsu et al. “Buruli Ulcer: a Review of the Current Knowledge”. In: *Current Tropical Medicine Reports* 5.4 (2018), pp. 247–256.

[204] D. S. Snyder and P. L. C. Small. “Uptake and cellular actions of mycolactone, a virulence determinant for Mycobacterium ulcerans”. In: *Microbial Pathogenesis* 34.2 (2003), pp. 91–101.

[205] J. En et al. “Mycolactone is responsible for the painlessness of Mycobacterium ulcerans infection (Buruli ulcer) in a murine study”. In: *Infection and Immunity* 76.5 (2008), pp. 2002–2007.

[206] K. M. George et al. “A Mycobacterium ulcerans toxin, mycolactone, causes apoptosis in guinea pig ulcers and tissue culture cells”. In: *Infection and Immunity* 68.2 (2000), pp. 877–883.

[207] L. Guenin-Macé et al. “Mycolactone activation of Wiskott-Aldrich syndrome proteins underpins Buruli ulcer formation”. In: *Journal of Clinical Investigation* 123.4 (2013), pp. 1501–1512.

[208] S. Adusumilli et al. “Mycobacterium ulcerans toxic macrolide, mycolactone modulates the host immune response and cellular location of M. ulcerans in vitro and in vivo”. In: *Cellular Microbiology* 7.9 (2005), pp. 1295–1304.

[209] R. E. Simmonds et al. “Mycolactone Inhibits Monocyte Cytokine Production by a Posttranscriptional Mechanism”. In: *The Journal of Immunology* 182.4 (2009), pp. 2194–2202.

[210] L. Baron et al. “Mycolactone subverts immunity by selectively blocking the Sec61 translocon”. In: *Journal of Experimental Medicine* 213.13 (2016), pp. 2885–2896.

[211] B. S. Hall et al. “The Pathogenic Mechanism of the Mycobacterium ulcerans Virulence Factor, Mycolactone, Depends on Blockade of Protein Translocation into the ER”. In: *PLoS Pathogens* 10.4 (2014), e1004061.

[212] M. McKenna, R. E. Simmonds, and S. High. “Mechanistic insights into the inhibition of Sec61-dependent co- and post-translational translocation by mycolactone”. In: *Journal of Cell Science* 129.7 (2016), pp. 1404–1415.

[213] J. Ogbechi et al. "Inhibition of Sec61-dependent translocation by mycolactone uncouples the integrated stress response from ER stress, driving cytotoxicity via translational activation of ATF4". In: *Cell Death and Disease* 9.3 (2018), p. 397.

[214] M. McKenna, R. E. Simmonds, and S. High. "Mycolactone reveals the substrate-driven complexity of Sec61-dependent transmembrane protein biogenesis". In: *Journal of Cell Science* 130.7 (2017), pp. 1307–1320.

[215] F. Song et al. "Total synthesis of mycolactones A and B". In: *Tetrahedron* 63.26 (2007), pp. 5739–5753.

[216] S. Saint-Auret et al. "Modular total syntheses of mycolactone A/B and its [2H]-isotopologue". In: *Organic and Biomolecular Chemistry* 15.36 (2017), pp. 7518–7522.

[217] S. Saint-Auret et al. "Synthetic strategies towards mycolactone A/B, an exotoxin secreted by: Mycobacterium ulcerans". In: *Organic Chemistry Frontiers* 4.12 (2017), pp. 2380–2386.

[218] P. Walter and G. Blobel. "Preparation of microsomal membranes for cotranslational protein translocation". In: *Methods in Enzymology* 96.C (1983), pp. 84–93.

[219] N. B. Matasova et al. "Isolation of ribosomal subunits containing intact rRNA from human placenta: Estimation of functional activity of 80S ribosomes". In: *Analytical Biochemistry* 198.2 (1991), pp. 219–223.

[220] A. Kelkar and B. Dobberstein. "Sec61β, a subunit of the Sec61 protein translocation channel at the Endoplasmic Reticulum, is involved in the transport of Gurken to the plasma membrane". In: *BMC Cell Biology* 10 (2009), p. 11.

[221] S. Q. Zheng et al. "MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy". In: *Nature Methods* 14.4 (2017), pp. 331–332.

[222] A. Rohou and N. Grigorieff. "CTFFIND4: Fast and accurate defocus estimation from electron micrographs". In: *Journal of Structural Biology* 192.2 (2015), pp. 216–221.

[223] D. Elmlund and H. Elmlund. "SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles". In: *Journal of Structural Biology* 180.3 (2012), pp. 420–427.

[224] J. Zivanov et al. "New tools for automated high-resolution cryo-EM structure determination in RELION-3". In: *eLife* 7 (2018).

[225] S. H. W. Scheres. "RELION: Implementation of a Bayesian approach to cryo-EM structure determination". In: *Journal of Structural Biology* 180.3 (2012), pp. 519–530.

[226] R. M. Voorhees et al. "Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution". In: *Cell* 157.7 (2014), pp. 1632–1643.

[227] S. H. W. Scheres and S. Chen. "Prevention of overfitting in cryo-EM structure determination". In: *Nature Methods* 9.9 (2012), pp. 853–854.

[228] L. Swint-Kruse and C. S. Brown. "Resmap: Automated representation of macromolecular interfaces as two-dimensional networks". In: *Bioinformatics* 21.15 (2005), pp. 3327–3328.

[229] N. W. Moriarty, R. W. Grosse-Kunstleve, and P. D. Adams. "Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation". In: *Acta Crystallographica Section D: Biological Crystallography* 65.10 (2009), pp. 1074–1080.

[230] P. D. Adams et al. “PHENIX: A comprehensive Python-based system for macromolecular structure solution”. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (2010), pp. 213–221.

[231] A. T. Balaban, O. Mekenyan, and D. Bonchev. “Unique description of chemical structures based on hierarchically ordered extended connectivities (HOC procedures). I. Algorithms for finding graph orbits and canonical numbering of atoms”. In: *Journal of Computational Chemistry* 6.6 (1985), pp. 538–551.

[232] P. Emsley and K. Cowtan. “Coot: Model-building tools for molecular graphics”. In: *Acta Crystallographica Section D: Biological Crystallography* 60.12 I (2004), pp. 2126–2132.

[233] S. F. Gérard et al. “Structure of the Inhibited State of the Sec Translocon”. In: *Molecular Cell* 79.3 (2020), 406–415.e7.

[234] M. Gehringer et al. “Configurationally Stabilized Analogs of M. ulcerans Exotoxins Mycolactones A and B Reveal the Importance of Side Chain Geometry for Mycolactone Virulence”. In: *Organic Letters* 21.15 (2019), pp. 5853–5857.

[235] J. Wang et al. “Development and testing of a general Amber force field”. In: *Journal of Computational Chemistry* 25.9 (2004), pp. 1157–1174.

[236] K. Lindorff-Larsen et al. “Improved side-chain torsion potentials for the Amber ff99SB protein force field”. In: *Proteins: Structure, Function and Bioinformatics* 78.8 (2010), pp. 1950–1958.

[237] J. P. M. Jämbeck and A. P. Lyubartsev. “An extension and further validation of an all-atomistic force field for biological membranes”. In: *Journal of Chemical Theory and Computation* 8.8 (2012), pp. 2938–2948.

[238] W. L. DeLano. “Pymol: An open-source molecular graphics tool”. In: *CCP4 Newsletter on protein crystallography* 40 (2002), pp. 82–92.

[239] B. S. Hall et al. “The Pathogenic Mechanism of the Mycobacterium ulcerans Virulence Factor, Mycolactone, Depends on Blockade of Protein Translocation into the ER”. In: *PLoS Pathogens* 10.4 (2014).

[240] E. Palovcak et al. “A simple and robust procedure for preparing graphene-oxide cryo-EM grids”. In: *Journal of Structural Biology* 204.1 (2018), pp. 80–84.

[241] F. Wang et al. “General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM”. In: *Proceedings of the National Academy of Sciences of the United States of America* 117.39 (2020), pp. 24269–24273.

[242] A. Kumar, N. Sengupta, and S. Dutta. “Simplified approach for preparing graphene oxide tem grids for stained and vitrified biomolecules”. In: *Nanomaterials* 11.3 (2021), pp. 1–22.

[243] E. D’Imprima et al. “Protein denaturation at the air-water interface and how to prevent it”. In: *eLife* 8 (2019).

[244] M. Gehringer et al. “Configurationally Stabilized Analogs of M. ulcerans Exotoxins Mycolactones A and B Reveal the Importance of Side Chain Geometry for Mycolactone Virulence”. In: *Organic Letters* 21.15 (2019), pp. 5853–5857.

[245] E. Chevet et al. “Phosphorylation by CK2 and MAPK enhances calnexin association with ribosomes”. In: *EMBO Journal* 18.13 (1999), pp. 3655–3666.

[246] D. Schubert et al. “Plasma cell deficiency in human subjects with heterozygous mutations in Sec61 translocon alpha 1 subunit (SEC61A1)”. In: *Journal of Allergy and Clinical Immunology* 141.4 (2018), pp. 1427–1438.

[247] T. Junne et al. “Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology”. In: *Journal of Biological Chemistry* 282.45 (2007), pp. 33201–33209.

[248] S. F. Trueman, E. C. Mandon, and R. Gilmore. “A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function”. In: *Journal of Cell Biology* 199.6 (2012), pp. 907–918.

[249] M. A. Smith et al. “Modeling the effects of prl mutations on the Escherichia coli SecY complex”. In: *Journal of Bacteriology* 187.18 (2005), pp. 6454–6465.

3

Structural studies of the 260-kDa antigen of *Entamoeba histolytica*

3.1 Summary

Entamoeba histolytica is an invasive parasite which ingests intestinal cells and causes tissue damage of the epithelium. To achieve this, amoebae stick to the surface of the epithelium through a 260-kDa heterodimer expressed on their surface, which consists of a heavy subunit and a light subunit. The protein is antigenic and it is the main vaccine candidate against human intestinal amoebiasis. This third chapter describes the purification and structure determination of the 260-kDa antigen by cryo-EM. The structure shows that heterodimerisation occurs between the N-terminal half of the heavy subunit and the light subunit, whereas the C-terminal half of the heavy subunit is flexibly linked and poorly resolved. In addition, the protein contains multiple domains of different folds, which have been associated with binding to a variety of ligands. The chapter also provides preliminary structural data on its C-terminal half, and its main interacting partner, the 150-kDa antigen.

3.2 Introduction

3.2.1 Intestinal infection by *Entamoeba histolytica*

Human amoebiasis is the third most common cause of deaths from parasitic diseases after malaria and schistosomiasis, with up to 500 million infections causing between 40,000 and 110,000 deaths annually [250, 251]. These are prevalent in developing countries with poor sanitation, mostly in Central and South America, Africa and Asia.

Amoebiasis is caused by the intestinal pathogen *Entamoeba histolytica*, a single-celled amoeba with a relatively simple life cycle (Figure 3.1.a). Mature quadrinucleate cysts infect human hosts through ingestion of contaminated food or water. They can survive harsh acidic conditions of the stomach to reach the intestinal lumen. There, excystation occurs, during which a cyst undergoes nuclear and cytoplasmic division to form eight rapidly growing trophozoites. Trophozoites feed on bacteria and nutrients present in the colonic environment [252] and can reproduce by binary fission, or encyst to form cysts that are passed through feces to repeat a new infection cycle [253].

In 10% of all infections, trophozoites become invasive through an unknown molecular switch. In this case, they adhere to and degrade mucins that coat the colonic epithelium. Then, trophozoites contact exposed epithelial cells and ingest bits of them until they die, through a process called trogocytosis [257] (Figure 3.1.b). This leads to tissue destruction and formation of intestinal ulcers (Figure 3.1.c). In some cases, parasites enter the blood and reach secondary sites of infection such as the liver, where they can cause amoebic liver abscesses (ALA), which contain cellular debris, liquefied cells and dead hepatocytes [256]. Therefore, patients with amoebiasis may suffer from dysentery and/or extra-intestinal disease.

Currently, treatment of patients with intestinal amoebiasis relies on a combination of anti-amoebicidal tissue-active agents to kill invasive trophozoites and luminal cysticidal drugs to prevent parasite invasion and transmission [258]. Although they cause a wide range of side effects, such as nausea, anorexia, metallic taste, headaches and diarrhea, they remain widely used since there is no vaccine against amoebiasis.

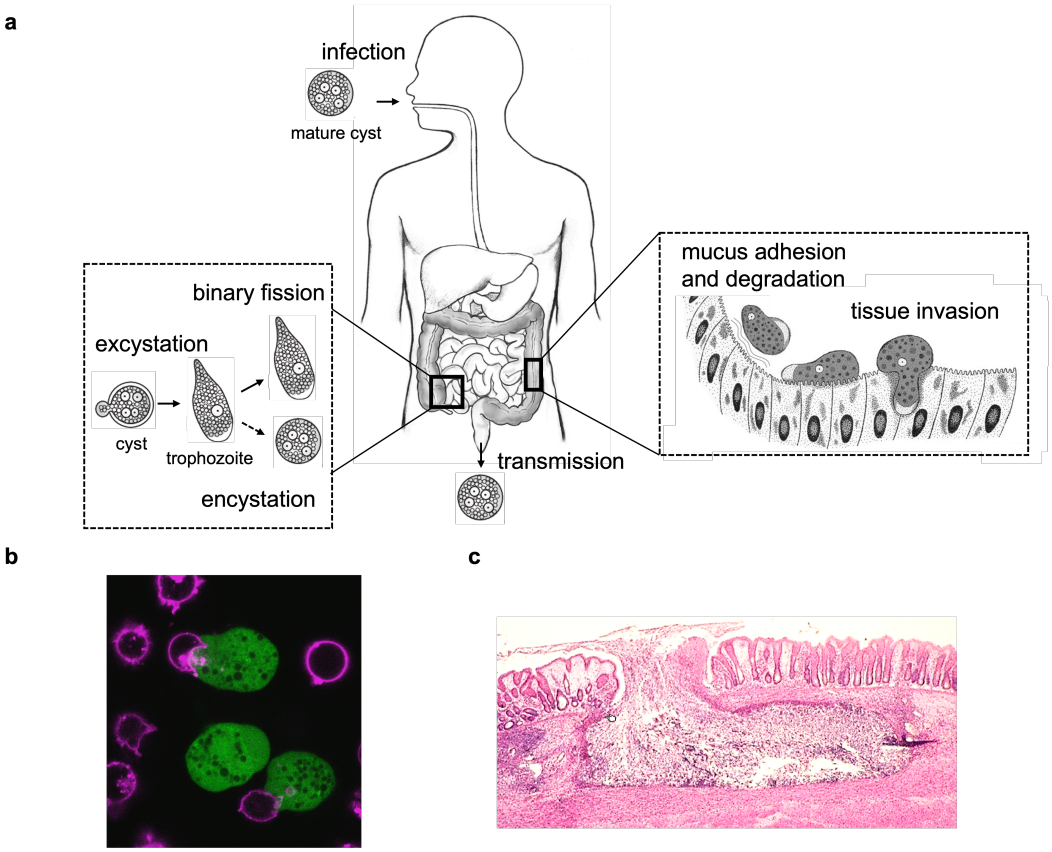


Figure 3.1: Infection by *Entamoeba histolytica*. a) Life cycle of *Entamoeba histolytica* (adapted from [254]). b) Trogocytosis of human Jurkat T cells and their membranes (in purple) by *Entamoeba histolytica* trophozoites (in green, taken from [255]). c) Typical flask-shaped ulcer, a hallmark of human amoebiasis (taken from [256]).

3.2.2 The 260-kDa surface antigen of *Entamoeba histolytica*

3.2.2.1 Identification of the 260-kDa antigen

Invasion by *Entamoeba histolytica* starts with adherence of trophozoites to the protective mucin coat on the surface of colonic cells. There, mucins form a dense network of O-glycosylated proteins. In early *in vitro* studies of *Entamoeba histolytica*, pre-treatment of trophozoites with galactose (Gal) and N-acetylgalactosamine (GalNAc) prevented adherence and cytolysis to target cells [259]. This suggested the presence of a Gal/GalNAc-specific surface lectin important in virulence.

Later, a panel of mouse monoclonal antibodies (mAbs) against trophozoites was obtained, four of which inhibited trophozoite binding to Chinese hamster ovary (CHO) cells [260]. Using these mAbs, immuno-affinity chromatography allowed purification

of a detergent-solubilised lectin, identical to that obtained by Gal-affinity chromatography [261].

The protein was characterised as a 260-kDa heterodimer, which consists of a heavy subunit (HgL) of 170 kDa covalently associated with a 31/35-kDa light subunit (LgL) [262] (Figure 3.2). In *Entamoeba histolytica*, five isoforms of HgL share 89-95% sequence homology, while seven isoforms of LgL share 79-85% homology [263]. The sequence of HgL features a large extracellular domain of more than 1,200 residues, a single transmembrane span, and a short cytoplasmic tail [264, 265]. Notably, a large C-terminal stretch of the extracellular domain is rich in cysteines, in contrast to the N-terminus. As for LgL, it is exclusively extracellular, though a 31-kDa isoform may be GPI-anchored [266].

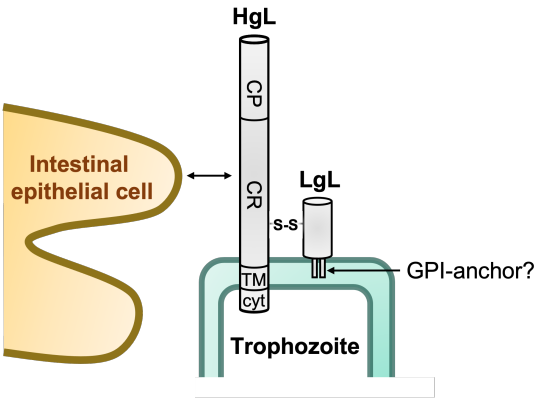


Figure 3.2: Schematic representation of the 260-kDa surface heterodimer. HgL contains a cytoplasmic domain (cyt), a transmembrane domain (TM), a cysteine-rich region (CR) and an N-terminal region poor in cysteines (CP). It is covalently linked to LgL, which may be GPI-anchored.

3.2.2.2 Functions of the 260-kDa antigen

Following its identification, the roles of the 260-kDa heterodimer in amoebic virulence were investigated.

3.2.2.2.1 Adhesion to colonic mucins

3.2.2.2.1.1 Lectin specificity

Though the 260-kDa heterodimer is a Gal/GalNAc lectin, requirements for optimal carbohydrate binding was studied for complex ligands that are more biologically relevant than monosaccharides. For instance, asialofetuin and asialoorosomucoid, which contain

terminal Gal residues, were 1,000-fold more potent inhibitors of trophozoite adherence to target cells than Gal. By contrast, fetuin and orosomucoid, for which Gal is masked by terminal sialic acid moieties, had no inhibitory effect [261]. In addition, the lectin displayed at least 100,000-fold increase in affinity for Gal and GalNAc polyvalent bovine serum albumin (BSA) neoglycoproteins compared to monosaccharides [267].

While the nature of the minimal endogenous carbohydrate ligand of the 260-kDa lectin remains unknown, various ligands were tested for their inhibitory effects on binding of polyvalent GalNAc neoglycoprotein to trophozoite membranes. In particular, 3-OH, 4-OH and 6-OH groups of GalNAc were important for carbohydrate binding, since their deletion or modification prevented inhibition [268]. Besides, as opposed to high affinity observed for GalNAc polyvalent neoglycoproteins, the 260-kDa heterodimer bound to monovalent GalNAc and 'miniclusters' of GalNAc displayed on polyvalent dendrimers with similar affinities. Therefore, optimal carbohydrate binding appears to occur with ligands that display multiple terminal Gal or GalNAc carbohydrates that are spaced by a large enough distance. This is observed in mucins, which harbour extensive O-linked glycans with terminal Gal and GalNAc residues (Figure 3.3). In support of this, partially purified colonic mucins are the most potent inhibitors of amoebic adherence to epithelial cells [269, 270].

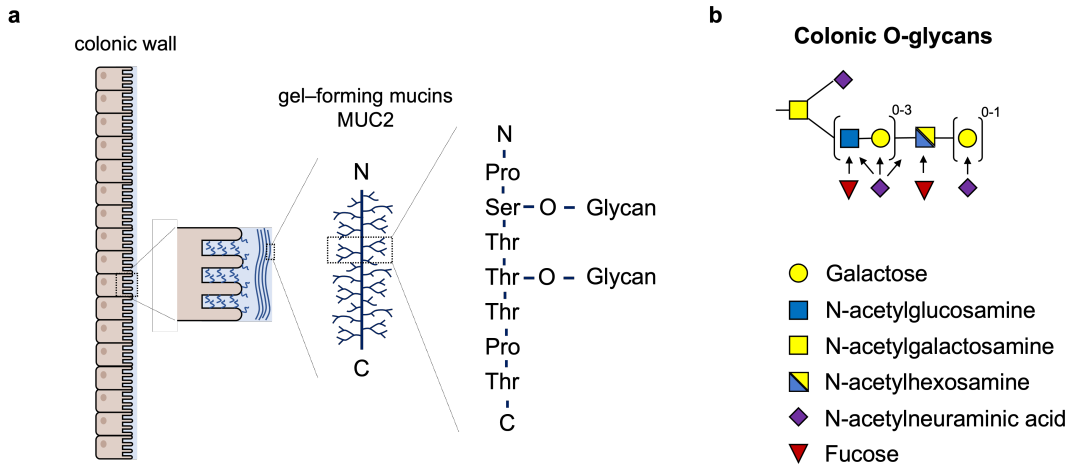


Figure 3.3: Colonic mucins and their O-glycans. a) Schematic representation of the mucus coat on the surface of colonic cells, from cell to molecule, with a close-up on the predominant intestinal mucin (MUC2). b) Schematic representation of human colonic O-glycans, as observed in [271, 272]

3.2.2.2.1.2 Carbohydrate recognition domains

To date, a single carbohydrate recognition domain (CRD) has been identified as a 107 amino-acid fragment (residues 892 to 998) included in the cysteine-rich region of HgL [273]. The CRD was shown to bind to BSA$(GalNAc)_{19}$ and to Gal-GalNAc disaccharide [273, 274]. Nonetheless, some inhibitory mAb epitopes map to regions distinct to the CRD (residues 596 to 818, 1033 to 1082 and 436 to 624) [275, 276]. This suggests the presence of other CRDs, though carbohydrate binding remains uncharacterised.

Sugar binding is attributed solely to HgL. However, determination of LgL carbohydrate binding has been hampered by difficulties to express it recombinantly. The absence of LgL carbohydrate binding was suggested by pull-down experiments, in which GPI-deficient LgL from cell lysates failed to bind to HgL and Gal beads [277]. However, the behaviour of GPI-defficient LgL in lysates remains unclear and residues truncated may be important for its structure and function. Regardless, sequence analysis of Hgl and LgL did not reveal any sequence similarities with other known carbohydrate binding domains.

3.2.2.2.1.3 Modulation by monoclonal antibodies

Different antibodies against non-overlapping epitopes of HgL showed different effects on trophozoite adherence to CHO cells and human colonic mucins [276]. Indeed, out of three mAbs binding to the putative CRD, one was inhibitory, whereas two increased adherence by two to five fold (Figure 3.4.a). This suggests that antibody binding may stabilise conformations of the 260-kDa antigen in which carbohydrate binding is modulated. Modulation of carbohydrate binding through conformational changes could be a mechanism by which trophozoites disengage from carbohydrates of mucins and glycoproteins on the surface of epithelial cells during invasion [278].

Such mechanism is reminiscent to that adopted by integrins through "inside-out" signalling [282], for which activation leads to large conformational changes and increase in affinity for ligand binding. Remarkably, the cytoplasmic tail of HgL shares homology with that of β_2 and β_7 integrins, which are important for "inside-out" signalling (Figure 3.4.b). In *Entamoeba histolytica*, over-expression of HgL cytoplasmic domain led to 50% decreased adherence of trophozoites to CHO cells, whereas this effect was reverted by

a

Designation	Adherence	Location on HgL
3F4	Increases	895-998
8A3	Increases	895-998
8C12	Inhibits	895-998
1G7	Inhibits	596-818
H85	Inhibits	1033-1082
CP33	Inhibits	578-1154
Anti-r170PR	Inhibits	436-624

b

```
                          ***       *
HgL1    KTYQLVSSAMKNAITITNENAEYVGADNEATNAATFNG
β7      K--QDSNPLYKSAITTTI-NPRFQEADSPTL
β2      KSQWNNDNPLFKSATTTVMNPKFAES
```

Figure 3.4: Mapping of antibodies that affect adherence and similarities with integrins. a) List of antibodies affecting amoebic adherence and their corresponding epitopes [275, 279, 280]. b) Sequence alignment of the cytoplasmic sequences of β integrins and HgL. Asterisks highlight residues essential for 'inside-out' signalling [281].

mutating residues shared with β_2 integrin at the cytoplasmic tail [281]. This indicates signalling is important for adherence and virulence.

3.2.2.2.2 Other roles in amoebic pathogenicity

The 260-kDa antigen mediates contact-dependent cell killing. This was first hinted by abrogation of trophozoite adherence to and cytolysis of CHO cell mutants lacking terminal Gal moieties on their N- and O-linked glycans [283]. Furthermore, antibodies against three different epitopes in the cysteine-rich region of HgL blocked both host cell binding and killing [276]. However, one antibody was also found to block cytotoxicity, while increasing adherence [284] and also inhibited trogocytosis, the nibbling and killing of host cells [257]. This antibody was suggested to modulate the ability of the 260-kDa antigen to transduce cell signalling towards cytolysis, and signalling through the 260-kDa antigen was proposed to be essential for amoebic trogocytosis and phagocytosis [285]. Overall, cell adherence and killing by the amoebic heterodimer appear to be distinct, but linked, possibly through signalling.

During infection, *Entamoeba histolytica* is exposed to the complement system of the host, which can lyse parasites. Interestingly, invasive parasites were found to be resistant to membrane attack complexes C5b-9 deposited on the surface of amoebae

[286, 287]. In contrast, non-invasive parasites were susceptible to complement-mediated killing by C5b-9. Importantly, two antibodies raised against parasites adapted to human sera drastically increased lysis of *Entamoeba histolytica* by C5b-9 and bound to HgL [288]. Likewise, HgL cross-reacted with an antibody against CD59, a surface protein that inhibits formation of the membrane attack complex, and both proteins shared sequence similarities. Last, the 260-kDa antigen purified from serum-adapted trophozoites interacted with complement proteins C8 and C9. Altogether, these data suggest a role for the 260-kDa complex in evading complement-mediated killing.

3.2.3 Interacting partners of the 260-kDa antigen

Co-purifications, co-localisation assays, genetic manipulations and proteomic analyses reveal that the 260-kDa antigen engages with various proteins. This affects parasite interactions with its environment and protection from its host.

3.2.3.1 The intermediate subunit of 150 kDa (IgL)

Among these interacting partners, a 150-kDa protein was purified from the surface of *Entamoeba histolytica* trophozoites using an antibody that inhibited adherence and cytotoxicity to host cells [289, 290]. The protein was co-purified with the 260-kDa amoebic antigen and both were estimated to form a 380-kDa complex by native gel electrophoresis. In addition, the 150-kDa protein is consistently detected in preparations of the 260-kDa heterodimer by Gal-affinity and immuno-affinity chromatographies using antibodies against HgL. Furthermore, HgL, LgL and IgL co-localise on the surface of trophozoite membrane [291].

Therefore, the 150-kDa protein was named the intermediate subunit and appears to form a non-covalent heterotrimeric complex with LgL and HgL (Figure 3.5). However, the subunit with which IgL associates remains unknown. Moreover, IgL separates from the 260-kDa antigen by SEC, suggesting a relatively weak interaction in solution [289].

IgL consists of two isoforms that share 84% sequence homology [291]. It is GPI-anchored and rich in cysteines, with characteristic CXXC and CXC motifs found across the whole extracellular sequence. IgL was proposed to belong to a large family of

over 80 transmembrane kinases, which all share CXXC motifs, but contain additional cytoplasmic kinase domains [292]. Interestingly, isoforms IgL1 and IgL2 share 97 and 99% sequence homology with a β_1 integrin-like fibronectin receptor, respectively [293]. This suggests a role of IgL in parasite adhesion to the extracellular matrix.

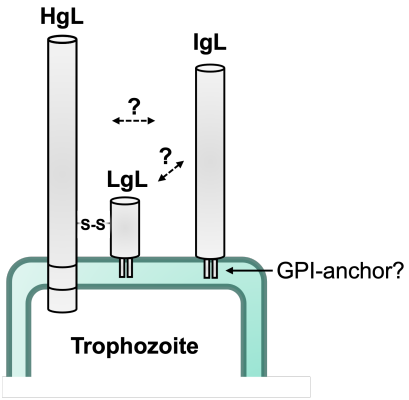


Figure 3.5: A bigger surface complex. The intermediate subunit IgL associates with the 260-kDa antigen in an unknown manner.

Passive immunisation of hamsters with a mAb against IgL prevented formation of amoebic liver abscesses (ALA) [294]. In addition, immunisation with an IgL-containing fraction purified by immuno-affinity chromatography with the same antibody led to increased protection against development of ALA and a significant decrease in their size [295].

Later, mAbs raised against IgL were mapped to different regions of the extracellular domain [296]. One antibody interacted with the N-terminus of IgL1 (residues 14 to 382) with no effect on amoebic adherence to CHO cells. In contrast, three antibodies (XEhI_20, XEhI_B5 and XEhI_H2) inhibited adherence, one of which (XEhI_20) dramatically reduced the size of ALA in hamsters. XEhI_H2 epitope was found between residues 653 and 753, but the other two inhibitory antibodies only bound to full-length IgL rather than fragments of N-terminus (residues 14 to 382), middle (residues 294 to 753) and C-terminus (residues 603 to 1088), suggesting conformational epitopes. The C-terminus displayed hemolytic and cytotoxic activities *in vitro* [297] and immunisation with this construct prevented formation of ALA [298]. Therefore, it is considered as a promising vaccine candidate.

3.2.3.2 Other potential extracellular interacting partners

Entamoeba histolytica trophozoites activate Toll-like receptor (TLR) signalling in target cells [299, 300]. A direct interaction between the putative CRD of HgL and TLR-2 and TLR-4 was reported, although binding to glycans on the receptor could not be excluded. Activation of TLR signalling mediated through the 260-kDa antigen led to increased expression of TLRs and secretion of cytokines, which contributed to inflammatory response that damaged epithelial cells *in vitro* [300]. Nonetheless, the importance of CRD-induced activation of TLRs in intestinal amoebiasis remains to be determined.

Furthermore, HgL putative CRD competes with the hepatocyte growth factor (HGF) for binding to its receptor (cMet) [273]. Intriguingly, the CRD shares sequence identity with the receptor binding domain of HGF. The molecular details of this interaction have yet to be characterised, but this could have implications in the development of ALA and tropism of trophozoites to the liver, since HGF/cMet signalling promotes liver repair and regeneration [301, 302].

During invasion, *Entamoeba histolytica* interacts with components of the extracellular matrix. For instance, incubation of parasites with Gal inhibited adhesion to collagen, suggesting that collagen binding requires prior lectin-mediated contact or that the 260-kDa antigen acts as a collagen receptor [303]. Notably, lectin-dependent collagen binding of trophozoites correlated with increased association of HgL and LgL with IgL on cholesterol-rich membranes and led to increased intracellular calcium levels [304]. Thus, collagen binding may modulate the activity of the amoebic antigen and trigger signalling.

3.2.4 Prospect for a vaccine against amoebiasis

Amoebiasis is not a vector-borne disease and it is mostly transmitted through contaminated food or water. Therefore, the disease could be controlled by improvement in water management systems and sanitary practices. However, this is often unaffordable for endemic countries. Thus, vaccination represents an attractive alternative. Different stages in disease could be targeted. For instance, blocking binding of trophozoites to the surface epithelium, or trophozoite formation through excystation could prevent

invasion. In addition, blocking encystation may contribute to reduce transmission. An effective vaccine against *Entamoeba histolytica* should trigger a strong immune response and immunological memory.

The 260-kDa heterodimer is responsible for the first steps of parasite invasion, therefore its ability to trigger an effective immune response against amoebiasis was investigated. Antibodies specific to the antigen were found in more than 95% of patients with amoebic colitis or ALA [305–307], indicating a systematic immune response. Moreover, in children from Bangladesh, a country strongly affected by amoebiasis, presence of mucosal IgA against the 260-kDa antigen was associated with protection from severe disease and colonisation of parasites [308].

The findings of naturally acquired immunity against the 260-kDa antigen inspired multiple vaccination trials in animals. Immunisation of rodents with parasite-extracted native antigen [309–311], recombinant fragments [273, 275, 307, 310, 312–317] or synthetic peptides [318] of HgL all led to some protection against amoebic colitis and liver abscesses. In baboons, which are natural hosts of *Entamoeba histolytica*, immunisation with HgL-based polylysine-linked synthetic peptide vaccine elicited IgA against the 260-kDa antigen, which inhibited amoebic binding to host cell and prevented colitis [319].

In addition to mucosal IgA response, vaccination with the 260-kDa antigen was able to induce production of Th1 cytokines *in vitro* and *in vivo*, such as tumor necrosis factor-α [320] in macrophages, interleukin-2 and interferon-γ in T cells [321, 322], and interleukin-12 in dendritic cells [322]. The complex also stimulated production of interleukin-17, which are Th17 cytokines known to contribute to mucosal immunity and inflammation [323]. Importantly, neutralisation of interleukin-17 or interferon-γ production in vaccinated mice abrogated protection [323, 324].

These studies show that the 260-kDa antigen is an exploitable vaccine candidate against amoebiasis. Nonetheless, recombinant fragments of the cysteine-rich region of HgL were less immunogenic and effective at eliciting an immune response, when compared to the native parasite complex. One such fragment, LecA, contains HgL residues 578 to 1154, which includes the putative CRD and epitopes of adhesion-inhibitory and -enhancing mAbs. While protection against amoebiasis was associated

with fecal IgA against the native antigen, no such association existed for IgA against LecA [310]. In addition, immunisation with LecA gave lower levels of protection in a gerbil model than with the native complex [310]. Increased protective efficacy of LecA could only be achieved through the use of synthetic TLR ligands as vaccine adjuvants, which increased levels of fecal IgA, interferon-γ and interleukin-17A in mice [317, 325]. However, comparison with the native antigen is lacking in these studies.

Since the native complex is challenging to produce in decent quantities, new approaches are required to design cheap and scalable vaccine immunogens which match, or even exceed its efficacy. These could be guided by structural studies of the multiple domains of the 260-kDa antigen and assessment of their immunogenicity. Moreover, since different mAbs against the same epitope of HgL have opposing effects on adherence of trophozoites, structural insights into regulation by these mAbs is desirable.

3.2.5 Aims

The main goal of this project is to structurally characterise the 260-kDa antigen. Given the size of the complex, the protein is unlikely to act only as a lectin. Likewise, we aim to solve the structure of IgL and study its interaction with the 260-kDa antigen. Ultimately, such structures could reveal further insights into the functions of these antigens and guide targeted vaccine design in the fight against amoebiasis.

3.3 Materials and methods

3.3.1 Construct design and molecular cloning

All primers and synthetic genes were purchased from Sigma Aldrich and GeneArt (Invitrogen), respectively. PCR reactions were performed using PCR mastermix 2X (#K0171, ThermoFisher Scientific) or KOD Xtreme Hot Start DNA polymerase kit (#71975, Novagen) for fragments exceeding 3 kbp in size. Sequence insertion into linear vector was done with Gibson Assembly cloning kit (#E5510, New England BioLabs). Sequencing reactions were performed by Source Bioscience and sequence analysis was done with SerialCloner 2.6 (SerialBasics).

Signal peptide, transmembrane domain and GPI anchor predictions were performed in SignalP 5.0 [326], TMHMM [327] and PredGPI [328], respectively. Protein disorder and N-linked glycosylation sites were predicted in PONDR [329] and NetNGlyc 1.0 [330], respectively. Protein molecular weights and extinction coefficients were calculated in ProtParam (Expasy). Structure-homology modelling was performed using a combination of Phyre2 [331], I-TASSER [332], Swissmodel [333], trROSETTA [334]. This was later complemented by modelling in RosettaFold [335] and AlphaFold [336].

All constructs and primers used in this thesis are listed in Appendix Table A.1.

3.3.1.1 Constructs of heavy, light and intermediate subunits

Two genes encoding the N-terminal (1-532) and the C-terminal (512- 1219) portions of *Entamoeba histolytica* HgL isoform 3 (EHI_133900), three synthetic genes encoding for the full-length isoforms 1, 3, and 6 (EHI_035690, EHI_148790 and EHI_183400 respectively) and a synthetic gene encoding for IgL (residues 14 to 1082) isoform 1 (EHI_006980), all codon-optimised for mammalian expression, were purchased.

Constructs were amplified by PCR and extracted from agarose gel using Monarch DNA Gel Extraction Kit (#T1020, New England BioLabs) for insertion into pHLSec, a vector suitable for protein secretion [337].

The pHLSec plasmid includes a translation initiation site (Kozak motif), the µ-phosphatase secretion signal and a C-terminal hexahistidine tag. The vector was digested by overnight incubation with AgeI-HF and KpnI-HF (#R3552 and #R3142, New England BioLabs, respectively) and purified with the QIAquick PCR purification kit (#2814, QIAGEN). For each construct, a Gibson assembly of the PCR-pure sample into pHLsec was performed. For surface plasmon resonance (SPR) coating trials, several constructs were cloned into pHLSec vector modified with a C-terminal AVI tag preceding the hexahistidine tag.

3.3.1.2 Constructs of fragments of antibodies

Synthetic genes encoding the variable portions of the heavy and light chains of CP33/LA22 (referred to as CP33 here) [280], XEhI_20, XEhI_28, XEhI_H2 [296] and EH3015 [338] were purchased.

As described in section 3.2.1.1, these variable domains were amplified and assembled into pOPINVH and pOPINVL [339], digested by overnight incubation with KpnI and SfoI (#R0142S and #R0606S, New England BioLabs) or KpnI and SacI (#R0142S and #R0156S, New England BioLabs), respectively. pOPINVH and pOPINVL contain mouse IgG1 CH1 heavy chain and IgG1 K light chain constant regions, respectively, C-terminal to the inserted variable domains. For SPR experiments, an AVI-tag was added to the C-terminus of the constant region of pOPINVL by Gibson assembly.

CP33 scFv contains the variable region of the heavy chain of CP33 at the N-terminus, joined to the variable domain of the light chain LA22 by a flexible linker of 18 residues (GGSSRSSSSGGGGSGGGG), as described previously [340]. Both variable regions were PCR-amplified and gel-purified separately, as described in section 3.2.1.1, with overlapping primers containing sequence encoding for the linker. Both fragments were inserted into pHLSec with a C-terminal hexahistidine tag by Gibson assembly.

3.3.1.3 Constructs for bacterial refolding of HgL putative CRD and LgL

HgL putative CRD (residues 892 to 998) and LgL putative β-trefoil domains were amplified by PCR and inserted by Gibson assembly into pEt15b (Novagen), modified by the insertion of DNA encoding an N-terminal hexahistidine tag and a cleavage site for the TEV protease, digested by BamHI and NdeI (#R3136 and #R0111, New England BioLabs, respectively).

3.3.1.4 Construct of HgL putative VWD fused to glutathione S-transferase

HgL putative VWD (residues 211 to 388) was amplified by PCR and inserted by Gibson assembly into pGEX-2T (Novagen), previously digested by BamHI and EcoRI (#R3136 and #R3101, New England BioLabs, respectively). The resulting plasmid contains an N-terminal sequence for glutathione S-transferase (GST), followed by a thrombin cleavage site and the insert.

3.3.2 Protein expression, purification and modification

3.3.2.1 Protein expression and purification from Expi293F cells

Unless otherwise stated, all constructs were expressed by Expi293F cells (GIBCO, Life Technologies), a cell line based on human embryonic kidney (HEK) cells and adapted in suspension.

Transfections were performed at a density of 3.0 to 4.0 million cells per mL in the presence of 1 µM kifunensine [341], an α-mannosidase I inhibitor that prevents formation of complex N-glycans. After 16-18 h, transfection enhancers were added. After 5-7 days, cells were pelleted by centrifugation at 4,000 rpm for 45 min. Supernatants were filtered and applied on Ni-NTA beads (#NB-45-00042, Generon) and purified by IMAC using 50 mM Tris-HCl pH 8.0, 300 mM NaCl, with 10 mM imidazole during equilibration and washes, and 250 mM imidazole during elution. For deglycosylation, Endo-Hf (#P0703, New England BioLabs) was added and incubated overnight at room temperature before application on SEC column (Superdex S200 or S75 Increase 10/300 GL, GE Healthcare) in 20 mM HEPES pH 7.4, 150 mM NaCl.

3.3.2.2 Bacterial refolding of HgL putative CRD and LgL

HgL putative CRD was expressed in BL21-CodonPlus (DE3)-RIPL competent cells (#230280, Agilent). A 50-mL overnight starter culture in LB and 50 µg/mL ampicillin, 50 µg/mL chloramphenicol was used to inoculate 2 L of 2-YT medium containing 100 µg/mL ampicillin. Protein expression was induced at an optical density at 600 nm of 0,6-0,8 by addition of 100 µM IPTG, and cells were shaken at 160 rpm, 20°C overnight.

After centrifugation at 4,000 rpm for 20 min, cell pellets were resuspended in 20 mM Tris pH 8.0, 500 mM NaCl, and lysed by sonication in the presence of protease inhibitors (#A32965, ThermoFischer Scientific). After centrifugation at 20,000 rpm, 4°C for 1 h, the supernatant was discarded and inclusion bodies were resuspended in 6 M guanidine chloride, 20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol using a Dounce homogeniser and incubated overnight on a roller shaker. After centrifugation at 22,000 rpm for 1 h, the supernatant was applied to Ni-NTA agarose beads, equilibrated with wash buffer (20 mM Tris pH 8.0, 500 mM NaCl,

20 mM imidazole, 6 M guanidine chloride). Extensive washes were followed by dropwise elution with 20 mL of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole, 6 M guanidine into a 1-L flash dilution buffer (400 mM L-arginine pH 8.0, 10 mM EDTA, 0.1 mM phenylmethanesulfonylfluoride, 3 mM reduced glutathione, 0.3 mM oxidized glutathione). After stirring at 4°C overnight, the mix was dialysed (3,500 MW cut-off) overnight against 20 mM Tris-HCl pH 8.0, 500 mM NaCl with four exchanges of 10 L.

HgL putative CRD was purified by IMAC with Ni-NTA beads using 20 mM Tris-HCl pH 8.0, 300 mM NaCl, supplemented with 10 mM imidazole during equilibration and washes, and 250 mM imidazole during elution. The eluate was concentrated and applied on SEC column (Superdex S75 Increase 10/300 GL, GE Healthcare) in 20 mM HEPES pH 7.4 and 150 mM NaCl. For crystallographic studies, the hexahistidine tag was removed by overnight incubation with TEV protease (1:100 molar ratio) at room temperature, collected by reverse IMAC and purified by SEC.

For LgL β-trefoil domains, the same protocol was used but refolding buffers were first screened using Quickfold refolding kit (#MD1-28, Molecular Dimensions). The selected refolding buffer contained 50 mM MES, pH 6.0, 240 mM NaCl, 10 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 750 mM guanidine chloride and 5 mM DTT. Buffer was exchanged against the same buffer but without guanidine nor DTT by dialysis, tangential flow filtration or IMAC.

3.3.2.3 Bacterial expression of GST-fused HgL VWD

GST-fused HgL VWD was expressed as described in section 3.3.2.2. However, after centrifugation of the lysate, the supernatant was filtered and applied to glutathione agarose beads (#16101, ThermoFisher Scientific), equilibrated and washed with 20 mM Tris-HCl pH 7.5, 150 mM NaCl, and eluted with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 50 mM reduced glutathione. The eluate was concentrated and applied on SEC (Superdex 75 Increase 10/300 GL, GE Healthcare) in 20 mM HEPES pH 7.4 and 150 mM NaCl.

3.3.2.4 SDS-PAGE and immunoblotting

Protein samples were separated by SDS-PAGE on homemade gels or 4–20% acrylamide gradient gels (Bio-Rad) and stained with Coomassie (InstantBlue, #ISB1L, Expedeon). Molecular weights were estimated using a protein marker (#26623, ThermoFisher Scientific)

For Western blots, samples were transferred onto polyvinylidene difluoride membranes, followed by blocking with 5% skim milk dissolved in phosphate-buffered saline (PBS) with 0.1% Tween-20 (PBS-T) for 1 h at room temperature. After incubation with penta-His HRP conjugate (#34460, Qiagen) diluted in PBS-T for 1 h at room temperature, membranes were washed with PBS-T and incubated with Amersham ECL detection reagents (#RPN3004, Cytiva). Membranes were developed on films or imaged with iBright FL1500 (#A44241, ThermoFisher Scientific).

3.3.2.5 Protein methylation

Lysine methylation was used to alter surface charge and solubility for crystallisation and cryo-EM sample modification. The reaction was carried out as previously described [342]. Deglycosylated proteins purified by SEC in 20 mM HEPES pH 7.4, 150 mM NaCl were concentrated to 1 mg/mL. 20 µL of freshly prepared 1 M dimethylamine-borane complex (#15584, Fluka) and 40 µL of 1 M formaldehyde were added per mL of protein solution. After incubation in the dark for 2 h at 4°C, the same amount of reagents was added. After 2 h, 10 µL of 1 M dimethylamine-borane complex was finally added, and the reaction was incubated overnight at 4°C. Precipitation was removed by high speed centrifugation and an equal amount of 20 mM Tris pH 7.5, 150 mM NaCl was added to quench the reaction before application on SEC in 20 mM Tris pH 7.5, 150 mM NaCl.

3.3.2.6 Biotinylation of Fab fragments

GST-fused BirA ligase, purified as described in section 3.3.2.3 (Figure 3.6.a), was incubated *in vitro* with AVI-tagged constructs (Figure 3.6.b) in the presence of an excess of biotin, 5 mM $MgCl_2$ and 2 mM ATP for 2 h. Biotinylated proteins were separated from GST-BirA using glutathione agarose beads (#16101, ThermoFisher

Scientific) and further dialysed against PBS overnight to ensure complete removal of biotin. Biotinylation efficiency was assessed by streptavidin shift assay (Figure 3.6.c,d). In short, an excess of streptavidin (#S4762, Sigma Aldrich) was added to protein samples after denaturation at 95°C and a shift of the protein band to higher bands indicated their association with streptavidin via biotinylated tags.

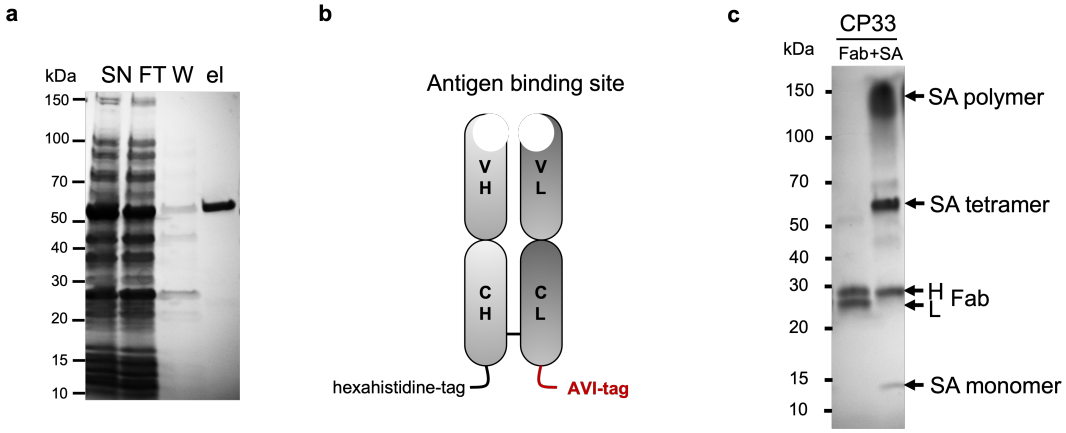


Figure 3.6: Selective biotinylation of Fabs. a) Reducing SDS-PAGE of the steps during purification of GST-BirA. Samples includes supernatant (SN), flow-through (FT), wash (W) and elution (el). b) Schematic representation of a Fab with a biotinylation site at the C-terminal end of the light chain, and a polyhistidine tag at the C-terminal end of the heavy chain. c) Streptavidin (SA) shift assay of biotinylated CP33. The light chain (L) has migrated, indicative of association with streptavidin, therefore successful biotinylation.

3.3.2.7 Limited proteolysis and mass spectrometry analysis

Limited proteolyses were carried out with trypsin, chymotrypsin, staphylococcus aureus protease V8 (GluC), flavastacin (AspN) and carboxypeptidase B added in a 1:100 molar ratio (protein:protease) and incubated at room temperature. Samples collected over time points were analysed on reducing SDS-PAGE and the reaction was scaled up, applied on SEC, and fractions were analysed on reducing SDS-PAGE. Coomassie-stained gel bands were excised and sent for analysis by mass spectrometry (MS, Advanced Proteomics Facility, University of Oxford) after in-gel tryptic digestion.

3.3.3 Purification of the 260-kDa antigen from amoebae

3.3.3.1 Culture of trophozoites in axenic conditions

Entamoeba histolytica HM1:IMSS trophozoites were cultured in a biological safety cabinet (containment level 2), wearing eye protection, lab coat and double gloves. Cells were lysed in the cabinet before removal and liquid waste was decontaminated by incubation in 1% Virkon.

3.3.3.1.1 Adaptation of amoebae in axenic medium

A 13-mL cell starter culture of HM1:IMSS trophozoites was kindly provided by Pr Graham Clark (London School of Hygiene and Tropical Medicine), together with a 100-mL aliquot of adult bovine serum (#P30-0601, lot P132607, PAN Biotech) and axenic medium. These were used to maintain healthy parasites while progressively adapting them to components of homemade LYI-S2 axenic medium. Unless otherwise stated, all reagents were purchased from Sigma Aldrich.

LYI-S medium was first prepared by mixing 1.0 g potassium phosphate dibasic, 0.6 g potassium phosphate monobasic, 1.0 g cysteine, 0.2 g ascorbic acid, 2.0 g sodium chloride, 10.0 g glucose, 5.0 g Oxoid liver digest neutralised (#LP0027, ThermoFisher Scientific), 25.0 g yeast extract, 45.6 mg ferric ammonium citrate (#0215804080, MP Biomedicals) in water to a total volume of 880 mL, with pH adjusted to 6.8. This medium was autoclaved and stored at -20°C.

A 200-mL vitamin mix stock was prepared by mixing four solutions S1, S2, S3 and S4. S1 contained 45 mg niacinamide, 4 mg pyridoxal hydrochloride, 23 mg calcium pantothenate, 5 mg thiamine hydrochloride and 1.2 mg vitamin B12 dissolved in 25 mL water. S2 consisted of 7 mg riboflavin and 5.5 mg folic acid dissolved in 45 mL of water with minimal amount of 0.1 M sodium hydroxide. S3 was prepared by dissolving 2 mg D-biotin in 45 mL of water with minimal amount of 0.1 M sodium hydroxide. S4 contained 1 mg thioctic acid dissolved in 5 mL of 95% ethanol, combined with 500 µL of Tween-80 in water to a total volume of 30 mL. S1, S2 and S3 were mixed together, then S4 was added. The resulting vitamin cocktail was filtered through a 0.22 µm filter and stored at 4°C.

950 mL of fresh LYI-S2 medium was prepared by mixing 880 mL of LYI-S with 20 mL of vitamin mix stock and 150 mL of heat-inactivated adult bovine serum.

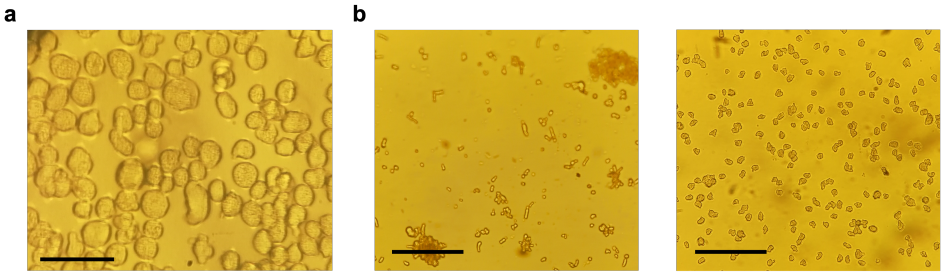


Figure 3.7: Parasite culture of *Entamoeba histolytica* HM1:IMSS. a) Light microscopy of *Entamoeba histolytica* trophozoites. Scale bar is equal to 100 µm. b) Light microscopy of *Entamoeba histolytica* trophozoites that fail to adapt to serum (left, scale bar is equal to 600 µm) compared to serum-adapted trophozoites (left, scale bar is equal to 300 µm)

Initially, parasites were grown in glass tubes (#9826-16X, Pyrex), filled with 13 mL of medium, tightly sealed (anaerobic conditions) and tilted by 15° from vertical in a static incubator at 35.5°C. After 3 days, confluent parasites on the surface of glass tubes were passaged by gently removing clumps of dead cells, placing the tube on ice for 5 min and transferring aliquots (100 µL to 2 mL) to fresh medium in autoclaved tubes filled up to 13 mL. Parasites were gradually adapted to homemade LYI-S2 medium by 20% increase in composition for each passage and cell growth was monitored with a light microscope (Yenway ICX41) each day for each aliquot size, sera and yeast extracts tested (Figure 3.7). After selection and adaptation, healthy trophozoites were routinely passaged every 3 days in fresh LYI-S2 axenic medium. Successful adaptation was obtained with two lots of adult bovine sera (#B9433, lot 16N203, Sigma Aldrich; #16170086, lot 1671327, ThermoFisher Scientific) and one yeast extract (#Y1625, lot SLBX6973, Sigma Aldrich).

3.3.3.1.2 Scale up of axenic cultures

500 µL of trophozoites were passaged from 15-mL test tubes, as described in section 3.3.3.1.1, to T-25 flasks filled with 50 mL fresh LYI-S2 medium and tightly sealed with cap and parafilm. T-flasks were placed vertically and incubated in a static incubator at 35.5°C for 3 days. Passage from a T-flask to another was performed by carefully discarding medium and washing cell clumps with PBS at the bottom of vertical T-flasks. A minimal amount of fresh medium (5 mL for T-25, 15 mL for T-75, 25 mL

for T-175) was added to the surface and flasks were placed on ice for 5 min. Flasks were tapped to ensure detachment of most parasites, and these were transferred to new flasks filled with fresh medium (500 µL in T-25, 2 mL in T-75 or 5 mL in T-175). Large scale ups were performed in multiple T-175 flasks and harvested after 3 days. The medium was discarded, and PBS was added to flasks that were placed on ice for 5 min. Detached parasites were transferred to falcon tubes and centrifuged at 1,000 rpm for 10 min. Parasite pellets were washed with PBS followed by centrifugation at 1,000 rpm for 10 min, then pellets were stored at -80°C.

3.3.3.2 Purification of the 260-kDa antigen

A 5 g pellet of trophozoites was resuspended in PBS supplemented with EDTA-free protease inhibitors, Dounce-homogenised and sonicated on ice (60 mA, 3 s pulse on, 10 s pulse off, 30 pulses). Lauryl maltose neopentyl glycol (LMNG, 3% stock solution in PBS) or octyl β-D-glucopyranoside (OG) was added to a final concentration of 1%. Alternatively, for extraction in lipids, the lysate was incubated with 2% DIBMA dissolved in PBS. The sample was incubated at 4°C overnight on a tube roller, then centrifuged at 20,000 rpm for 1 h at 4°C . The supernatant was filtered and further purified by affinity chromatography.

3.3.3.2.1 Galactose-affinity chromatography

The supernatant was incubated with agarose resin coupled to Gal (#20372, ThermoFisher Scientific) at room temperature for 2 h on a tube roller. Beads were washed with PBS with 1.2 critical micellar concentration (CMC) of LMNG or OG and elution was done with PBS, 0.5 M Gal, 1.2 CMC LMNG or OG. The eluate was concentrated to 500 µL using a 100-kDa cutoff concentrator and applied on SEC (S200 Increase 10-300 GL, GE Healthcare) in 20 mM HEPES pH 7.4, 150 mM NaCl, 1.2 CMC LMNG or OG. The yield obtained before SEC ranged from 50 to 75 µg for 5 g of starting pellet.

For DIBMA-solubilised sample, the same protocol was followed without addition of detergent.

3.3.3.2.2 Immuno-affinity chromatography

25 mg of CP33 Fab were coupled per g of cyanogen bromide-activated sepharose 4B beads (#C9142, Sigma Aldrich) following instructions from the manufacturer. The resin was equilibrated in PBS, added to the supernatant and incubated at room temperature for 1 h on a tube roller. Beads were washed with PBS with 1.2 CMC LMNG or OG and elution into 500 µL fractions was done with 0.1 M glycine pH 2.7, 1.2 CMC LMNG or OG. Samples were eluted into Eppendorf tubes with 50 µL Tris pH 9.0, 1.2 CMC LMNG or OG for neutralisation. Fractions were analysed on gel and lectin fractions were pooled together. After concentration to 500 µL using a 100-kDa cutoff concentrator, the sample was applied on SEC (S200 Increase 10-300 GL, GE Healthcare) in 20 mM HEPES pH 7.4, 150 mM NaCl, 1.2 CMC LMNG or OG. The yield obtained before SEC ranged from 150 to 250 µg for 5 g of starting pellet.

Alternatively, CP33 Fab was added to parasite supernatant and the Fab-bound 260-kDa antigen was purified by IMAC using 20 mM Tris pH 8.0, 200 mM NaCl, 1.2 CMC LMNG supplemented with 20 mM imidazole for equilibration and washes, and with 300 mM imidazole for elution, before application on SEC. For DIBMA-solubilised sample, the same protocol was followed without addition of detergent.

3.3.4 Functional studies of the 260-kDa antigen

3.3.4.1 Interaction with antibodies and ligands by surface plasmon resonance

SPR measurements were performed at 25°C on a BIAcore T200 machine (GE Healthcare) with a buffer composed of 20 mM HEPES pH 7.5, 150 mM NaCl, 0.05% Tween-20 and a CM5 chip (#29104988, Cytiva) for covalent coating or a CAPture chip (#28920234, Cytiva) for capture of biotinylated proteins on streptavidin.

Binding of HgL constructs to CP33 Fab was investigated. A CM5 chip surface was activated with a 1:1 molar mixture of 0.4 M (3-dimethylamino-propyl)-ethyl-carbodiimide hydrochloride and 0.1 M N-hydroxysuccinimide for 7 min at 5 µL/min. CP33 Fab was added at 5 µL/min to reach 442 response units (RU). The first flow channel remained uncoupled for signal subtraction and assessment of non specific binding. Reactive groups

were deactivated with 1 M ethanolamine hydrochloride, pH 8.5. HgL constructs were used as analytes and injected at 100 nM at 20 µL/min for 120 s and dissociated for 300 s, followed by surface regeneration with 100 mM glycine hydrochloride pH 2.5 for 60 s until baseline RU was restored.

To obtain affinity parameters, biotinylated Fab CP33 was immobilised on a CAPture chip to reach 500 RU. HgL_09 (LecA, residues 579-1154) was injected for 240 s with a 300 s dissociation time before regeneration between each run. One dilution series with 11 dilutions was performed with a starting concentration of 500 nM.

Similarly, binding of the native lectin to a panel of mAbs (3F4, H85, 8A3 and 1G7) provided by Pr W. Petri Jr (University of Virginia) was assessed. This time, parasite-extracted 260-kDa antigen was coupled to 250 RU on a CM5 chip. mAbs were used as analytes, associated for 700 s and dissociated for 1400 s. Regeneration was complete after 180 s. One dilution series with 11 dilutions was performed, with a starting concentration of 100 nM and two-fold dilutions between consecutive series.

The same chip was used to investigate binding of immobilised native lectin to IgL, C8 (#C3535, Sigma Aldrich) and C9 (#C3660, Sigma Aldrich) analytes. IgL at 1 µM was injected at 20 µL/min for 100 s and dissociated for 300 s. C8 and C9 at 1 µM were injected at 20 µL/min for 240 s and dissociated for 300 s.

Data were fitted and affinities estimated with the BIAevaluation (GE Healthcare) software using a one site-binding model for Fab binding to analytes, and a two-site binding model for mAb analytes.

3.3.4.2 Enzyme-linked immunosorbent assay of HgL VWD binding to collagens

Enzyme-linked immunosorbent assays (ELISA) were performed with Nunc Maxisorp 96-well plates (#442404, ThermoFisher Scientific) coated with collagens I, III and IV (#C-7774, #C-4407, #C-5533, respectively, Sigma Aldrich) at 10 µg/mL in 10 mM acetic acid overnight, blocked for 1 h with 200 µL of 50 mg/mL BSA (#A7638, Sigma Aldrich) in TBS, then washed three times with TBS containing 2 mM $MgCl_2$, 10 mM $CaCl_2$ and 1 mg/mL BSA. Analytes were dissolved at 10 µg/mL in 100 µL of TBS containing 1 mg/mL BSA supplemented with 2 mM $MgCl_2$ and 10 mM $CaCl_2$, or 5 mM EDTA . Analytes

included GST-fused HgL VWD and the positive control A-domain fusion protein (gift from Samir Hamaia). Plates were incubated for 1 h at 20°C, then washed three times as above. Horseradish peroxidase-linked anti-GST antibody (#RPN1236, GE Healthcare) was incubated at 1:500 in TBS for 1 h. The reaction was developed by addition of 100 µL per well of TMB substrate (Pierce) and quenched by addition of an equal volume of 0.2 M sulfuric acid. Plates were read at 450 nm with an Envision multiplate reader.

3.3.4.3 Size-exclusion chromatography multi-angle light scattering of the recombinant lectin heterodimer

Size-exclusion chromatography multi-angle light scattering (SEC-MALS) experiments were performed on a SEC column (Superdex 200 HR10/30, GE Healthcare) and signal was detected using a multi-angle light scattering detector (DAWN HELEOS-II-8, Wyatt Technology) and a differential refractometer (Optilabr EX, Wyatt Technology). Recombinant HgL_19:LgL1_05, HgL_05 and HgL_09 were injected at 1 mg/mL on column equilibrated with 20 mM HEPES pH 7.4, 150 mM NaCl. Molecular mass were estimated using ASTRA 6.1 software package (Wyatt Technology).

3.3.4.4 SAXS analysis of constructs of the heavy chain (HgL)

In SAXS, the scattering pattern obtained for a solution of protein particles is related to the shape and size of these particles. Useful information such as molecular weight or radius of gyration (R_g) can be obtained from scattering patterns. R_g gives an estimation of the overall size of the protein. A globular compact protein is characterised by a smaller R_g when compared to an extended protein of the same size. The measured scattering intensity is also related to the distance distribution function $p(r)$, which describes the distances between all pairs of atoms within the protein. This function provides information about shape of scattering particles. For globular particles, the distance distribution is a single bell-shaped peak. Conversely, for an unfolded protein, $p(r)$ shows a longer tail towards high r values.

Similarly, globularity and flexibilibity can be compared between different proteins from normalised Kratky plots, $(sR_g)^2 I(s)/I(0)$ vs sR_g, independently of their size. Typically, single domain globular proteins exhibit a bell-shaped curved with a maximum

of 1.104 for a sR_g value of 1.73, whereas unfolded proteins display a rising curve with no local maximum. Multidomain proteins are hybrid of both cases.

SAXS data were collected in SEC-SAXS mode on beamline B21 (Diamond Light Source, UK) with an EIGER 4 M detector. Before the experiment, samples were buffer-exchanged by dialysis into the same PBS buffer used during data collection. Samples were concentrated to 5 mg/mL in 50 µL and injected on a SEC column (Superdex Increase S200 3.2/300, GE Healthcare) pre-equilibrated in PBS using an Agilent 1200 HPLC system. Scattering data were collected as the sample eluted from the column and analysed using the ATSAS suite [343]. In particular, sample selection and buffer subtractions were performed in Chromix, and molecular weights, R_g, distance distributions and Kratky plots were calculated in Primus.

3.3.5 X-ray crystallography of recombinant proteins

3.3.5.1 Crystallisation trials

For crystallisation, glycosylated, deglycosylated, methylated proteins were concentrated to 5-50 mg/mL and subjected to sitting drop vapour diffusion using a Mosquito (TTP Labtech) and commercial crystallisation screens (Hampton Research, Molecular Dimensions). For each trial, 100 nL of protein was mixed with 100 nL of well solution in a SWISSCI 96-well MRC-2 lens plate for sitting drop vapour diffusion at 4°C.

All crystallisation conditions tested are listed in Appendix Table A.2.

3.3.5.2 Crystallisation of HgL VWD, CP33 scFv and Fab

HgL VWD at 15 mg/mL yielded crystals, which were harvested after 37 days at 4°C in 0.2 M $MgCl_2$, 31% polyethylene glycol (PEG) 2000, 0.1 M sodium cacodylate, pH 6.5 (Memgold condition E7). CP33 scFv at 5 mg/mL formed crystals, which were harvested after 15 days at 4°C in 0.1 M CHES, 20% v/v PEG 8000, pH 9.5 (JCSG+ condition A7). CP33 Fab crystallised at 9 mg/mL protein, and crystals were harvested after 5 days at 4°C in 15% PEG 3350 and 0.15 M ammonium citrate dibasic in the presence of a seeds stock obtained from incubation of 9 mg/mL protein in 20% PEG 3350 and 0.2 M ammonium citrate dibasic (JCSG+ condition A3).

All crystals were cryo-protected in solutions identical to that of the mother liquor, supplemented with 25% glycerol, and flash-frozen in liquid nitrogen.

3.3.5.3 Data collection and structure determination

X-ray diffraction data were collected on beamlines IO4 for HgL VWD and scFv CP33 and I24 for Fab CP33 (Diamond Light Source, UK). Before data collection, optimal crystal orientation, exposure time, crystal rotation and oscillation degree were determined using EDNA and Mosflm.

Diffraction data were indexed and integrated using xia2-dials [344, 345], before scaling and merging with AIMLESS in CCP4i2 suite [346, 347]. 5% of randomly selected diffraction data were used for R_{free} determination [348]. Phaser was used for molecular replacement (MR) [349]. For CP33 scFv, a polyalanine model of a scFv (from Thomas Harrison, University of Oxford, unpublished) was used for MR, and one copy was found in the asymmetric unit of the crystal. Using the MR solution, automatic building was performed in Bucaneer [350]. For CP33 Fab, a composite model of CP33 scFv and the constant region of PDB code 3DGG was used for MR, and one copy was found in the asymmetric unit.

For HgL VWD, MR was performed using homology models and ensembles of VWDs but yielded only poor solutions in which a few β-strands could be assigned. Since this domain was one of the only regions of the whole antigen predicted to contain helical features, we could identify the corresponding region in the cryo-EM reconstruction. In Chimera [351], the homology model was fitted there and a surface map was generated at a resolution of 20 Å around the model. The overlapping cryo-EM density was extracted and used as an input in MR, and one copy was found in the asymmetric unit.

Solutions were visualised and edited in Coot [352] and further refined in Phenix [353]. Models were validated with MolProbity [354].

	CP33 Fab	CP33 scFv	HgL VWD
Data collection			
Wavelength (Å)	1.00	0.98	0.98
Resolution (Å)	58.39-2.75 (2.90-2.75)	60.77-1.37 (1.39-1.37)	51.67-3.31 (3.58-3.31)
Space group	C 2 2 2_1	P 6_1 2 2	I 4_1 2 2
Cell dimensions			
a, b, c (Å)	80.31, 163.29, 90.14	70.17, 70.17, 212.17	121.22, 121.22, 129.55
α,β,γ (°)	90, 90, 90	90, 90, 120	90, 90, 90
Unique reflections	27878 (14807)	121112 (64760)	13710 (7472)
Multiplicity	1.9 (1.9)	1.9 (1.9)	1.8 (1.9)
Completeness (%)	99.7 (97.8)	98.6 (96.3)	99.9 (99.9)
R_{pim} (%)	19.6 (68.4)	3.5 (77.3)	13.1 (82.8)
$CC_{1/2}$ (%)	91.4 (29.3)	99.9 (33.1)	0.986 (39.3)
Average I/σ(I)	1.2 (0.3)	8.6 (0.5)	4.8 (0.9)
Refinement			
R_{work}/R_{free}	20.39/25.26	19.39/21.22	28.36/32.65
RMSD			
Bond length (Å)	0.00	0.01	
Bond angles (°)	0.63	1.19	
Validation			
Poor rotamers (%)	0.00	0.45	0.00
Clashscore	6.29	4.70	
Ramachandran plot			
Favoured (%)	94.16	97.01	84.08
Allowed (%)	5.61	2.99	12.10
Disallowed (%)	0.23	0.00	3.82

Table 3.1: Current crystallographic statistics. Refinement of HgL VWD is on-going.

3.3.6 Cryo-EM of the 260-kDa and 150-kDa antigens

3.3.6.1 Cryo-EM of the native 260-kDa antigen

3.3.6.1.1 Sample modification and grid preparation

After SEC in 20 mM HEPES pH 7.4, 150 mM NaCl, the 260-kDa native complex was either untreated or PEGylated before application on grid. PEGylation was used to reduce denaturation and aggregation at the air-water interface. Samples were PEGylated at 0.5 mg/mL by incubation with methyl-PEG-NHS-esters $MS(PEG)_4$, $MS(PEG)_{12}$ or $MS(PEG)_{24}$ (#22341, #22685, #22687, ThermoFisher Scientific) with a 40 molar excess of PEG reagent for 2 h on ice. This ensured PEGylation of a maximum of one third of the total lysines (118), thereby minimising change in stability. Reactions were quenched by addition of 50 mM Tris pH 7.5, 100 mM NaCl. PEGylation was also carried out

with $MS(PEG)_{12}$ in the presence of 20 mM GalNAc in all buffers.

Untreated and PEGylated samples were diluted to 0.11-0.15 mg/mL and applied on carbon-coated holey grids (Quantifoil, R1.2/1.3, 300 gold mesh) after plasma cleaning with a Quorum device (#Q150R ES, Quorum Technologies) at 50 mA for 60 s. Grids were blotted for 1.5 s with a blot force of -5 at 100% humidity and plunge frozen in liquid ethane using a Vitrobot (FEI). Untreated samples concentrated at 3.3 mg/mL in 0.1% fluorinated octylmaltoside were applied on same grids and blotted for 1 s with a blot force of -25 at 100% humidity.

Untreated samples were also applied at 0.02 mg/mL on R2/4 grids with a suspended monolayer of graphene (#AGS179-GO, Agar Scientific), previously glow-discharged at 20 mA for 10 s. These grids were blotted for 5 s with a blot force of -15 at 100% humidity.

3.3.6.1.2 Data acquisition

Cryo-EM data of native 260-kDa antigen either untreated, bound to CP33 Fab, PEGylated with $MS(PEG)_4$, $MS(PEG)_{12}$, $MS(PEG)_{24}$ or applied on a graphene monolayer were collected on a Titan Krios transmission electron microscope (FEI) operating at 300 kV. Data were recorded either on a K2 or K3 detector under low dose conditions with a nominal pixel size of 0.822 Å/pixel or 0.832 Å/pixel, and a total exposure of approximately 50 or 60 electrons per $Å^2$, respectively. Conditions are summarised in Table 3.2.

Cryo-EM data of the first SEC fraction of the 260-kDa antigen, of sample extracted with DIBMA, and of sample diluted in fluorinated detergents were collected on a Talos Arctica microscope (FEI) operating at 200 kV. 400 movies of each sample were recorded on a Falcon 4 detector, with a pixel size of 1.192 Å/pixel and a total exposure of approximately 45 electrons per $Å^2$.

	Untreated	CP33-bound	Graphene	
Voltage (kV)	300	300	300	
Pixel size (Å)	0.822	0.822	0.832	
Detector	K2	K2	K3	
Total dose ($e^-/Å^2$)	48	49	60	
Number of frames	32	32	40	
Defocus range (µm)	-0.7 to -2.2	-0.7 to -2.2	-1 to -2.5	
Grid	Au 300 R1.2/1.3	Au 300 R1.2/1.3	Au 300 R2/4 graphene	
Number of movies	1480	1427	720	
	PEG_4	PEG_{12}	PEG_{24}	PEG_{12} + GalNAc
Voltage (kV)	300	300	300	300
Pixel size (Å)	0.832	0.822	0.832	0.832
Detector	K3	K2	K3	K3
Total dose ($e^-/Å^2$)	60	49	60	60
Number of frames	40	32	40	40
Defocus range (µm)	-0.7 to -2.2	-0.7 to -2.2	-0.7 to -2.2	-0.7 to -2.2
Grid	Au 300 R1.2/1.3	Au 300 R1.2/1.3	Au 300 R1.2/1.3	Au 300 R1.2/1.3
Number of movies	580	2380	420	12806

Table 3.2: Cryo-EM settings for imaging of the 260-kDa antigen.

3.3.6.1.3 Data processing

For each dataset, motion correction, CTF estimation, particle picking, particle extraction and 2D classifications were done in SIMPLE 3.0 [355]. Best 2D class averages were selected and exported into RELION 3.1 for further processing [356].

Because of preferred orientation, datasets collected at 300 kV were combined. Particles were subjected to a new round of 2D classification and re-centering during extraction with a box size of 256 × 256 pixels. Selected particles were used to generate a 3D ab-initio model, which was filtered to 15 Å and used as a reference map for 3D classification into three classes. The best class was refined and further classified into 10 classes without alignment. The best class was refined to a resolution of 2.96 Å after CTF refinement and post-processing. Sharpening was also performed in DeepEMhancer [357] to guide structure building, and local resolution estimates were determined in RELION 3.1.

Processing of PEGylated sample in the presence of GalNAc was carried out similarly and yielded a map at 3.07 Å. In addition, focused classification of the protrusion was

performed with a regularisation parameter (T) of 20 after particle subtraction, followed by 3D refinement. Alternatively, particles were extracted in Cryosparc [358] with box sizes of 512 × 512 or 720 × 720 pixels and subjected to 2D classifications, 3D ab-initio modelling, non uniform and local refinements [359].

3.3.6.1.4 Preliminary structure determination

The crystal structure of HgL VWD was fitted into the cryo-EM reconstruction using Chimera [351] and used as a starting point in structure building in Coot [352]. From both N-terminal and C-terminal ends, polyalanine backbones were built into the EM density.

At the N-terminus (residues 16 to 210), the presence of β-strands could be observed and strands were placed there. The partial polyalanine model of the N-terminus was used as an input in DALI server [360], which determines structure similarities regardless of protein sequence. The server revealed similarity to proteins adopting a lipoprotein-like fold. The conserved topology observed in these domains was used as a guide to unambiguously confirm or establish connections between each β-strand. Last, sequence was assigned based on secondary structure prediction using PROMALS3D server [361] and β-turn prediction from BetaTPred3 [362].

At the C-terminus of HgL VWD (residues 389 to 586), a polyalanine backbone was built, but DALI failed to identify similar 3D structures. Models of pseudo-domains predicted by Alphafold [336] were aligned to the polyalanine backbone and used to guide sequence assignment. The rest of HgL C-terminus could not be built with high confidence due to poorer resolution estimates and deviation far from the core of the reconstruction.

The globular EM density adjacent to the partial model of HgL was attributed to LgL. There, the β-trefoil domain of ricin-type agglutinin (pdb code 4G9N) was fitted in the EM density. Sequence from the domain was assigned based on secondary structure prediction in PROMALS3D and identification of QXW and QXF β-trefoil motifs in LgL sequence. The rest of the protein, consisting of about 60 residues N-terminal to the β-trefoil domain, was built manually and partially. Cysteine 75 was located in close proximity to cysteine 415 from HgL, indicative of a disulfide bond.

The sequence assigned is that of HgL isoform EHI_133900 and LgL isoform EHI_035690 based on their identification with highest scores obtained by MS (Appendix Table

A.3), and because HgL isoform EHI_133900 failed to co-express with LgL isoforms EHI_148790 nor EHI_183400.

3.3.6.2 Cryo-EM of the cysteine-rich region of the 260-kDa antigen

3.3.6.2.1 Grid preparation and sample modification

After SEC in 20 mM HEPES pH 7.4, 150 mM NaCl, recombinant HgL_20 in complex with CP33 Fab was diluted to 0.05 mg/mL, applied on holey grids as described in 3.3.6.1.1, and blotted for 0.5 s with a blot force of -5.

3.3.6.2.2 Data acquisition

Cryo-EM data were collected as described in section 3.3.6.1.2 and summarised in Table 3.3.

	CP33-bound HgL_20
Voltage (kV)	300
Pixel size(Å)	0.832
Detector	K3
Total dose($e^-/Å^2$)	60
Number of frames	40
Defocus range (μm)	-0.7 to -1.7
Grid	Au 300 R1.2/1.3
Number of movies	5520

Table 3.3: Cryo-EM settings for imaging of HgL_20.

3.3.6.2.3 Data processing

Motion correction was performed in SIMPLE 3.0 [355]. Integrated micrographs were imported into CryoSPARC [358] for further processing. There, particles were extracted with a box size of 200 × 200 pixels and subjected to several rounds of 2D classifications. The best 2D classes were used in 3D ab-initio reconstruction into three classes, followed by heterogenous refinement and homogenous refinement. Particles that gave the best maps were used in non-uniform refinement [359] to yield reconstructions at 3.40 Å.

3.3.6.3 Cryo-EM of IgL

3.3.6.3.1 Grid preparation and sample modification

After SEC in 20 mM HEPES pH 7.4, 150 mM NaCl, recombinant IgL was either untreated, PEGylated, or methylated before application on grids. IgL complexed with Fabs XEhI_H2 and XEhI_X20 was also applied on grid without any modification. PEGylation was carried out with a 30 molar excess of $MS(PEG)_{12}$, as described in section 3.3.6.1.1. Methylation was carried out as described in section 3.3.2.5. Samples diluted to 0.10-0.13 mg/mL were applied on holey grids as described in 3.3.6.1.1, and blotted for 2.5 s with a blot force of -5.

3.3.6.3.2 Data acquisition

Cryo-EM data were collected as described in section 3.3.6.1.2 and summarised in table 3.4.

	Untreated	Untreated	Fab-bound
Voltage (kV)	300	300	300
Pixel size (Å)	0.822	0.822	0.822
Detector	K2	K2	K2
Total dose ($e^-/Å^2$)	48	48	49
Number of frames	32	32	32
Defocus range (µm)	-0.7 to -2.2	-0.7 to -1.5	-0.7 to -2.2
Grid	Au 300 R1.2/1.3	Au 300 R1.2/1.3	Au 300 R1.2/1.3
Number of movies	1320	6411	3594
	Methylated	PEG_{12}	
Voltage (kV)	300	300	
Pixel size (Å)	0.822	0.822	
Detector	K2	K2	
Total dose ($e^-/Å^2$)	49	49	
Number of frames	32	32	
Defocus range (µm)	-1.0 to -2.5	-1.0 to -2.5	
Grid	Au 300 R1.2/1.3	Au 300 R1.2/1.3	
Number of movies	533	1200	

Table 3.4: Cryo-EM settings for imaging of IgL.

3.3.6.3.3 Data processing

Each dataset was processed as described in section 3.3.6.2.3. The best angular distribution and 3D reconstruction was obtained after non-uniform refinement of

PEGylated IgL, at an overall resolution of 4.53 Å.

Particles of IgL from all datasets recorded with a K2 detector were combined after 2D classifications for heterogenous refinement using the reference map of PEGylated IgL, non-uniform refinement and CTF refinement. This yielded a final 3D reconstruction at 3.01 Å. The same pipeline was used for collection of Fab-bound IgL, to generate a reconstruction at 2.85 Å.

3.4 Results

3.4.1 Recombinant expression of fragments of the amoebic antigen and antibodies for X-ray crystallography

3.4.1.1 Structural homology modelling, bioinformatic analyses and construct design

Initially, recombinant extracellular fragments of the amoebic antigen were produced for crystallisation trials. A set of bioinformatic tools were used to predict domains of HgL (EHI_133900), IgL (EHI_006980) and LgLs (isoforms EHI_035690, EHI_148790 and EHI_183400) and guide construct design.

3.4.1.1.1 Analysis of HgL

HgL contains an N-terminal signal peptide (residues 1 to 15), an extracellular domain (residues 16 to 1232), a C-terminal transmembrane domain (residues 1233 to 1255) and a small cytoplasmic domain (residues 1256 to 1293).

Poor homology modelling is observed at the N-terminus (residues 16 to 210) and the cysteine-rich domain (residues 395 to 1232). Between both domains, a VWD (residues 211-388) is predicted with high confidence (Figure 3.8.a). Although residues 892 to 998 contain the region identified as a CRD, the structure there is the most poorly predicted. EGF-like domains are also predicted with low confidence for most of the C-terminal half of HgL, but none of the signature motifs found in all types of EGF-like domains (prosite PS01186, EGF-like type 2; PS00022, EGF-like type 1; PS01248, laminin type; PS01187, calcium-binding type [363]) can be found in the sequence of HgL. These predictions were combined with protein disorder prediction to design constructs suitable for X-ray crystallography (Figure 3.8.b).

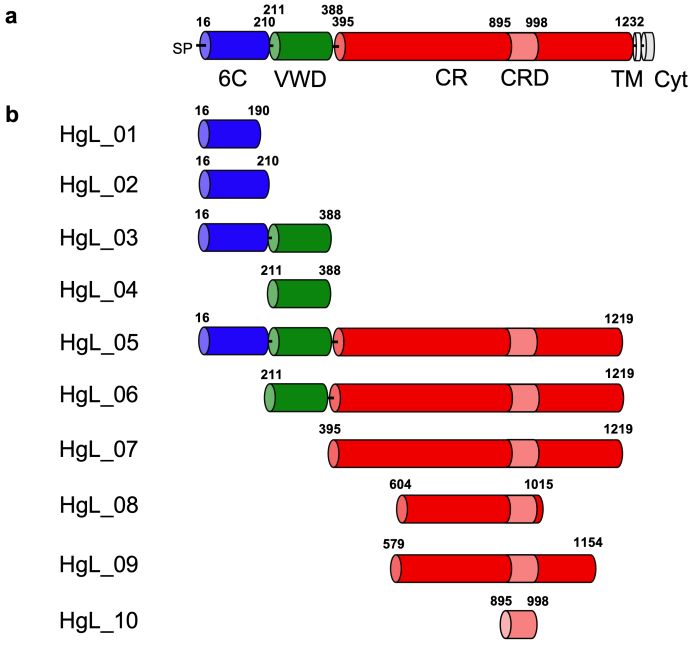


Figure 3.8: Constructs of HgL for recombinant expression. a) Schematic representation of the domains of HgL predicted with bioinformatic tools. These include a small N-terminal domain of six cysteines (6C, in blue), a putative VWD (in green) and the cysteine-rich region (CR, in red), which contains the putative CRD (in pink). The extracellular domain of HgL is flanked by a small transmembrane domain (TM, in white) and a cytoplasmic tail (Cyt, in grey). b) Schematic representation of HgL constructs, with colours matching those in a).

Constructs of the N-terminal domain of 6 cysteines (6C, HgL_01 and HgL_02), of the VWD (HgL_04) and of both domains (HgL_03) were designed. In addition, constructs of the full extracellular portion of HgL (HgL_05), and of fragments of the cysteine-rich region (CR) around the putative CRD (HgL_06-10) were designed, among which the vaccine candidate LecA (HgL_09) was added.

3.4.1.1.2 Analysis of LgL

LgL, contains an N-terminal signal peptide (residues 1 to 15 for EHI_035690, 1 to 18 for EHI_148790, 1 to 17 for EHI_183400). As opposed to a previous study, LgLs are not predicted to contain any C-terminal GPI anchor [266]. Interestingly, LgL isoforms seem to share structural features with ricin-type β-trefoil domains (Figure 3.9.a).

β-trefoil domains consist of two sets of six β-strands which form a cap and a barrel, respectively (Figure 3.9.b). They also contain a conserved sequence motifs QXW repeated three times in β-strands 4, 8 and 12. Similarly, in all isoforms of LgL, 12 β-strands that may form the characteristic cap and barrel of this domain are predicted

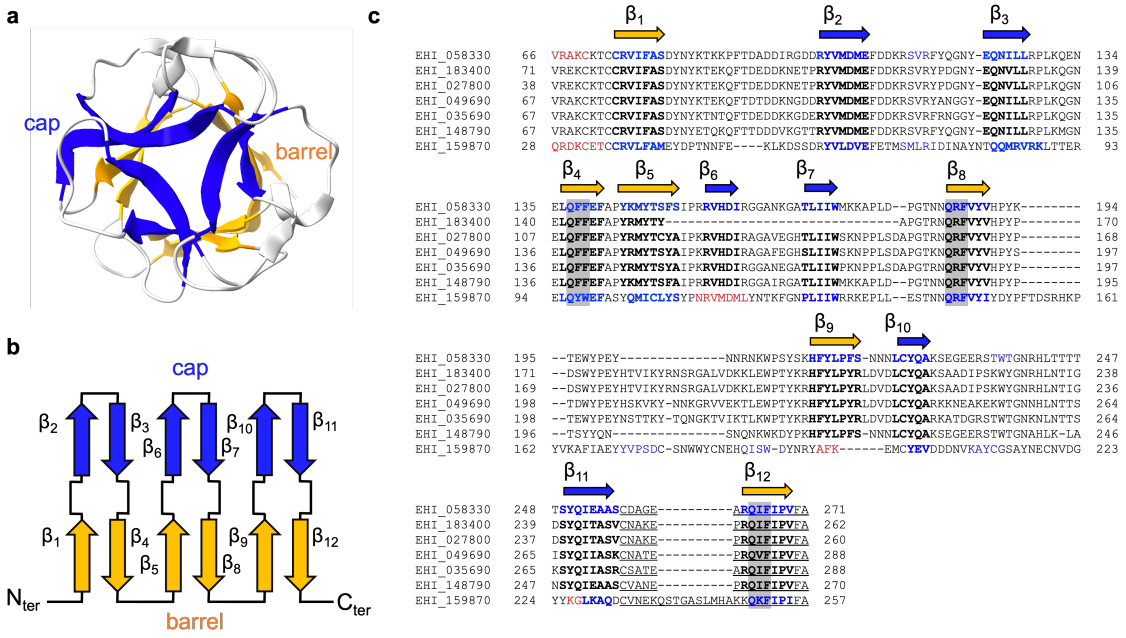


Figure 3.9: LgL shares homology with β-trefoil domains. a) Ribbon representation of PDB template 4G9N, a ricin-type β-trefoil lectin which shares partial homology with LgL. β-strands of the cap (in blue) and barrel (in orange) subdomains are highlighted. The same colour code is used in b) and c). b) Schematic representation of the topology of a β-trefoil domain, with two halves of six β-strands forming the cap and the barrel in a pseudo three-fold symmetry. c) Sequence alignment of LgLs based on secondary structure prediction using PROMALS3D. Residues predicted to form helices (in red) and β-strands (in blue), QXW and QXF motifs (in grey) are highlighted. Sequences matching the reported GPI anchor are underlined.

(Figure 3.9.c). In addition, LgLs contain three QX(F/W) motifs located in the predicted β-strands 4, 8 and 12. Intriguingly, the location of the last β-strand of the predicted β-trefoil domain overlaps with that of the reported GPI anchor.

Besides, the structure formed by residues N-terminal to the putative β-trefoil domain is poorly predicted. For each of the three isoforms of LgL studied, constructs of the restricted β-trefoil domain and full extracellular proteins, with and without the reported GPI anchor domain, were designed and included all cysteines.

3.4.1.1.3 Analysis of IgL

IgL contains an N-terminal signal peptide (residues 1 to 13) and may be GPI-anchored at its C-terminus (Figure 3.10.a). Remarkably, the full extracellular protein is rich in cysteines, with 45 CXXC motifs (13 from residues 41 to 398, 4 from residues 525 to 578, 10 from residues 636 to 778 and 18 from residues 820 to 1082) and sequence signature

motifs of EGF-like domains found across the sequence. In all cases, structural homology tools failed to predict domains with sensible cysteine positions.

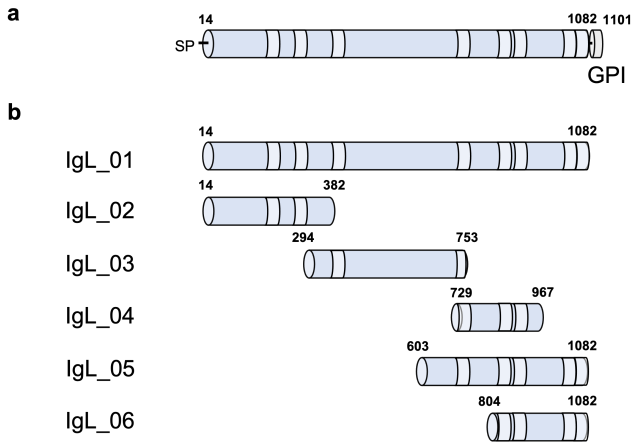


Figure 3.10: Constructs of IgL for recombinant expression. a) Schematic representation of the domains of IgL predicted with bioinformatic tools. EGF-like domain signature motifs (in blue) are found between residues 199 to 245, 275 to 315, 379 to 410, 733 to 776, 847 to 882, 890 to 930, 1014 to 1050 and 1051 to 1083. A GPI anchor (in light grey) is predicted at the C-terminus. b) Schematic representation of HgL constructs, with colours matching those in a).

For structural studies, a construct of the full extracellular portion (IgL_01) was designed (Figure 3.10.b). In addition, several fragments of the extracellular domain were designed based on studies of hemagglutinating, hemolytic and cytotoxic activities of IgL [297]. Among these, the C-terminal domain (IgL_05) is a vaccine candidate with most hemolytic activity, in which two fragments (IgL_04 and IgL_06) appear to contain six of the eight CX_8C motifs of IgL, which may be responsible for hemolysis. Moreover, antibodies against the middle (IgL_03) and C-terminal domains inhibit trophozoite adherence *in vitro*, as opposed to antibodies against the N-terminus (IgL_01).

3.4.1.2 Protein production and modification of amoebic constructs

3.4.1.2.1 Protein purification in Expi293F cells

Most constructs were expressed transiently in Expi293F cells, which are optimised for production and secretion of extracellular proteins. Importantly, these support formation of disulfide bonds and glycosylations. Since the amoebic antigen is rich in cysteines, and because *Entamoeba histolytica* contains genes of oligosaccharyltransferases

involved in N-glycosylation, this system seems more relevant than bacterial systems used in most studies.

Screening of protein expression by Western blot showed expression for all constructs, except LgLs and the region defined as a CRD (HgL_10). Recombinant proteins secreted in the medium were purified by IMAC and SEC. Notably, a poor yield was obtained for the full cysteine-rich region of HgL (HgL_07) and the middle portion of IgL (IgL_03). Moreover, some contaminants were present at the end of purification of two constructs (HgL_06 and HgL_07).

All other constructs were successfully expressed and purified in milligram quantities suitable for crystallisation trials (Figure 3.11.a,b). Various conditions were tested to induce crystal formation (see Appendix Table A.2). Trials were performed with glycosylated and deglycosylated proteins, and, in some cases, lysine methylation was carried out to alter surface properties and solubility (Figure 3.11.c,d). Of these, crystals of deglycosylated HgL VWD (HgL_04) were obtained, whereas other constructs failed to crystallise.

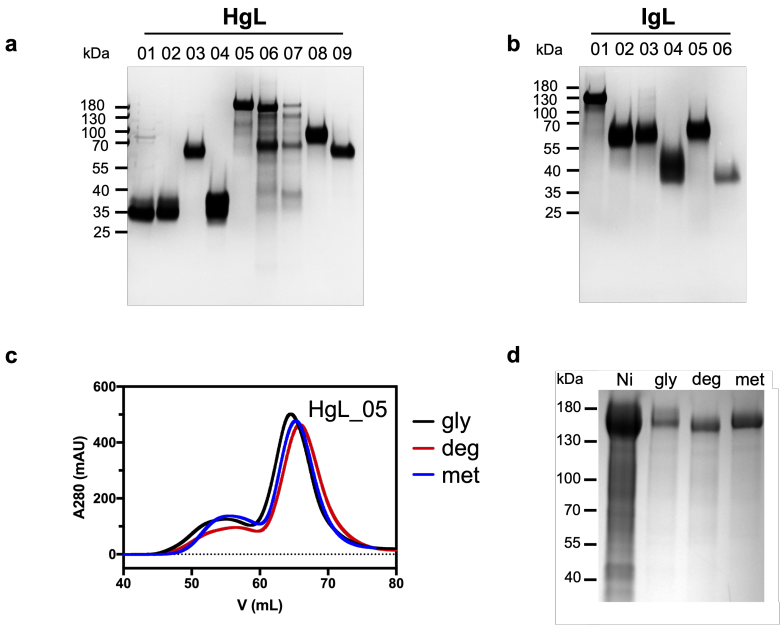


Figure 3.11: Protein production in Expi293F cells and modification. a) Reducing SDS-PAGE gel of HgL constructs expressed in Expi293F cells. b) Reducing SDS-PAGE gel of IgL constructs expressed in Expi293F cells. c) Overlay of SEC (S200 HiLoad 16/600) curves of HgL_05 glycosylated (gly, in black), deglycosylated (deg, in red) and methylated (met, after deglycosylation, in blue). The void volume is around 55 mL, while the elution volume for a typical IgG (158 kDa) is around 65 mL. The expected molecular weight of deglycosylated HgL_05 is around 136 kDa. d) Reducing SDS-PAGE of full extracellular HgL after IMAC (Ni), SEC (gly), deglycosylation (deg) and methylation (met).

3.4.1.2.2 New construct design based on limited proteolysis

To identify stable intermediates that may be more suitable for crystallisation, full extracellular HgL (HgL_05) was used in limited proteolysis experiments with a variety of proteases. In small scale experiments, multiple intermediates were formed over time in the presence of chymotrypsin. Given the close proximity of bands on gel, intermediates could not be isolated without cross-contamination. Therefore, digestions were scaled up and the proteolysed reaction applied on SEC to minimise cross-contamination (Figure 3.12.a).

Two broad peaks were obtained and bands for four intermediates (1 to 4) were excised for in-gel tryptic digest and MS (Figure 3.12.b,e). Similar migration patterns in non-reducing conditions suggest the absence of covalent interactions between intermediates (Figure 3.12.c). Nonetheless, intermediate 1 co-purifies with species of higher molecular weight and may therefore interact non-covalently with other parts of HgL.

Likewise, limited proteolysis was carried out for the most restricted construct (HgL_08) surrounding the region defined as a CRD, and an intermediate (5) was

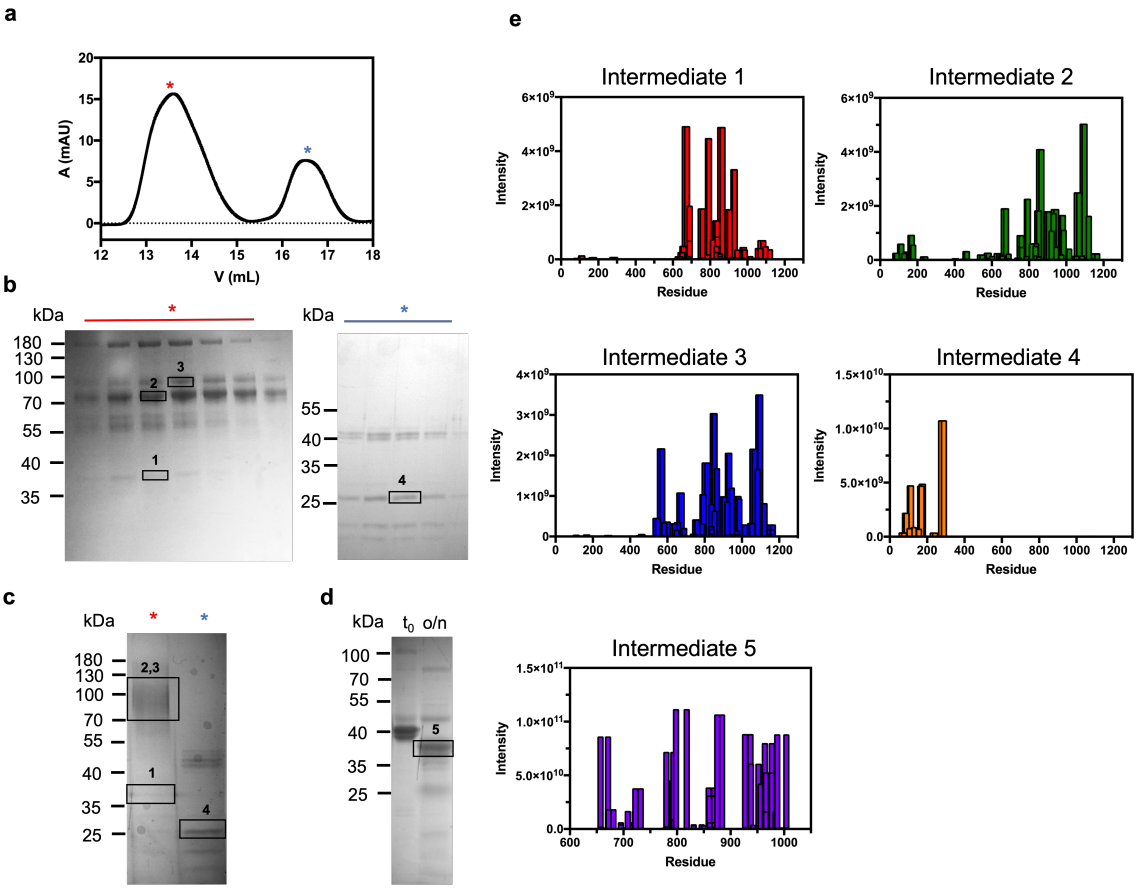


Figure 3.12: Limited proteolysis of HgL_05 and HgL_08 coupled with MS. a) SEC profile (S200 Increase 10/300 GL) of HgL_05 after overnight digest with chymotrypsin. Two peaks are obtained (red and blue asterisks). b) Reducing SDS-PAGE of the SEC fractions of both peaks shown in a). Intermediates 1 to 4 (black boxes) are highlighted. c) Non-reducing SDS-PAGE of fractions of highest absorbance of both peaks shown in a). Migration patterns are similar to that shown in b). d) Reducing SDS-PAGE of digest of HgL_08 by GluC. Samples before (t_0) and after overnight incubation (o/n) with GluC were loaded. e) Residue intensity plots obtained by MS analysis of in-gel tryptic digests of intermediates 1 to 5.

stable after overnight incubation with GluC.

Based on intensities obtained by MS (Figure 3.12.e), tryptic digests of intermediates 1, 2 and 3 isolated from the first SEC peak lacked residues N-terminal to positions 620, 664 and 560, respectively. For these fragments, the most C-terminal peptides of high intensities were found in positions 932 and 1121. Likewise, for the tryptic digest of intermediate 5, most proteolysis occured at the N-terminus of the construct, and peptides were identified between residues 657 and 1004. Notably, all four intermediates share a stable core that consists of residues 664 to 932.

Conversely, the tryptic digest of intermediate 4 present in the second SEC peak

lacked all peptides C-terminal to residue 290. This suggests that proteolysis consistently occurs between N-terminal and C-terminal halves. In particular, the 300 N-terminal residues of HgL appear to act independently to the 600 C-terminal residues, and residues between these two regions are most vulnerable to proteolysis.

Although this method allows identification of stable components, there are several limitations that can hamper estimation of new boundaries for construct design. First, peptides with glycans are not identified by MS. However, in HgL, N-glycans are predicted on residues 642, 659, 890, 992, 1138, 1204 and 1214, some of which may be included in stable fragments but not detected due to mass differences. In addition, before MS, samples are digested by trypsin, therefore some end peptides may not be identified in the MS analysis. Last, molecular weights cannot be estimated with precision from a gel but are ideally determined by intact MS.

With these limitations in mind, several constructs were designed based on approximate boundaries, molecular weight estimation, and disorder predictions (Figure 3.13.a). Based on peptide intensities, intermediate 4 contains residues 63 to 290. All six cysteines found at the N-terminus are included in this fragment. However, residue 290 is in the middle of HgL VWD. Therefore no more restricted domains of this region were designed. Instead, a longer construct of the N-terminus was designed, starting at residue 63 and excluding peptides identified in intermediates 1, 2, 3 and 5. All other constructs are fragments of the cysteine-rich region that include the putative CRD and peptides identified in intermediates 1,2,3 and 5.

Although all constructs were expressed successfully in Expi293F cells (Figure 3.13.b), they failed to crystallise. Among these, HgL_16 formed covalent dimers on SEC and non-reducing gel, suggesting the presence of reactive cysteines, while others were monomeric in solution.

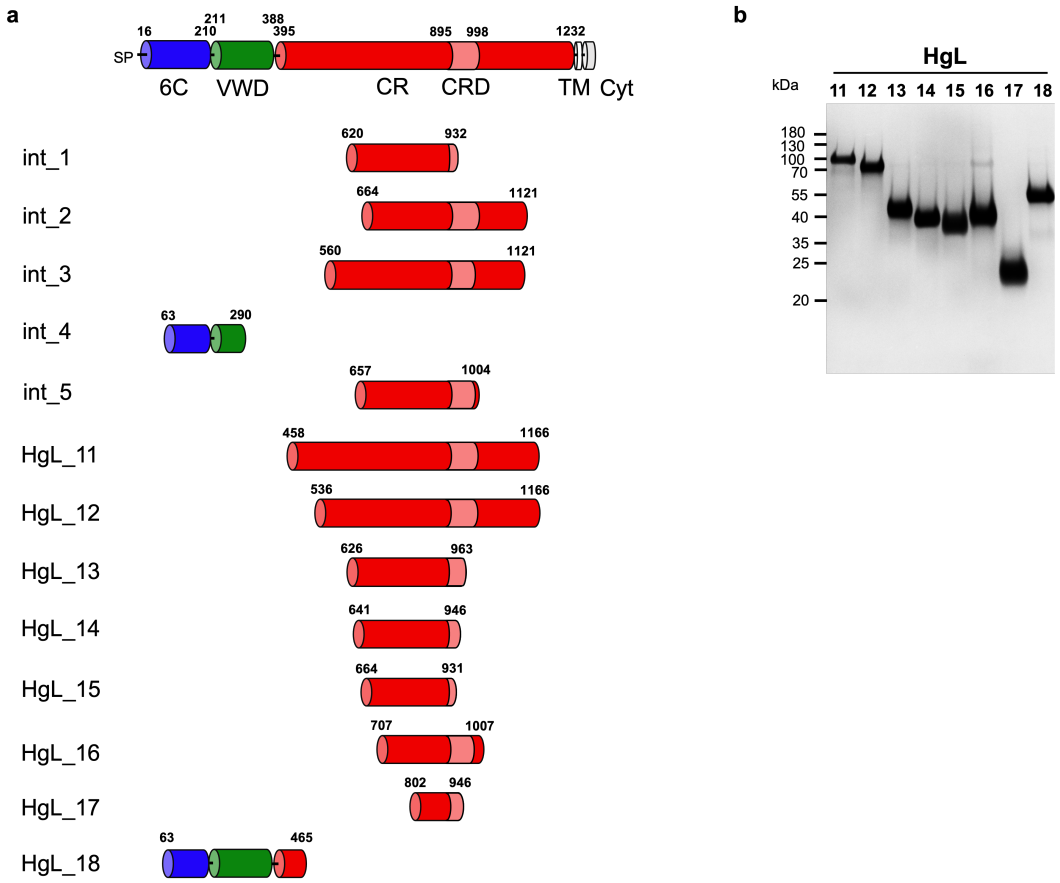


Figure 3.13: New constructs based on limited proteolysis. a) Schematic representation of minimum size intermediates (int_1 to int_5) obtained by MS after limited proteolysis and newly designed constructs (HgL_11 to HgL_18). b) Reducing SDS-PAGE of new constructs.

3.4.1.2.3 Bacterial refolding of LgL and HgL putative CRD

Since constructs of LgL and HgL putative CRD failed to express in mammalian cells, their refolding was investigated. The putative β-trefoil domains of three isoforms of LgL were enriched in inclusion bodies. After solubilisation in guanidine, on-column refolding only led to protein precipitation. Likewise, precipitation occured in all flash-dilution buffers, unless the buffer contained guanidine at 750 mM, which maintained partial denaturation (Figure 3.14.a). However, gentle removal of guanidine by dialysis, TFF or IMAC all led to protein precipitation.

Conversely, HgL putative CRD was successfully refolded from inclusion bodies using flash dilution. After purification by IMAC, cleavage of the N-terminal hexahistidine tag, and SEC, the protein was used in crystallisation trials (Figure 3.14.b,c). However, in

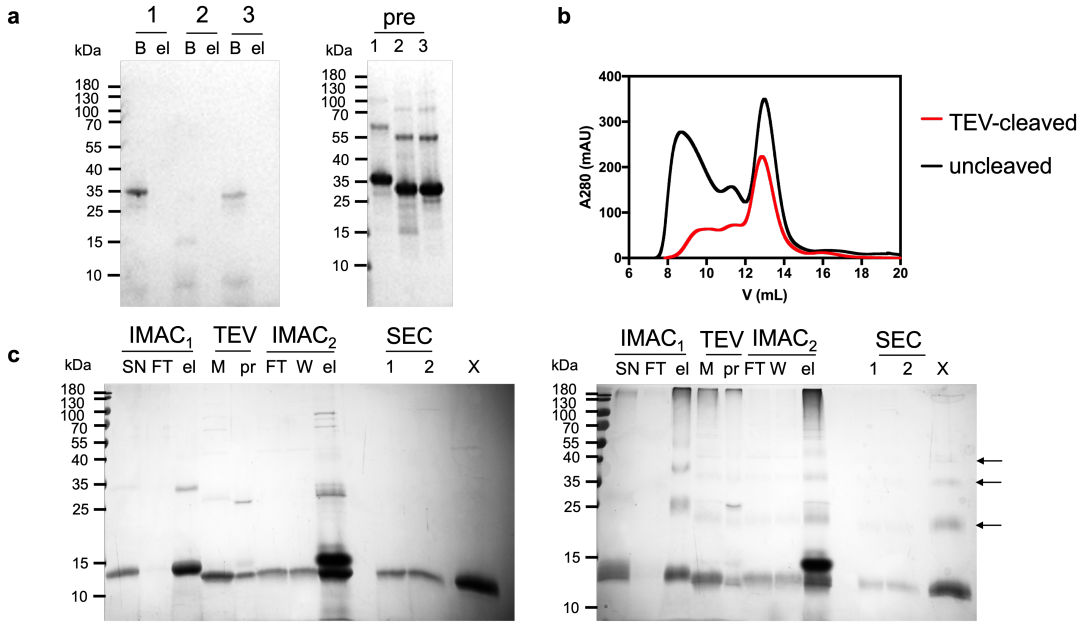


Figure 3.14: Refolding trials of HgL putative CRD and LgL predicted β-trefoil domains. a) Reducing SDS-PAGE of refolding trials of LgL β-trefoil domains (1 is EHI_035690, 2 is EHI_147980 and 3 is EHI_183400). Proteins in flash dilution buffer (B) and elution after IMAC (el) following removal of guanidine are shown on the left gel. Precipitates (pre) formed during guanidine removal by dialysis are shown on the right gel. b) SEC (S75 Increase 10/300GL) profile of refolded HgL putative CRD before (in black) and after cleavage of the affinity tag (in red). c) SDS-PAGE of protein fractions during refolding of HgL putative CRD in reduced (left gel) and non-reduced conditions (right gel). In a first IMAC ($IMAC_1$), solubilised inclusion bodies (SN) were loaded and the flow-through (FT) and elution (el) were recovered. The eluted protein was incubated with TEV protease and buffer exchanged overnight. Protein mix (M) and precipitates (pr) were loaded and a reverse IMAC ($IMAC_2$) allowed purification of cleaved CRD in flow-through (FT) and wash (W) fractions, while remaining proteins were in the elution (el, concentrated). After SEC, two fractions (1, 2) of the main peak were concentrated for crystallisation trials (X). Estimated molecular weights are 14.2 kDa for uncleaved CRD, 12.4 kDa for cleaved CRD and 27 kDa for TEV protease. Arrows on the right of the non-reduced gel show the location of higher order oligomers of uncleaved CRD

the presence and absence of GalNAc, the protein failed to yield crystals.

Importantly, in non-reducing conditions, higher order oligomers were observed on gel, indicating the presence of at least one reactive cysteine among the 10 cysteines present in the construct. This suggests that the minimal domain that defines a stable CRD remains to be determined, or that there are disulphide bonds linking the CRD to other domains.

3.4.1.3 Use of antibody fragments as chaperones for crystallisation

3.4.1.3.1 Purification of antibody fragments against HgL and IgL

Reported sequences of antibodies against HgL and IgL [280, 296, 338] were used to express antibody fragments, which can act as co-crystallisation chaperones (Figure

3.15.a) through stabilisation of protein targets and formation of crystal contacts. In addition to Fab antibody fragments, an scFv was designed by fusion of the variable heavy and light domains with a glycine-rich linker. These were expressed by Expi293F cells and purified by IMAC (Figure 3.15.b). CP33, an inhibitory antibody against HgL, was purified as Fab and scFv. In addition, three anti-IgL Fabs were purified, two of which inhibit trophozoite adherence to target cells (XEhI_H2 and XEhI_X20).

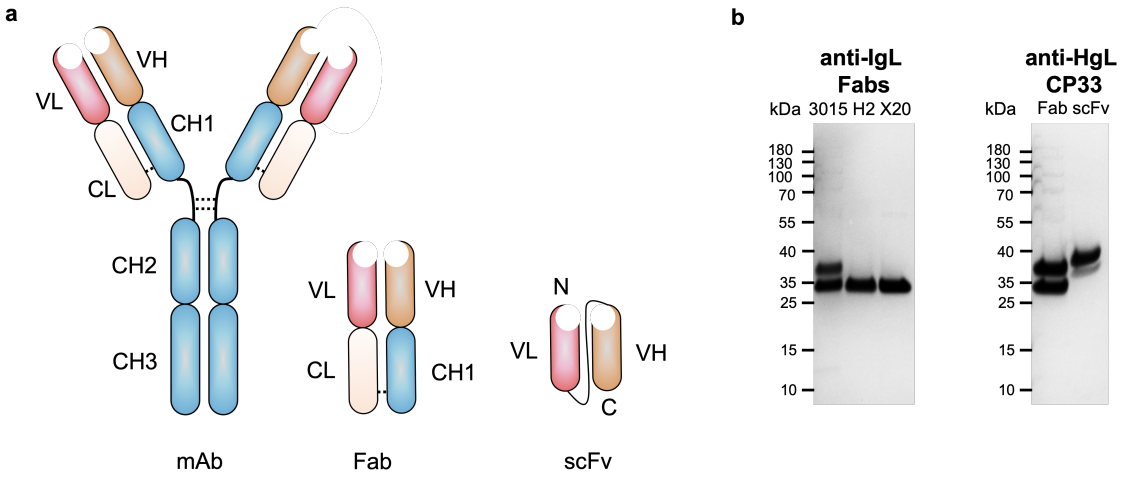


Figure 3.15: Purification of fragments of antibodies. a) Schematic representation of a mAb and its corresponding Fab and scFv fragments. b) Reducing SDS-PAGE of purified Fabs and scFv.

3.4.1.3.2 Antibody binding to recombinant constructs

Binding of Fabs to recombinant constructs was investigated by SPR. Intriguingly, IgL showed binding to immobilised XEhI_X20 and XEhI_H2, but failed to bind to EH3015. This was confirmed by SEC binding experiments, which showed that epitopes of both inhibitory Fabs were non-overlapping (shown later in Figure 3.34.b).

Besides, CP33 Fab bound to HgL_09 (LecA) with an estimated K_D of 4.9 nM, in agreement with the previously reported K_D of 7.73 nM against refolded LecA [280] (Figure 3.16.c). To further map the epitope of CP33, binding of several overlapping fragments of the cysteine-rich region was investigated (Figure 3.16.a). Interestingly, HgL_13 and HgL_16 bound to CP33, whereas no response was obtained for HgL_14 (Figure 3.16.b). This suggests that the stretch of 28 amino-acids between residues 946 and 973, which is contained in the region defined as a CRD, is essential for binding to the inhibitory antibody CP33 (Figure 3.16.d).

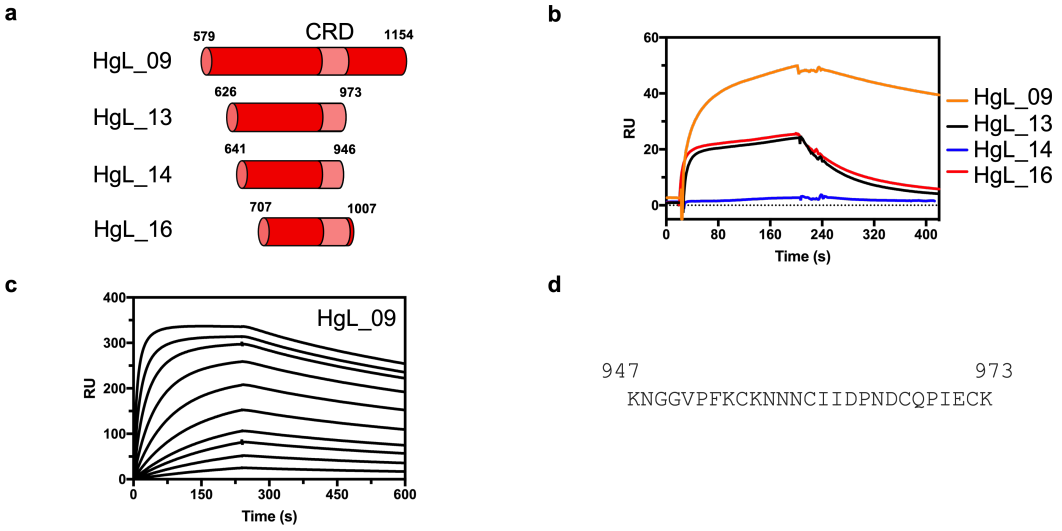


Figure 3.16: CP33 binding to HgL. a) HgL constructs screened for binding to CP33. b) SPR curves of constructs shown in a) flowed at 100 nM over CP33 Fab, immobilised on SPR chip. c) SPR curves for dilution series of HgL_09 (LecA) against immobilised biotinylated CP33 on streptavidin chip. HgL_09 concentration ranges from 0 to 500 nM. d) Sequence of HgL necessary for CP33 binding.

3.4.1.3.3 Structures of CP33 scFv and Fab

Structures of CP33 Fab and scFv were solved to provide phases in co-crystallisation experiments, since homology modelling of the cysteine-rich region is poor, or to potentially fit them in cryo-EM maps of antibody-bound amoebic antigen. Crystals of scFv and Fab were obtained and diffracted to 1.37 Å and 3.03 Å, respectively (Figure 3.17.a). The structure of the variable domain of CP33 Fab showed good agreement with that of CP33 scFv, with an overall RMSD of 0.490 Å (Figure 3.17.b,c).

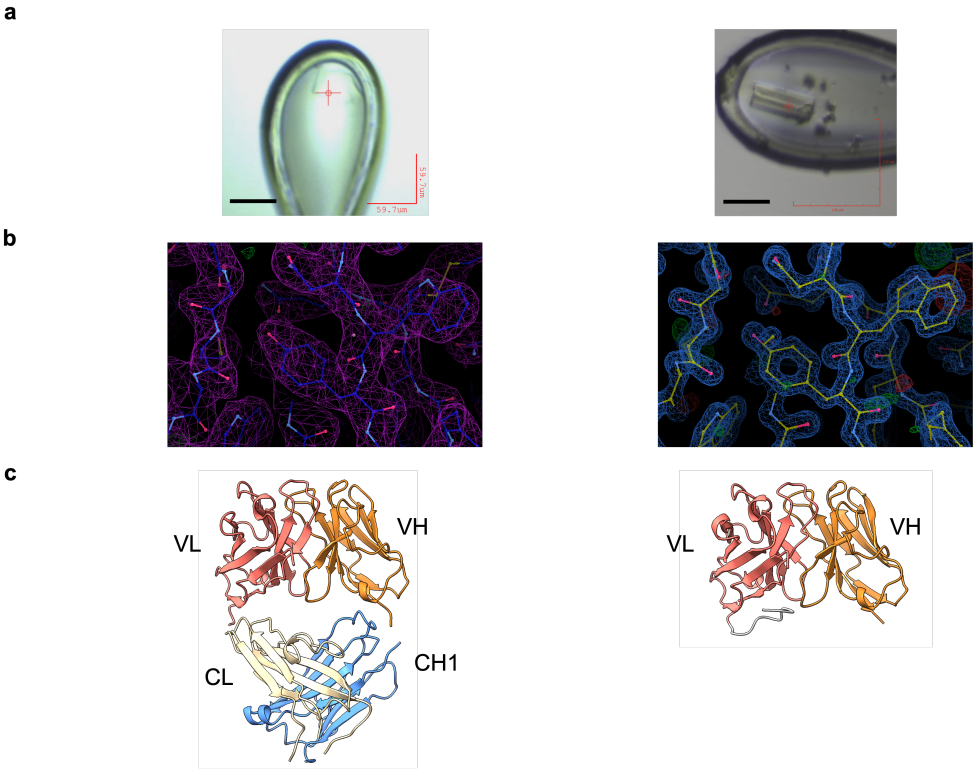


Figure 3.17: Crystal structures of CP33 Fab and scFv. a) Images of crystallography loops with crystals of CP33 Fab (left) and scFv (right). Scale bar is equal to 60 µm. b) Close-ups on electron densities and built model of CP33 Fab (left) and scFv (right), visualised in Coot. c) Ribbon representation of CP33 Fab (left) and scFv (right).

3.4.1.3.4 Complex formations for crystallisation

In some cases, antibody fragments can enhance solubility and/or stabilise target proteins, thereby helping crystals to form [364]. Therefore, antibodies were complexed with recombinant constructs that failed to crystallise. Complexes of full extracellular IgL (IgL_01) with XEhI_X20 and/or XEhI_H2 were purified and used in crystallisation trials. Similarly, various constructs of HgL were complexed to CP33 Fab or scFv (Figure 3.18). Nevertheless, no crystal hit was obtained in these conditions (Appendix Table A.2).

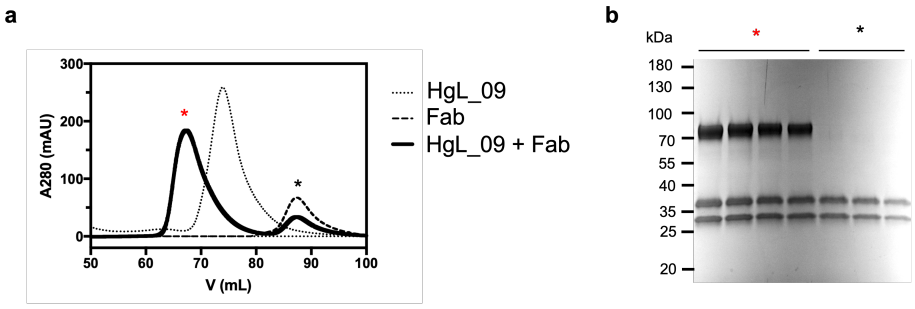


Figure 3.18: Complex formation with antibodies for crystallisation trials. a) SEC (S200 Increase 10/300 GL) profile of HgL_09 (LecA) complex formation with CP33 antibody (in black, plain line). Curves for CP33 Fab (dashes) and HgL_09 (round dots) are superimposed. b) Reducing SDS-PAGE of the corresponding fractions obtained by SEC. The estimated molecular weight of unglycosylated HgL_09 is 65.7 kDa.

3.4.1.3.5 Limited proteolysis coupled with MS

Since no crystals were obtained after complex formation with antibody fragments, limited proteolysis of an antibody-bound protein was performed. The most restricted construct that binds to CP33, expresses well, and does not contain reactive cysteines was selected. This construct, HgL_13 (residues 626 to 973), was complexed with CP33 scFv, incubated with various proteases and applied on SEC (Figure 3.19.a). The resulting fractions, analysed on gel, all contained intact complexes.

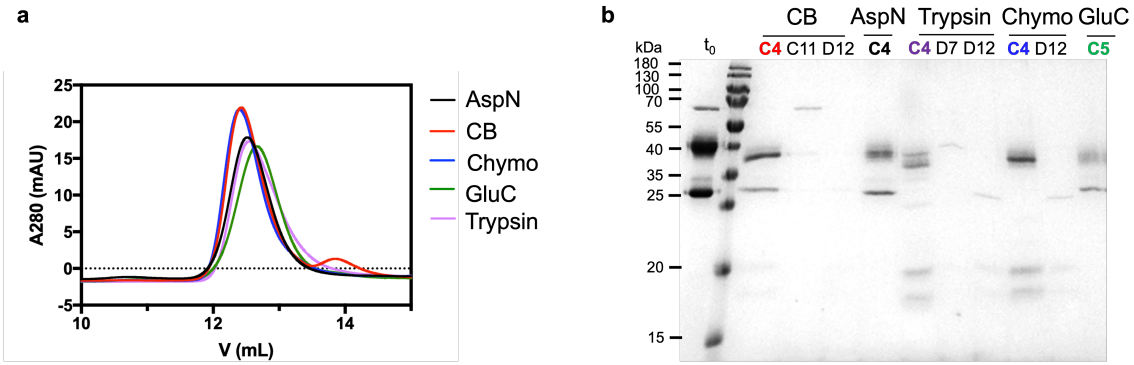


Figure 3.19: Limited proteolysis of scFv-bound HgL_13. a) SEC profiles of scFv-bound HgL_13 after overnight incubation with AspN (in black), carboxypeptidase B (CB, in red), chymotrypsin (chymo, in blue), GluC (in green), and trypsin (in purple). b) Reducing SDS-PAGE of the corresponding SEC fractions. Fractions of the main peak are labeled in bold, enzymes are labeled as in a). The t_0 lane corresponds to the complex deglycosylated with EndoHf before digestion. Expected molecular weights are 27.6 kDa, 40.7 kDa and 70 kDa for intact unglycosylated scFv, HgL_13 and EndoHf, respectively.

Limited proteolyses with trypsin, GluC and AspN yielded many intermediates of

similar sizes, whereas more homogenous samples were obtained with carboxypeptidase B and chymotrypsin (Figure 3.19.b). In the presence of chymotrypsin and trypsin, scFv was digested in two species, possibly at its glycine-rich linker. Digests with carboxypeptidase B and chymotrypsin were selected and scaled up for crystallisation purposes. However, no crystal was obtained in these conditions.

3.4.1.4 Crystal structure of HgL VWD

A crystal of deglycosylated HgL VWD diffracted to a resolution of 3.31 Å (Figure 3.20.a). As predicted, the structure solved by molecular replacement is that of a VWD and contains six central β-sheets surrounded by six α-helices (Figure 3.20.b). This fold is found in many extracellular proteins, such as von Willebrand and complement factors, integrins or collagens [365]. In most cases, VWDs are involved in formation of multi-protein complexes relevant for various biological processes.

Structure alignment with integrin α1-β1 I domain, the first hit obtained in DALI search, reveals that the C-terminal halves (β-strand 4 to helix 6) are well aligned, with an estimated RMSD of 3.7 Å, while the N-terminal halves deviate more, with an estimated RMSD of 6.6 Å (β-strand 1 to helix 3), which could explain why molecular replacement with homology models yielded poor solutions (Figure 3.20.c).

Many VWDs have been shown to interact with a variety of ECM ligands. Ligand binding often requires cation coordination at a well-conserved metal ion-dependent adhesion site (MIDAS). This motif consists of conserved residues DXSXS or DXTXS found at the C-terminus of β-strand 1, which coordinate cations such as magnesium, calcium or cadmium. Structure-sequence alignment with various VWD shows that HgL VWD lacks a MIDAS motif and signal for magnesium there was not found though crystals were obtained in the presence of $MgCl_2$ (Figure 3.20.d-f). In fact, the loop of HgL VWD there is much smaller.

Although most VWDs have been shown to interact with other proteins in a cation-dependent manner through their MIDAS, some VWDs lack a MIDAS but still associate with ligands. Among these, the von Willebrand factor A1 domain associates with glycoprotein Ib [366, 367] in a metal-independent manner. Likewise, HgL VWD may

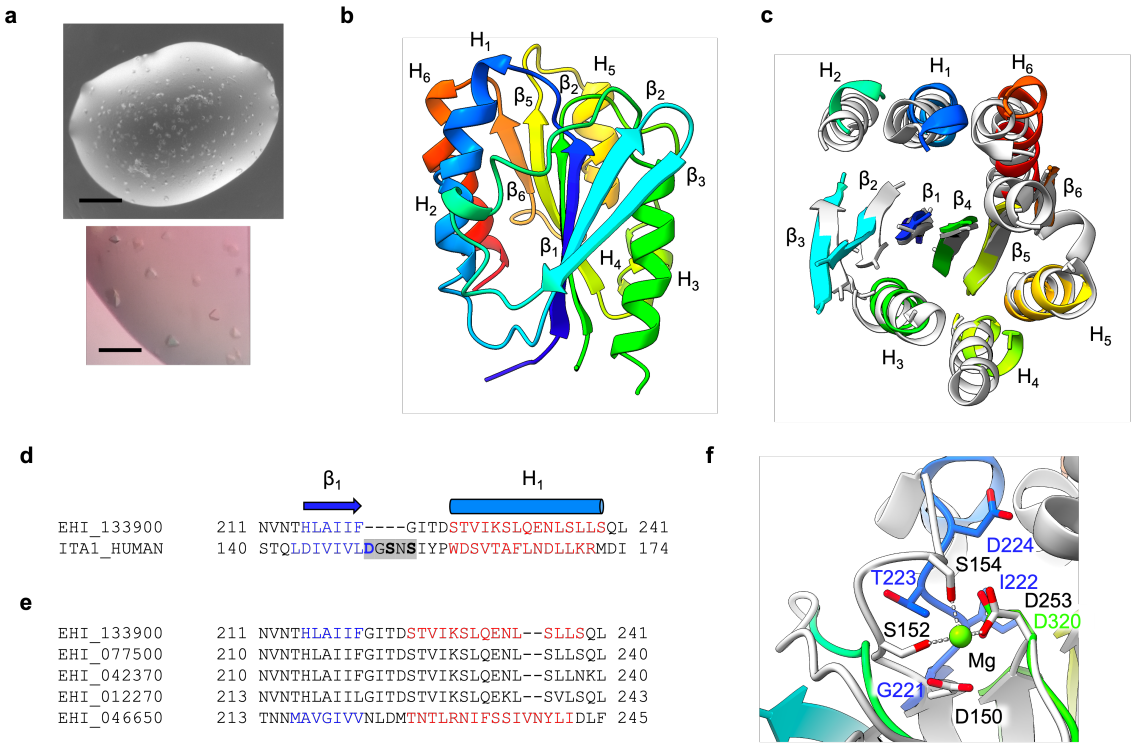


Figure 3.20: Crystal structure of HgL VWD. a) Images of drops in which crystals of HgL VWD grew. Scale bars are equal to 500 and 100 μm. b) Rainbow ribbon representation of the structure of HgL VWD. c) Structure alignment of HgL VWD (in rainbow) with integrin α1-β1 I domain (PDB code 1PT6, in light grey). Coils are not shown for clarity. d) Structure-based sequence alignment generated in Chimera of the region surrounding the location of a MIDAS motif for HgL VWD and integrin α1-β1 I domain (ITA1_HUMAN, PDB code 1PT6). The location of the MIDAS motif DXSXS (residues in bold) is highlighted in grey. HgL VWD helix 1 and β-strand 1 are depicted above the corresponding sequence with cylinders and arrows, respectively. Residues of helix 1 and β-strand 1 are coloured in red and blue, respectively. e) Sequence alignment based on secondary structure predictions generated in PROMALS3D of all isoforms of *Entamoeba histolytica* HgL VWD. Residues of helix 1 and β-strand 1 are coloured in red and blue, respectively. f) Ribbon and stick close-up of the MIDAS site of integrin α1-β1 I domain (PDB code 1PT6, in light grey) coordinated with a magnesium ion (Mg), aligned with HgL VWD (in rainbow).

form complexes with extracellular ligands, which could regulate processes important for parasite invasion, such as cell adhesion or signal transduction.

3.4.1.5 Purification of the heterodimer of LgL in complex with HgL

While LgL could not be purified on its own, it was successfully co-purified with HgL as a heterodimer, which includes all cysteines of both subunits. For this, sequences of helices of a parallel coiled-coil were introduced at the C-terminus of both subunits to help stabilise heterodimerisation, as previously observed in integrins [365]. In this case, the coiled-coil mimics the transmembrane domain of HgL and the potential GPI anchor of

LgL. Alternatively, co-expression was also performed without coiled-coils (Figure 3.21.a).

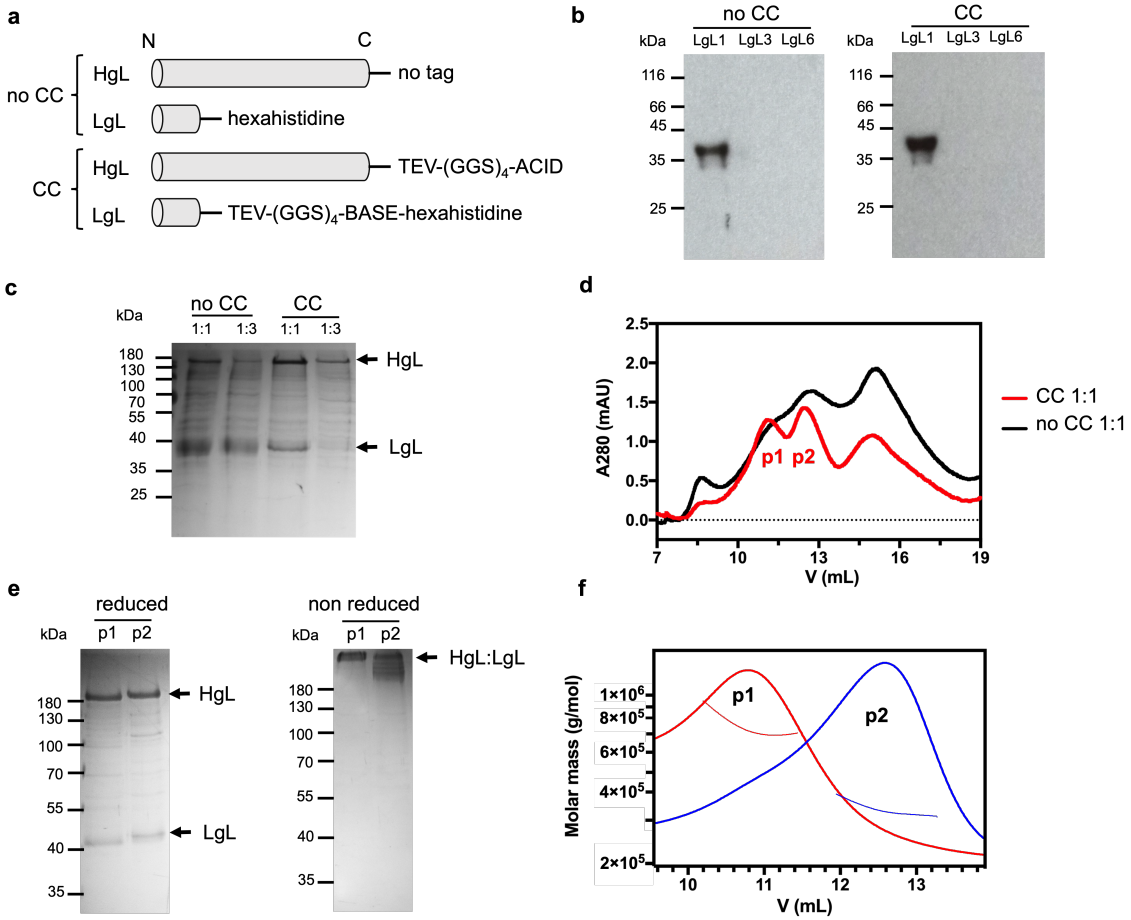


Figure 3.21: Recombinant expression of LgL:HgL heterodimer in Expi293F cells. a) Schematic representation of LgL and HgL constructs used in co-expression trials. Without coiled-coil (no CC), HgL has no tag whereas LgL contains the polyhistidine tag (6His) used for affinity chromatography. With coiled-coil (CC), HgL contains a TEV cleavage sequence (TEV), followed by a $(GGS)_4$ linker and the sequence for ACID coiled-coil, while LgL contains a TEV cleavage site, a $(GGS)_4$ linker, a sequence for BASE coiled-coil and a C-terminal hexahistidine-tag for purification. b) Western blot of reduced supernatants after co-expression of HgL with LgL isoforms EHI_035690 (LgL1), EHI_148790 (LgL3) and EHI_183400 (LgL6). An antibody against pentahistidine coupled to horseradish peroxidase was used for detection. The estimated molecular weights of LgL with and without coiled-coil are 37.0 kDa and 31.7 kDa, respectively. c) Reducing SDS-PAGE for fractions eluted by IMAC during small scale purification of HgL:LgL heterodimer. HgL and LgL were co-transfected using plasmids in 1:1 and 1:3 molar ratios (HgL:LgL), with and without coiled-coils. The estimated molecular weights of unglycosylated HgL, with and without coiled-coil, are 142.5 kDa and 137.1 kDa, respectively. d) SEC (S200 Increase 10/300 GL) curves of HgL:LgL heterodimers after IMAC. Peaks p1 and p2 were obtained in the presence of coiled-coil. e) SDS-PAGE of p1 and p2 SEC fractions in reducing and non-reducing conditions. The estimated molecular weights of unglycosylated HgL:LgL complexes, with and without coiled-coil, are 179.5 kDa and 168.8 kDa, respectively. f) SEC-MALS profile of both species after scale up.

Intriguingly, in both cases, LgL isoform EHI_035690 was co-expressed with HgL (isoform EHI_133900), whereas LgL isoforms EHI_148790 and EHI_183400 failed to

express (Figure 3.21.b-d). This suggests that heterodimerisation is isoform-dependent.

Importantly, purified recombinant heterodimers behaved similarly to the native 260-kDa antigen. Indeed, both subunits were covalently linked (Figure 3.21.e). In addition, the complex eluted as two peaks on SEC, and both species had estimated apparent molecular weights of 330 and 740 kDa by SEC-MALS (Figure 3.21.f). This suggests that HgL and LgL are able to form dimer of heterodimers, as previously suggested for the native 260-kDa antigen [262].

Introduction of coiled-coil consistently led to better separation of both species (Figure 3.21.d). Therefore, expression was scaled up and both SEC species were used for crystallisation trials. However, these failed to form crystals.

These results indicate that LgL requires HgL to be stable in solution. Stabilisation could occur through covalent interactions or formation of hydrophobic interfaces that may be exposed in the absence of HgL.

3.4.2 Purification of the native amoebic antigen

In parallel to recombinant production of protein fragments for crystallography, the native 260-kDa antigen was extracted from amoebae for structural studies by cryo-EM.

3.4.2.1 Establishment of axenic culture of *Entamoeba histolytica*

Culture of *Entamoeba histolytica* has been reported to be tedious and lengthy. Parasites are first established xenically, *ie* in culture with a complex undefined bacterial flora [368], and later adapted to axenic conditions, in which no other metabolising cells are present. Axenic media that support growth of *Entamoeba histolytica* contain amino-acids from trypticase, casein digest or liver digest peptones, nucleic acids provided by yeast extract, glucose, lipids from serum and vitamins. Establishing and maintaining axenic culture of *Entamoeba histolytica* is challenging because many batches of sera, peptones and yeast extract do not support parasite growth for unclear reasons.

Therefore, much efforts were invested in adaptation and maintenance of *Entamoeba histolytica* in axenic conditions. In particular, several batches of serum did not support

parasite growth. Nevertheless, after screening, amoebae were successfully adapted in two batches of sera and their production was scaled up for protein extraction.

3.4.2.2 Isolation of the native 260-kDa antigen

The 260-kDa antigen was first purified using Gal-affinity chromatography and anti-CP33-affinity chromatography in the presence of OG. From 5 g of parasites, 250 µg of pure heterodimer were extracted by immuno-affinity chromatography. In contrast, five times less material was obtained by Gal-affinity chromatography, during which contamination by other putative lectins occured (Figure 3.22.a).

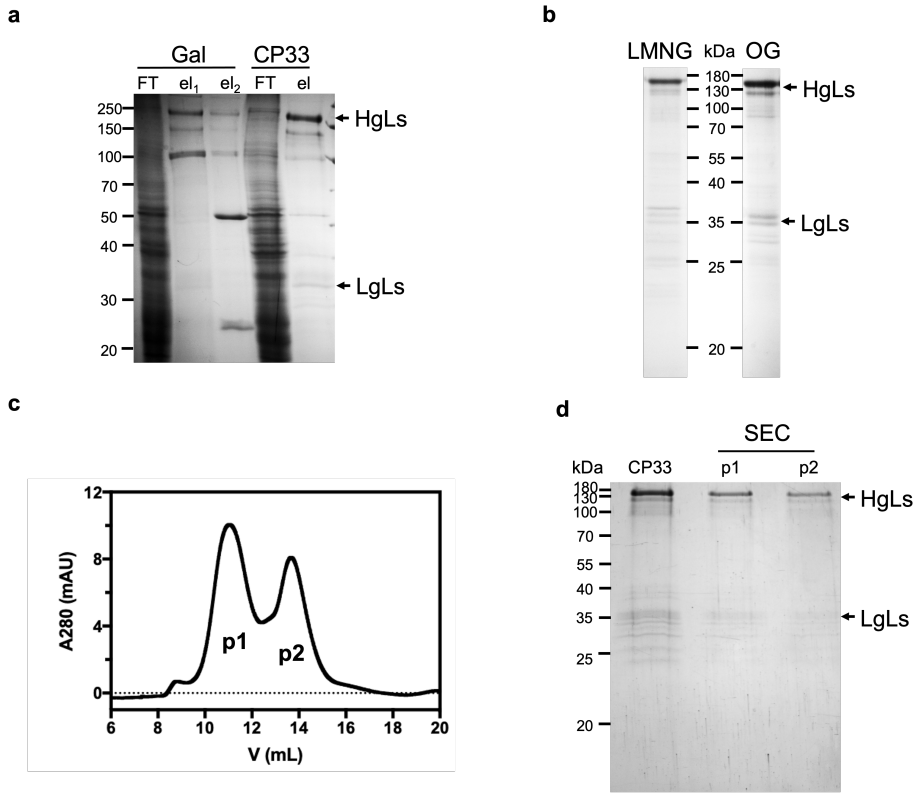


Figure 3.22: Purification of the 260-kDa native antigen in detergent. a) Reducing SDS-PAGE of samples purified during extraction of the native 260-kDa antigen from trophozoites using Gal-affinity chromatography and anti-CP33-affinity chromatography. In each case, the flow-through (FT) and elutions (el) are shown. Elutions were performed with Gal at 100 mM and 1 M during Gal-affinity chromatography (el_1 and el_2, respectively) and at low pH during anti-CP33-affinity chromatography. Estimated molecular weights are 170 kDa and 31-35kDa for HgL and LgL isoforms, respectively. b) Reducing SDS-PAGE of samples purified by immuno-affinity chromatography using OG and LMNG. c) SEC (S200 Increase 10/300GL) profile of the sample purified by immuno-affinity chromatography with LMNG. Two peaks are observed (p1 and p2). d) Reducing SDS-PAGE of the sample purified by immuno-affinity chromatography with LMNG, before (CP33) and after SEC (fractions p1 and p2 as in c)).

Yields obtained by immuno-affinity chromatography with OG were compared with those obtained during extraction with LMNG, a solvent more suited for cryo-EM applications (Figure 3.22.b). In this case, 150-200 µg of protein were consistently purified from 5 g of parasites and eluted as two peaks during SEC (Figure 3.22.c). Both samples looked similar on reducing gel, suggesting formation of a higher order oligomer, similarly to the recombinant heterodimer (Figure 3.22.c,d). Notably, the ratio between both peaks differed between purifications.

Tryptic digest of the top gel band followed by mass spectrometry identified unique peptides from 4 isoforms of HgL (EHI_133900, EHI_012270, EHI_077500, EHI_042370, from higher to lower intensity) in both samples, suggesting the presence of a mix of isoforms after purification (Appendix Table A.3).

Besides, the 260-kDa antigen was also successfully extracted using DIBMA, a synthetic polymer that can solubilise membrane proteins from native lipids (Figure 3.23). However, only traces of the 260-kDa antigen could be obtained by Gal-affinity chromatography (Figure 3.23.b). Moreover, elution at low pH triggered precipitation of DIBMA lipid particles during immuno-affinity chromatography. Nonetheless, the native lectin was successfully purified by IMAC after incubation of parasite lysates with CP33 Fab. Remarkably, the eluted complex behaved as a single specie during SEC and protein purification appeared cleaner (Figure 3.23.a). However, the yield obtained was 20 times lower than with LMNG, probably due to low extraction efficiency.

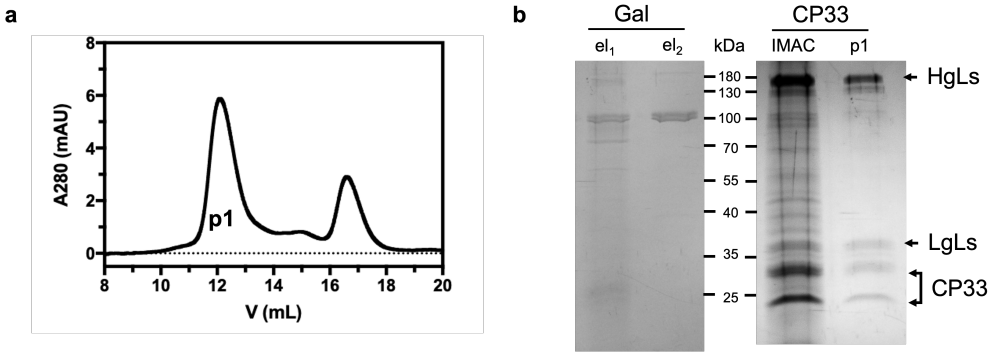


Figure 3.23: Purification of the 260-kDa native antigen in lipids. a) SEC (S200 Increase 10/300GL) profile of the sample extracted with DIBMA with an excess of CP33 Fab. One major peak (p1) is observed. b) Reducing SDS-PAGE of samples purified with DIBMA after Gal-affinity chromatography (Gal, left gel), or after IMAC and SEC (p1) using CP33 (right). Elutions were performed with Gal at 100 mM and 1 M during Gal-affinity chromatography (el_1 and el_2, respectively)

3.4.3 Cryo-EM studies of the native 260-kDa antigen

3.4.3.1 Negative staining

Both species obtained by immuno-affinity chromatography after extraction with LMNG were analysed by negative staining and showed homogenous particles in both cases (Figure 3.24). This indicated that samples were suitable for cryo-EM studies. Most particles from the first peak looked larger than that of the second peak.

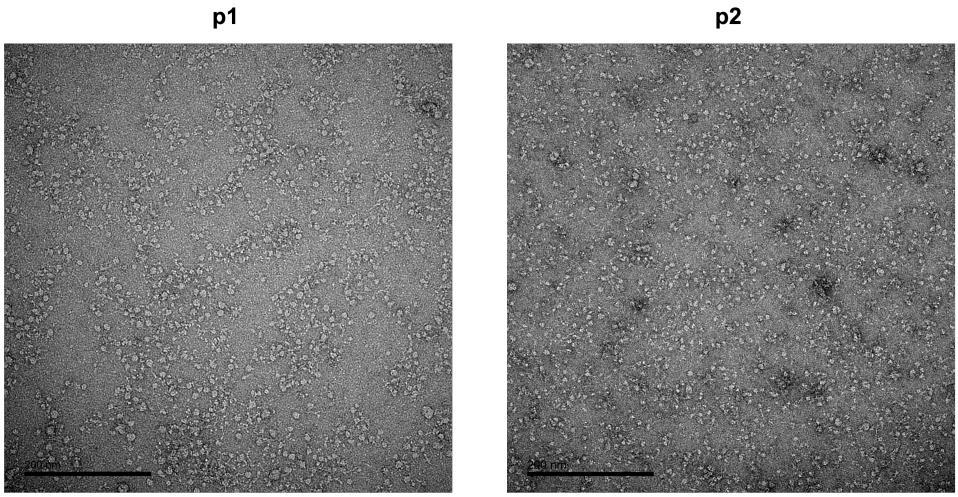


Figure 3.24: Negative staining after immuno-affinity chromatography. Samples from both fractions (p1 and p2) were analysed separately on grids. Scale bar is equal to 200 nm.

3.4.3.2 Optimisation of specimen for data acquisition

The native sample was applied on cryo-EM grids for structure determination. Most screenings were done with the sample obtained from the second peak on SEC, due to a better behaviour on grids (Figure 3.25.a). A first data collection yielded a dataset, from which a single view was predominant (Figure 3.25.b) and highlighted a rather compact protein with a domain acting as a cap, and a fuzzy intensity coming out of the core towards the outside of the box.

To reduce preferred orientation and increase weight and signal, the antigen was then complexed with CP33 Fab. While the same predominant class was observed, an additional view showed 2D structural features similar to that of a Fab, attached to a protrusion which extends towards the outside of the box.

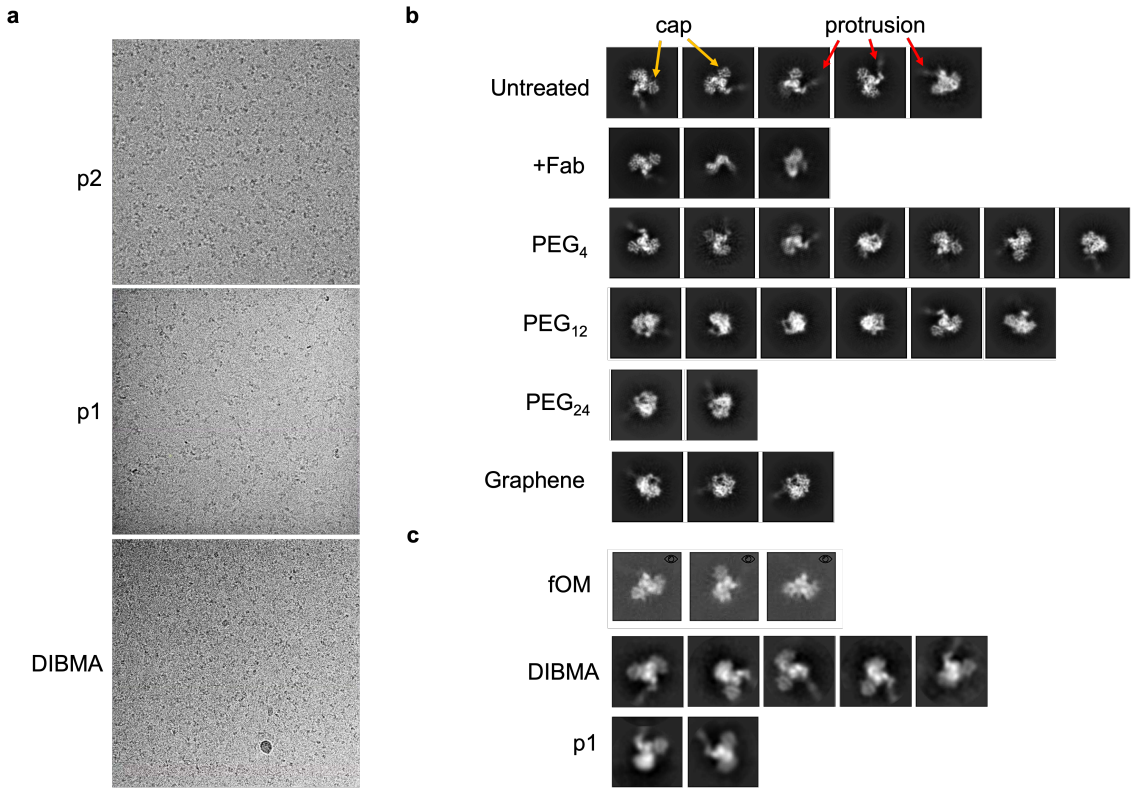


Figure 3.25: Sample optimisation for cryo-EM data collection. a) Micrographs of samples applied on grid, at a defocus of -2 μm, and a magnification of 190,000. Both samples purified by SEC after protein extraction with LMNG and immuno-affinity chromatography were screened (p1 and p2), in addition to the DIBMA-solubilised Fab-bound sample. b) 2D classes obtained for p2, with a box size of 256 × 256 pixels. Classes are shown for samples unbound (untreated), complexed with Fab (+Fab), PEGylated (PEG_n) or applied on graphene. Red and orange arrows point towards the fuzzy protrusion and the cap, respectively. c) 2D classes of p2 obtained in fluorinated octyl-maltoside (fOM), DIBMA-solubilised Fab-bound sample (DIBMA) and untreated p1, with a box size of 180 × 180 pixels.

Then, graphene support was used to alter particle distribution and reduce potential protein adsorption at the air-water interface. Though data quality was reduced, a single view that differs from previous 2D classes was observed. Nevertheless, at this stage, combination of all datasets failed to yield an acceptable 3D ab-initio volume.

Notably, screenings were also performed in fluorinated detergents, *ie* surfactants that can saturate the air-water interface while interacting poorly with proteins [369–372]. With these, ice behaviour changed drastically and data quality was much worse (Figure 3.25.c). Moreover, these detergents required 30 times more sample to observe particles distributed across foilholes. Since yields of native samples were limited, other methods that alter particle distribution in ice were favoured.

Eventually, preferential orientation was reduced after sample PEGylation. The effect

of lysine PEGylation by PEGs of different lengths was investigated. Remarkably, new views were obtained after PEGylation with PEG_4 and PEG_{24}, while PEG_{12} yielded the best angular distribution.

Besides, samples extracted in lipids with DIBMA, or contained in the first SEC peak during extraction with LMNG both displayed similar 2D classes (Figure 3.25.c). In the latter case, this suggests that potential higher order oligomers may fall off during freezing.

3.4.3.3 Cryo-EM processing

All datasets were combined to generate a 3D ab-initio model, from which 3D classifications and refinements allowed reconstruction of a map at an overall resolution of 2.96 Å (Figure 3.26.a,b). Notably, the 3D reconstruction is rather globular with very few helical secondary structures. Four domains can be distinguished. A globular domain sits on top of a central core, which contains long β-strands, and contacts a globular domain that displays helical features at the bottom of the reconstruction. Remarkably, a less well-resolved ring of consecutive small domains extends from the core and creates a cavity, when observed from the top (Figure 3.26.c).

Local resolution ranges from 2.5 to 4.0 Å, with the best resolution achieved at the central core of the protein. Less well-resolved areas include the ring of small domains and the bottom of the reconstruction (bottom view).

In addition, a bigger dataset was collected for native antigen incubated with GalNAc and PEGylated with PEG_{12}. This yielded a map at 3.07 Å, which displayed some degree of preferential orientation compared to the combined dataset (Figure 3.26.b,d,e).

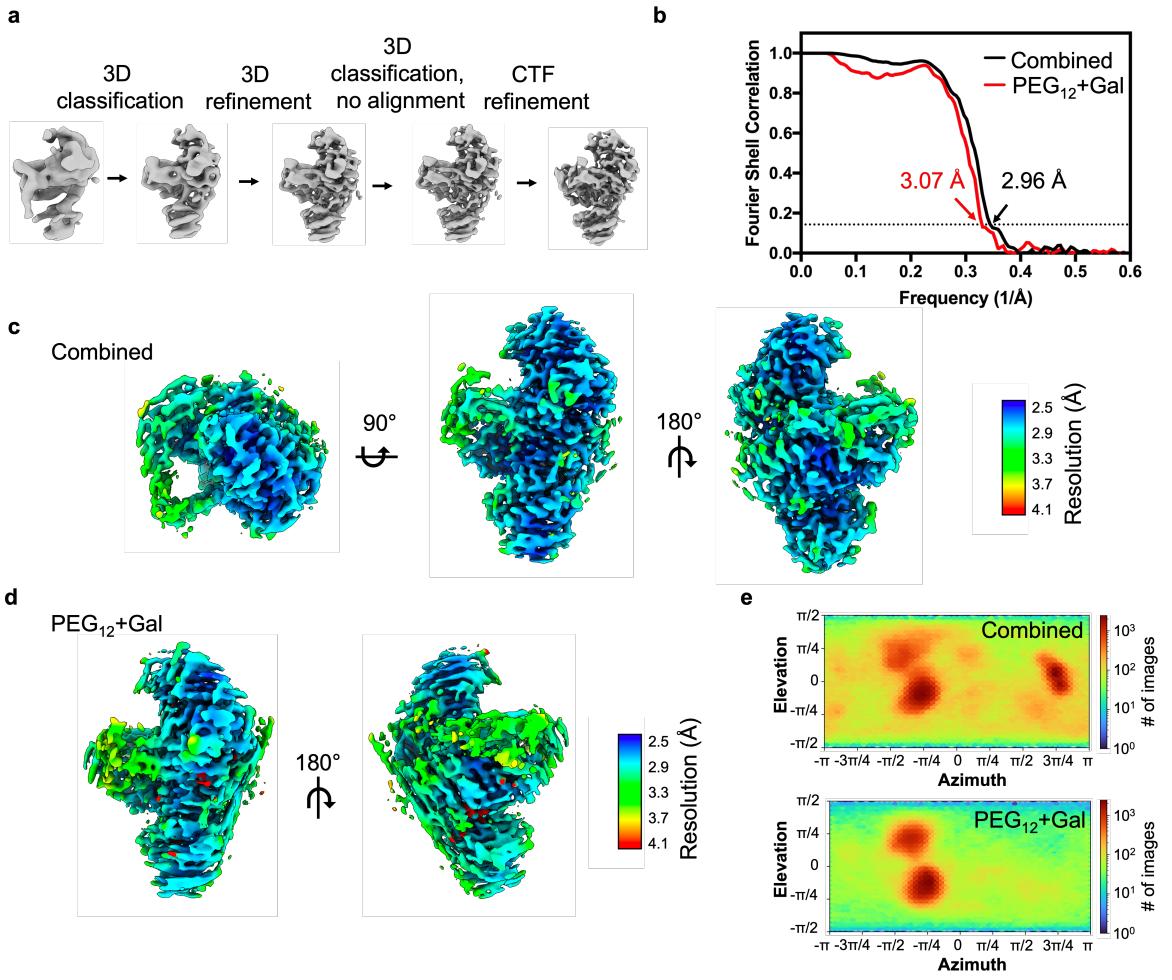


Figure 3.26: Processing pipeline and resolution estimates. a) Processing pipeline. b) FSC plot for the final 3D reconstructions of combined particles and sample in the presence of GalNAc and PEGylated with PEG_{12}. Resolution is determined at the FSC value of 0.143. c) Local resolution estimates of the 3D reconstruction of combined particles. d) Local resolution estimates of the 3D reconstruction of sample in the presence of GalNAc and PEGylated with PEG_{12}. Streaky artefacts can be observed. e) Orientation distribution maps for both reconstructions shown in c) (top) and d) (bottom).

3.4.3.4 Partial structure of the 260-kDa heterodimer

A preliminary model of the structure of the 260-kDa antigen was built in the cryo-EM reconstruction of the combined datasets. Using the model derived from the crystal structure of HgL VWD as a starting point, new domains were identified unambiguously and are presented here.

3.4.3.4.1 Structure of HgL

3.4.3.4.1.1 A lipoprotein-like domain

At the core of the electron density lies the N-terminal domain of HgL (Figure 3.27.a,b), a barrel formed by a series of eleven β-strands. Two halves (β_1 to β_6 and β_7 to β_{11}) are joined by a long loop, which wraps around the strands on one side and links the most external strands β_6 and β_7. In addition, a C-terminal helix wraps around the N-terminal half formed by β_1 to β_6 on the other side (Figure 3.27.c,d).

Interestingly, this domain shares a fold with bacterial lipoprotein transporters such as LolA [373] (Figure 3.27.e). Other proteins adopting this fold include human lysosomal ependymin-related protein 1 (EPDR1) [374], bacterial VioE and RseB [375, 376] and *Mycobacterium tuberculosis* LppX [377]. These are all single-domain proteins of about 200 residues that have been shown to interact with hydrophobic molecules such as lipids or PEGs. Similarly, the lipoprotein-like domain of the 260-kDa antigen forms a cavity of hydrophobic residues, but does not accommodate any hydrophobic ligand (Figure 3.27.f). This is the first report of such fold found in a multi-domain protein.

Interestingly, EPDR1 and VioE form homodimers. However, in structure alignments with the lipoprotein-like domain, the second copies overlap with the rest of the antigen.

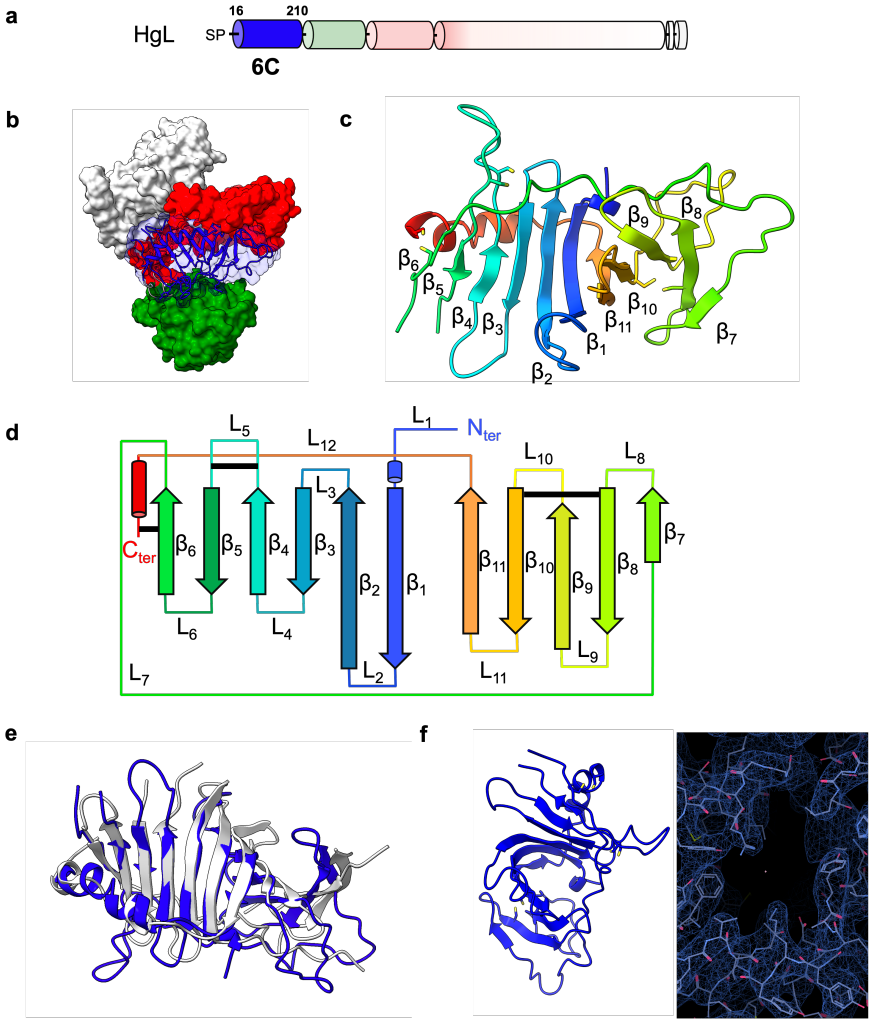


Figure 3.27: Structure of the N-terminal lipoprotein-like domain. a) Schematic representation of the domains of HgL, with emphasis on the N-terminal domain, in blue. b) Surface representation of the 260-kDa antigen, with transparency around the lipoprotein-like domain, depicted as a blue ribbon. c) Rainbow ribbon representation of the lipoprotein-like structure. d) Topological representation of the lipoprotein-like domain, with colours matching those in c). e) Ribbon representation of the lipoprotein-like domain (in blue) aligned to the crystal structure of EPDR1 (in light grey, PDB code 6E7O). f) Focus on the cavity formed by hydrophobic residues of the N-terminal domain. Ribbon representation of the cavity (left) and a close-up (right) in Coot with the corresponding EM density are shown. No extra density for potential hydrophobic molecules was observed.

3.4.3.4.1.2 A von Willebrand-like domain

The VWD structure is identical to the crystal structure of HgL VWD (described in section 3.4.1.4, Figure 3.20). Interestingly, at the C-terminal end of the last helix, density for an N-glycan on asparagine 390 is located at the interface between HgL VWD and the lipoprotein-like domain (Figure 3.28.d). This is one of the only glycosylation sites conserved across all sequences of HgL of *Entamoeba* species, suggesting an important

role in stabilising the structure.

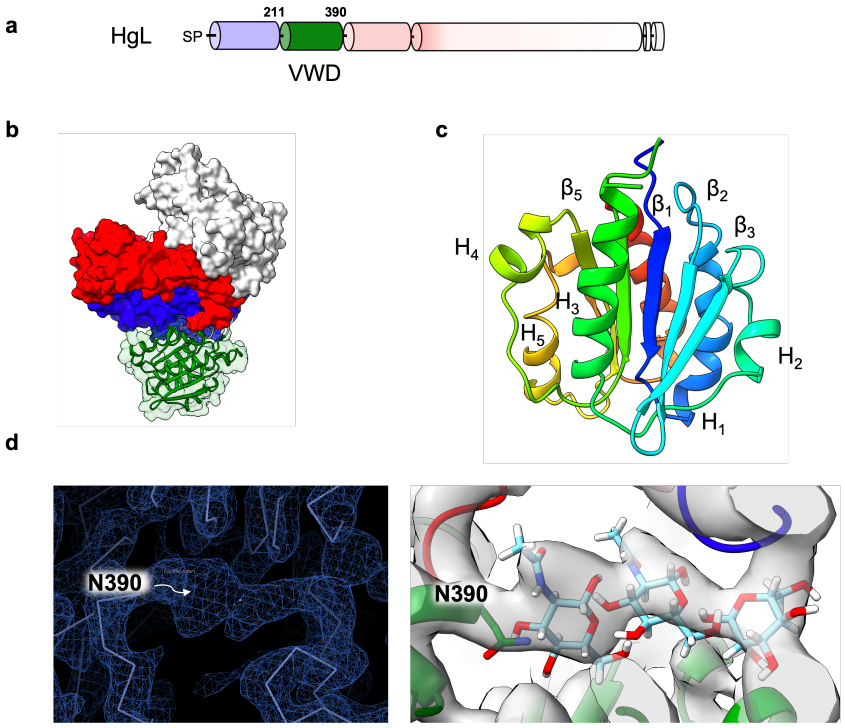


Figure 3.28: HgL VWD in the cryo-EM reconstruction. a) Schematic representation of the domains of HgL, with emphasis on the VWD, in green. b) Surface representation of the 260-kDa antigen, with transparency around HgL VWD, depicted as a green ribbon. The domain is at the bottom of the structure. c) Rainbow ribbon representation of HgL VWD structure. d) Focus on the interface between HgL VWD and the N-terminal lipoprotein-like domain, visualised in Coot (left) and ChimeraX (right). An extra density for glycosylation of asparagine 390 (N390) is apparent. There, two molecules of GalNAc and one molecule of mannose are manually fitted (right).

3.4.3.4.1.3 A partially resolved cysteine-rich region

Between the VWD and the transmembrane domain of HgL lies the cysteine-rich region, a stretch of 841 residues that consists of 91 cysteines. Of this, the structure of the N-terminal cysteine-rich region until residue 585 was built with confidence (Figure 3.29.a). Intriguingly, these residues form a ring that wraps around the core of the protein (Figure 3.29.b).

For each of these domains, no homologous structure was found in DALI. Nevertheless, sequence analysis reveals the presence of seven small pseudo-repeats of 21 to 34 residues that share a similar topology (Figure 3.29.c-f), which appears to consist of three anti-parallel strands. The N-terminal pseudo-repeat contains six cysteines, the first four

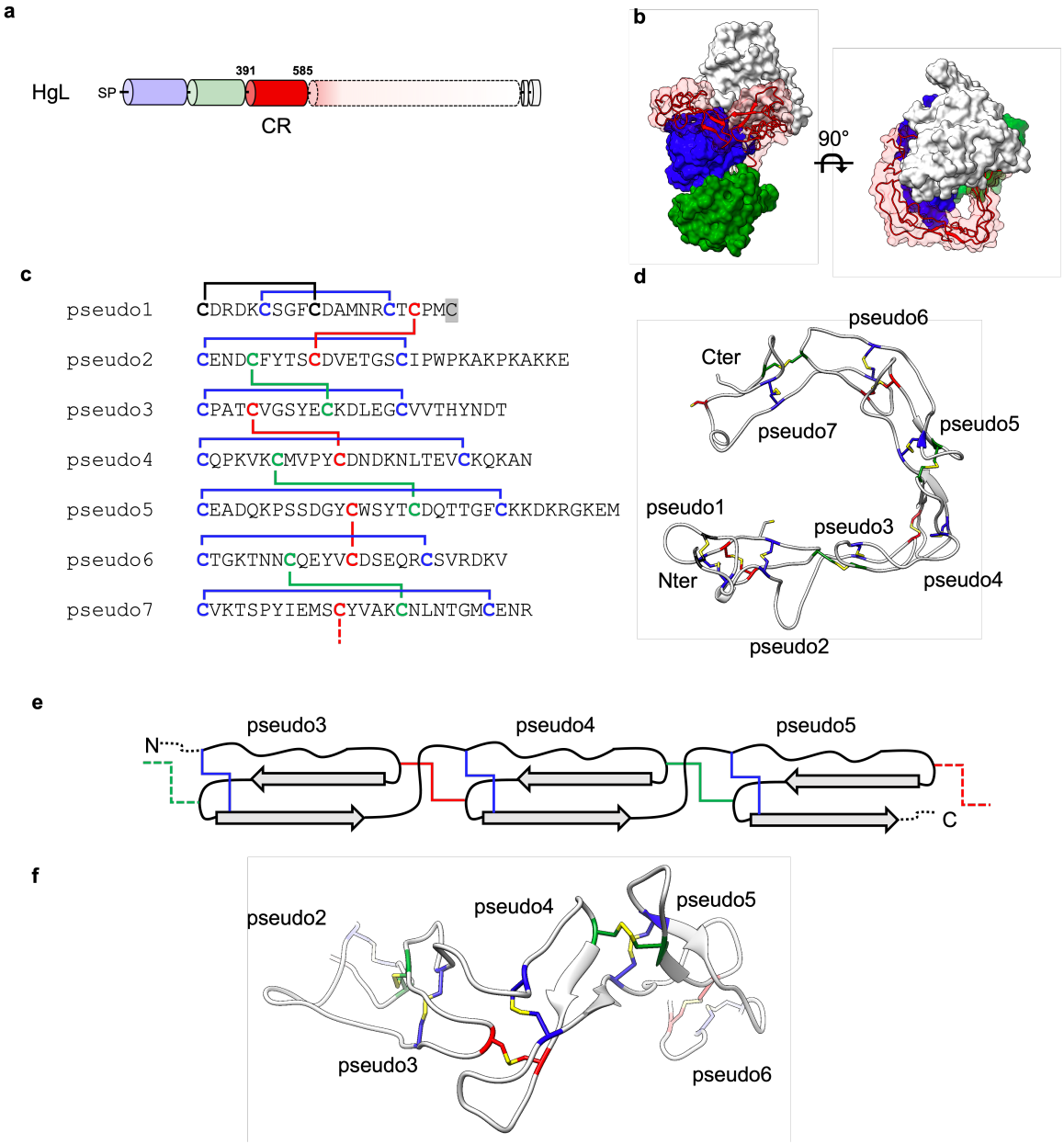


Figure 3.29: Structure of the cysteine-rich pseudo-domains. a) Schematic representation of the domains of HgL, with emphasis on the pseudo-repeats of the cysteine-rich region (CR), in red. b) Surface representation of the 260-kDa antigen, with transparency around the pseudo-repeats, depicted as a red ribbon. c) Sequences of each pseudo-repeat (pseudo1 to pseudo7). Cysteines that form disulfide bonds betwen domains are colored in red and green, while cysteines that form intra-domain covalent bonds are shown in blue (an black for pseudo1). For pseudo2 to pseudo7, the pattern formed by disulfide bonds is conserved. The cysteine of pseudo1, highlighted in grey, does not form any covalent bond with any cysteine of the pseudo-repeats. d) Ribbon representation of the pseudo-repeats, with cysteines coloured as in c). e) Schematic representation of the topology adopted by pseudo-repeats, with covalent bonds represented by coloured bridges between domains, using the same code as in c). f) Ribbon close-up representation of the region shown in e).

of which form intra-domain covalent bonds, while the last two cysteines are involved in inter-domain disulfide bonds. Each of the six other pseudo-repeats contains four cysteines, the first and fourth of which form covalent bonds between the first and the third β-strand, while the second and third cysteines are involved in disulfide bonds with adjacent C-terminal and N-terminal pseudo-repeats, respectively. These form a linear arrangement of interconnected domains that share a novel fold, which is distinct from EGF-like domains commonly found in cysteine-rich regions.

3.4.3.4.1.4 A flexibly-linked C-terminal half

Structure building at the C-terminal end of the pseudo repeats was hampered by poorly resolved density there, which could only be observed at high threshold. Remarkably, extraction of particles in larger boxes highlighted a long protrusion that moves away from the well-resolved core (Figure 3.30.a). At the tip of the elongated protrusion, a more globular domain can be distinguished (Figure 3.30.b). Linear cysteine-rich domains are likely to form this protrusion. Inter-domain linkers may contribute to flexibility and movement there, leading to poor averaging. Accordingly, focused classifications or local refinements only partially improved resolution for density proximal to the globular core (Figure 3.30.c).

Intriguingly, in 2D classes obtained in the presence of CP33 Fab, the characteristic projection of a Fab-like particle attached to a domain that sticks out is observed (Figure 3.30.d). Undoubtedly, this domain is not globular and may instead be flexible. This could explain why most constructs of the cysteine-rich domain failed to crystallise, regardless of the presence of antibody fragments.

Similarly, some 2D classes very different from that of the well-resolved portion of the 260-kDa antigen can be observed, but only account for a minority of particles. Nevertheless, these may correspond to domains of the C-terminus that could not be resolved. In support of this, most classes also display protrusions extending away from their globular core (Figure 3.30.e). Because of the small number of particles, only ab-initio models could be generated for these (Figure 3.30.f).

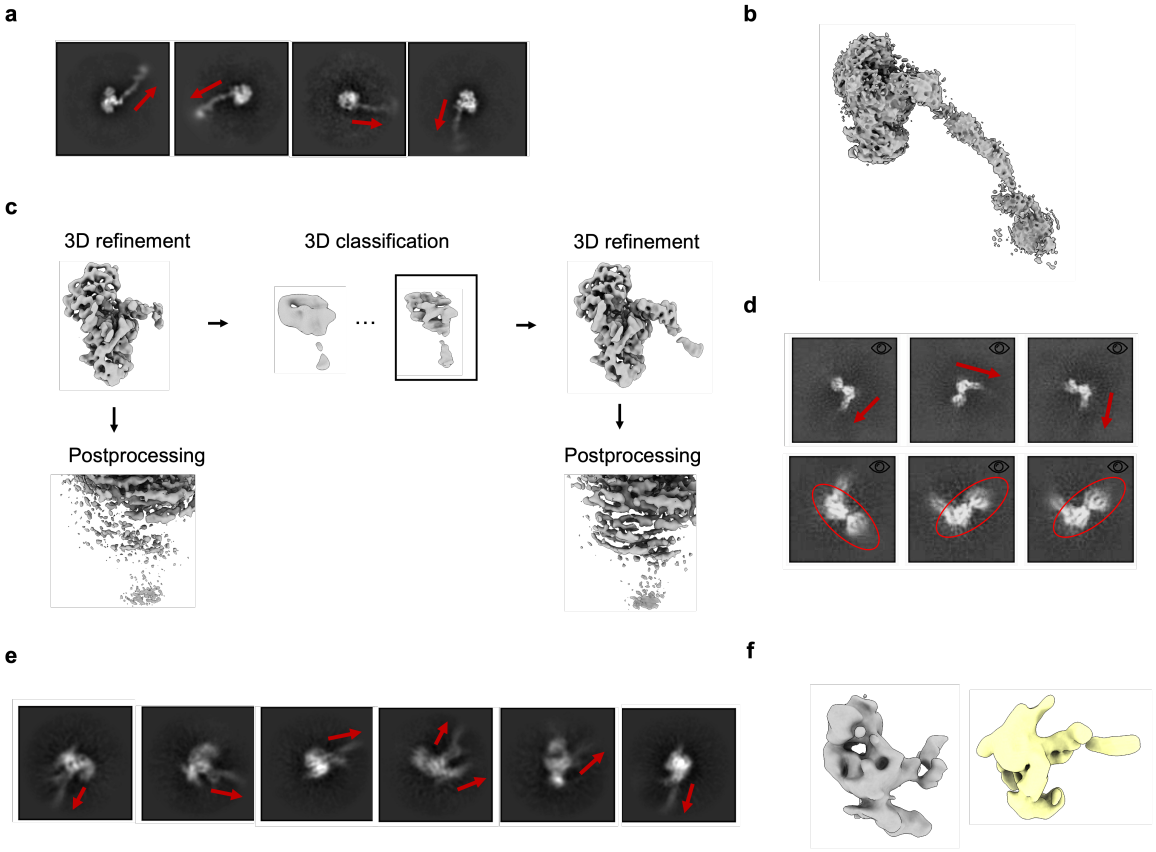


Figure 3.30: An elongated protrusion at the cysteine-rich region. a) 2D classes of the 260-kDa antigen, using a box size of 720 × 720 pixels. Red arrows show the direction taken by the protrusion (from N- to C-termini). b) Cryo-EM density of the 260-kDa antigen at high threshold after extraction with a box size of 512 × 512 pixels. c) Particle subtraction pipeline to better resolve the protrusion. The refined reconstructions and their corresponding protrusion densities are depicted, before (left) and after (right) particle subtraction. d) 2D classes of the 260-kDa antigen bound to Fab CP33, using a box size of 360 × 360 pixels (top) and 180 × 180 pixels (bottom). Intensity for Fab is observed with high contrast, while a less intense signal protrudes towards the outside of the box. Red arrows show the direction taken by the protrusion, and intensities for Fab are circled in red. e) 2D classes of the 260-kDa antigen, obtained for a minority of particles, using a box size of 256 × 256 pixels. f) Two 3D ab-initio reconstructions generated from particles in e).

Therefore, the cysteine-rich region of HgL seems to contain at least three subregions: an ordered ring of pseudo repeats, followed by a linear arrangement of elongated cysteine-rich domains, which join a more globular domain further at the C-terminus.

3.4.3.4.2 Structure of LgL

The light chain was built in the remaining density at the top of the reconstruction. The globular domain is in contact with the lipoprotein-like domain and the ring of pseudo-repeats (Figure 3.31.a,b). As predicted, residues 76 to 271 adopt the topology of a β-trefoil domain, with two halves of five β-strands forming a cap and a barrel in a pseudo three-fold symmetry, and the presence of three QXF motif (Figure 3.31.a-d).

The N-terminal residues are partially resolved but form a long coil which shields part of the otherwise solvent-exposed β-trefoil domain (Figure 3.31.c). Three cysteines are located at the junction between these two domains, one of which (cysteine 75) forms a disulfide bond with cysteine 415 of HgL first pseudo-repeat, while cysteines 71 and 74 form an intra-domain covalent bond (Figure 3.31.e). A third disulfide bond is formed between cysteines 241 and 274 located in the β-trefoil domain. In the structure, all C-terminal residues are observed, suggesting LgL is not GPI-anchored.

Alignment of LgL with the structure of a β-trefoil domain in *Entamoeba histolytica* [378] shows a similar arrangement of β-strands, with most changes occurring in the length of connecting loops (Figure 3.31.d). Strikingly, loops of LgL are much longer, especially between β_1 and β_2 (L_1), β_2 and β_3 (L_2), β_8 and β_9 (L_8), β_{10} and β_{11} (L_{10}). Interestingly, β-trefoil domains are lectins, which suggests that LgL plays a role in binding to carbohydrates.

Moreover, the interface between LgL and HgL is hydrophobic (Figure 3.31.f). This may explain why constructs of LgL failed to express in the absence of HgL.

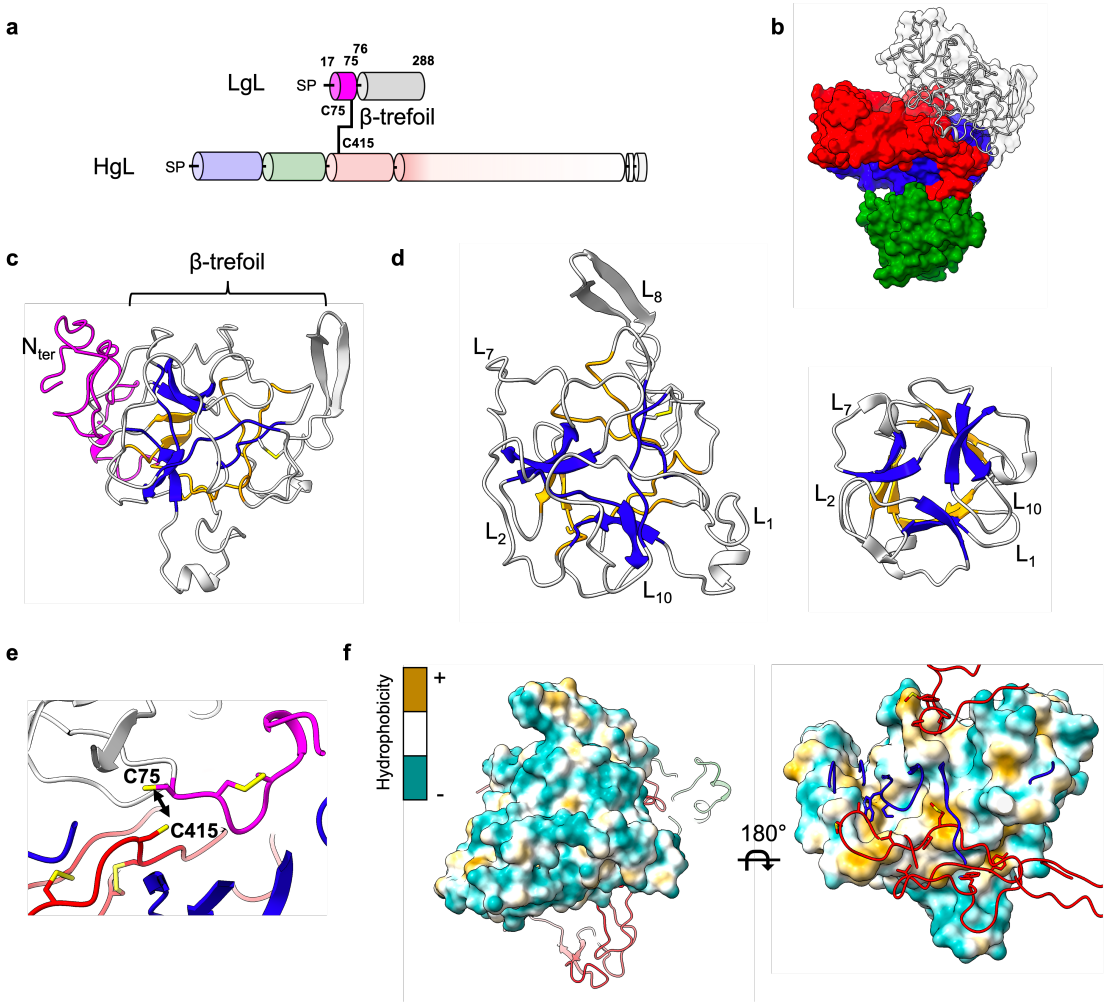


Figure 3.31: Structure of the light chain. a) Schematic representation of the domains of LgL and its covalent association with HgL. The β-trefoil domain is shown in grey (residues 76 to 288) and the N-terminal disordered domain is depicted in magenta (residues 17 to 75). LgL and HgL are covalently linked by a disulfide bond between cysteines 75 of LgL and 415 of HgL. b) Surface representations of the 260-kDa antigen, with transparency around LgL, depicted as a light grey ribbon. LgL is located on top of HgL N-terminal half, in contact with the lipoprotein-like domain and the cysteine-rich pseudo domains. c) Ribbon depiction of LgL. The β-trefoil domain is depicted in light grey, with β-strands of the cap domain shown in blue, and those of the barrel domain coloured in orange. The N-terminal domain is shown in magenta. d) Ribbon representation of LgL (left) and an other β-trefoil domain of *Entamoeba histolytica* (PDB code 6IFB). Structures were aligned and are depicted from the same angle, with cap β-strands at the front and barrel β-strands at the back. The colour code is the same as in c). Differences in loop sizes can be appreciated. e) Ribbon representation of a close-up on the covalent interaction between HgL cysteine 415 and LgL cysteine 75. The colour code is the same as in a). f) Surface hydrophobicity representations of LgL, from the top (left) and botom (right). HgL is shown as a ribbon, with the colour code as in a). Hgl and LgL interact through hydrophobic interactions, while the solvent-exposed residues of LgL are mostly hydrophilic.

3.4.3.5 Preliminary cryo-EM studies of HgL C-terminus

Since the C-terminal half of HgL was poorly resolved in the cryo-EM reconstruction, possibly due to movement between both halves of HgL, it was then studied independently from the rest of the 260-kDa antigen by cryo-EM.

3.4.3.5.1 Construct design and protein expression

Structure building in the cryo-EM map of the native 260-kDa antigen showed that the poorly resolved region of HgL starts after residue 585. In addition, models of the cysteine-rich region obtained in Alphafold and RosettaFold suggested that cysteines C-terminal to residue 641 are not involved in any disulfide bond with cysteines N-terminal to 641. Moreover, based on the cryo-EM reconstruction, residues between 585 and 641 are likely to contribute to the elongated protrusion.

Therefore, a construct (HgL_20) which spans residues 641 to 1220 was designed, and the protein was expressed in Expi293F cells, purified by IMAC, complexed to inhibitory Fab CP33 on SEC and vitrified without modification on cryo-EM grids (Figure 3.32).

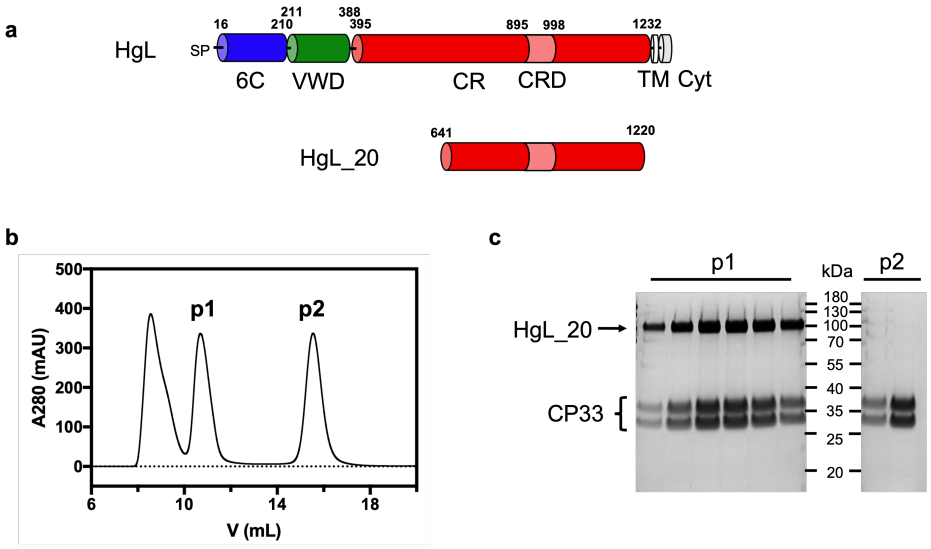


Figure 3.32: Construct design of HgL C-terminus and complex formation with CP33 Fab. a) Schematic representation of HgL domains and HgL_20. b) SEC (S200 Increase 10/300 GL) curve of HgL_20 purified by IMAC and complexed with an excess of Fab. Apart from the void, two peaks are observed (p1 and p2). c) Reducing SDS-PAGE of the fractions of p1 and p2 shown in b). p1 contains HgL_20 bound to Fab, p2 contains excess Fab. The estimated molecular weight of unglycosylated HgL_20 is 66.7 kDa.

3.4.3.5.2 Cryo-EM data collection, processing and limitations

Particles of untreated HgL_20 applied on grids could be observed by cryo-EM and a dataset was collected (Figure 3.33.a). Different views were observed during 2D classification, which could be separated in two groups (Figure 3.33.b). Interestingly, some minor views from cryo-EM imaging of the native sample were similar. After generation of ab-initio 3D reconstructions, two classes were identified and each was refined to a resolution of 3.4 Å (Figure 3.33.c-f).

Although the dataset suffers from preferential orientation, both reconstructions show features for β-strands, and protruding densities that extend away from their core. In the first reconstruction, CP33 Fab could be fitted on one side of the reconstruction, but was poorly fitted on the other side (Figure 3.33.g). This is either due to the absence of Fab, or lack of information due to preferential orientation, which could cause density artefacts, such as streaks where views are missing. Nonetheless, 2D classes obtained with these particles are similar to those obtained with native sample in the absence of Fab, suggesting that CP33 could bind somewhere else.

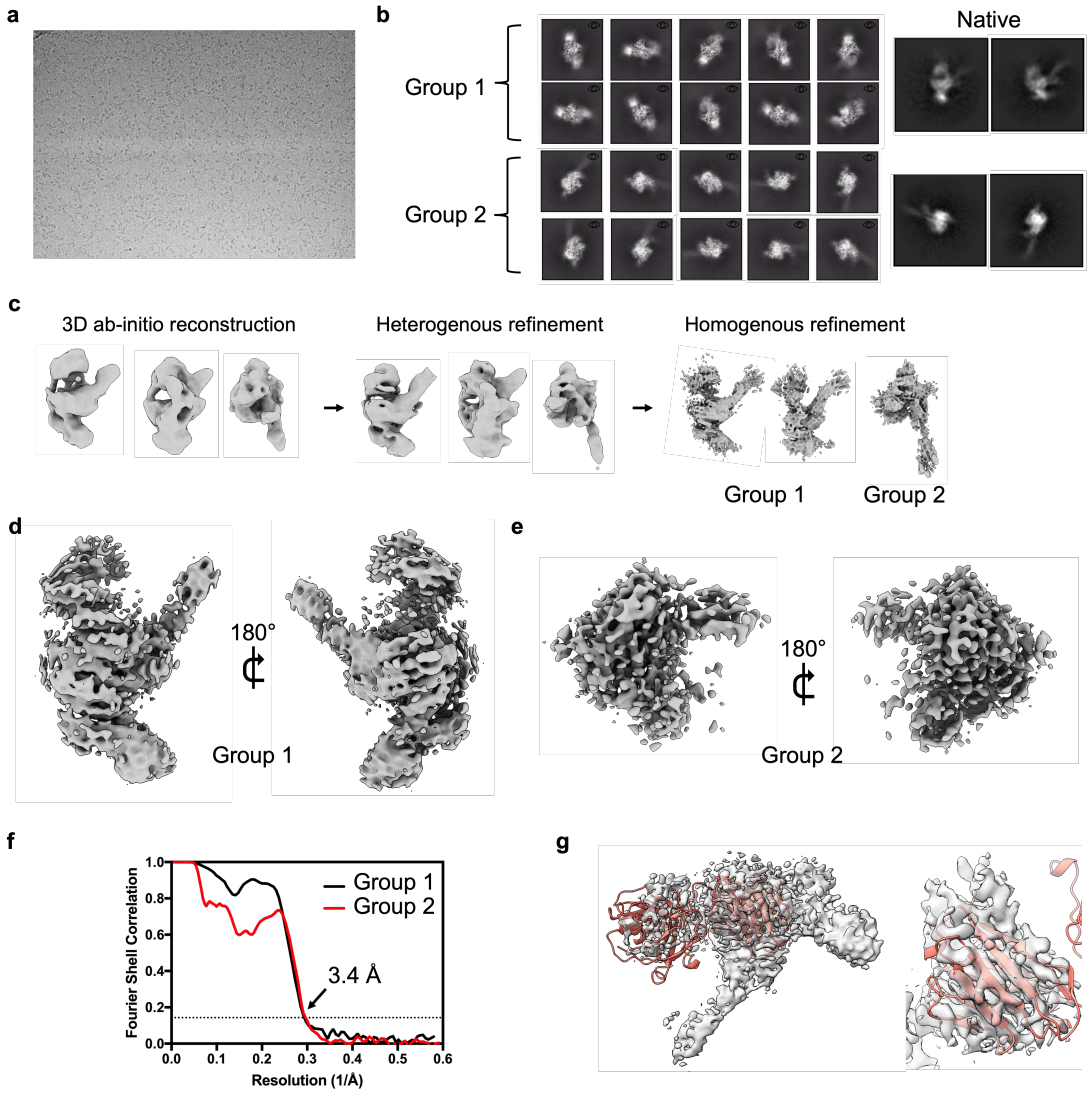


Figure 3.33: Cryo-EM studies of the C-terminal half of HgL in complex with CP33 Fab. a) Cryo-EM micrograph showing particles of HgL_20 in complex with CP33 Fab, collected at a magnification of 105,000 and a defocus of -1.5 µm. b) 2D classes generated after data collection (left, box size 200 × 200 pixels), with analogous classes from native sample (right, box size 256 × 256 pixels). Two groups can be distinguished. c) Pipeline for classification of particles. d) 3D reconstruction after non-uniform refinement of particles of group 1. e) 3D reconstruction after non-uniform refinement of particles of group 2. f) FSCs for d) and e). g) Fitting of CP33 Fab into the cryo-EM density of e).

3.4.3.6 Preliminary cryo-EM studies of the recombinant intermediate chain

IgL_01 was imaged by negative staining transmission electron microscopy and showed homogenous distribution of rather elongated particles (Figure 3.34.a). Cryo-EM datasets of untreated IgL_01 alone or bound to Fabs XEhI_X20 and XEhI_H2 (Figure 3.34.b-d) suffered from preferential orientation, which prevented generation of a reliable 3D ab-initio model (Figure 3.34.e).

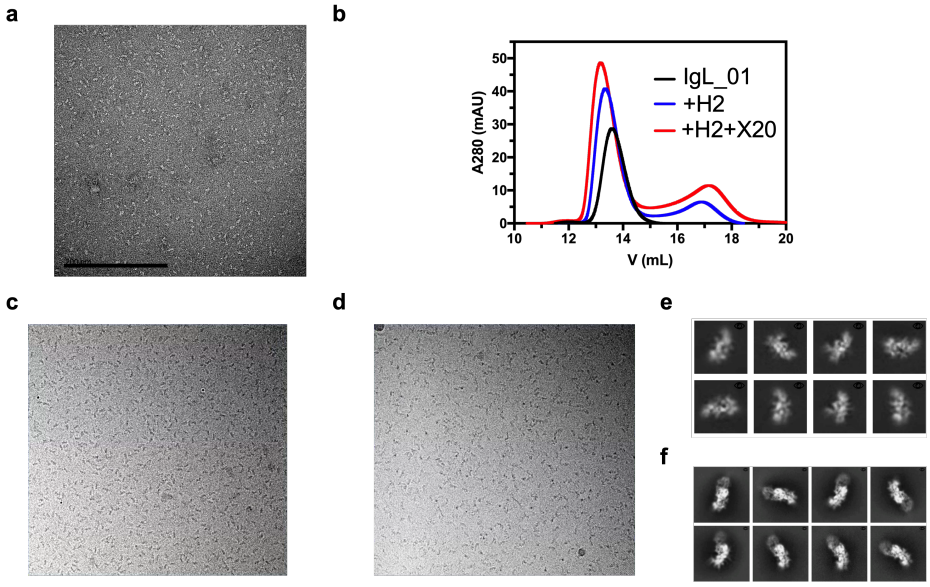


Figure 3.34: EM studies of IgL_01 with and without Fabs. a) SEC (S200 Increase 10/300 GL) curves of IgL_01 alone (in black) and complexed with XEhI_H2 (in blue) or XEhI_X20 and XEhI_H2 (in red). b) Electron micrograph of negatively stained IgL_01. Scale bar is equal to 200 nm. c) Cryo-EM micrograph of IgL_01, collected at a magnification of 165,000 and a defocus of -1.5 µm. d) Cryo-EM micrograph showing particles of IgL_01 bound to XEhI_X20 and XEhI_H2, collected at a magnification of 165,000 and a defocus of -1.5 µm. e) 2D classes of IgL_01 alone, with a box size of 144 × 144 pixels. f) 2D classes of Fab-bound IgL_01, viewed in SIMPLE with a box size of 200 × 200 pixels.

Then, surface properties of IgL_01 were altered through lysine methylation or PEGylation. PEGylation with PEG_{12} improved particle distribution in ice best, allowing generation of a 3D reconstruction at 4.7 Å (Figure 3.35.a,b,e,f). After combination with other datasets of IgL_01 alone, the volume was used as a reference to yield a reconstruction at 3.01 Å after non-uniform refinement (Figure 3.35.c,f). Density inspection suggests that the protein is rich in β-strands and also displays a protrusion that is less well resolved.

Likewise, using this volume as a reference, a map of IgL_01 bound to Fabs was generated at 2.85 Å, though anisotropy hampered map quality (Figure 3.35.d,f). Superimposition of reconstructions reveals extra density in the presence of Fabs, in which the variable domain of a Fab can be fitted (Figure 3.35.g). However, only one Fab could be fitted, suggesting that the other Fab either binds to the protrusion that is not well resolved, or that it collapses from the complex.

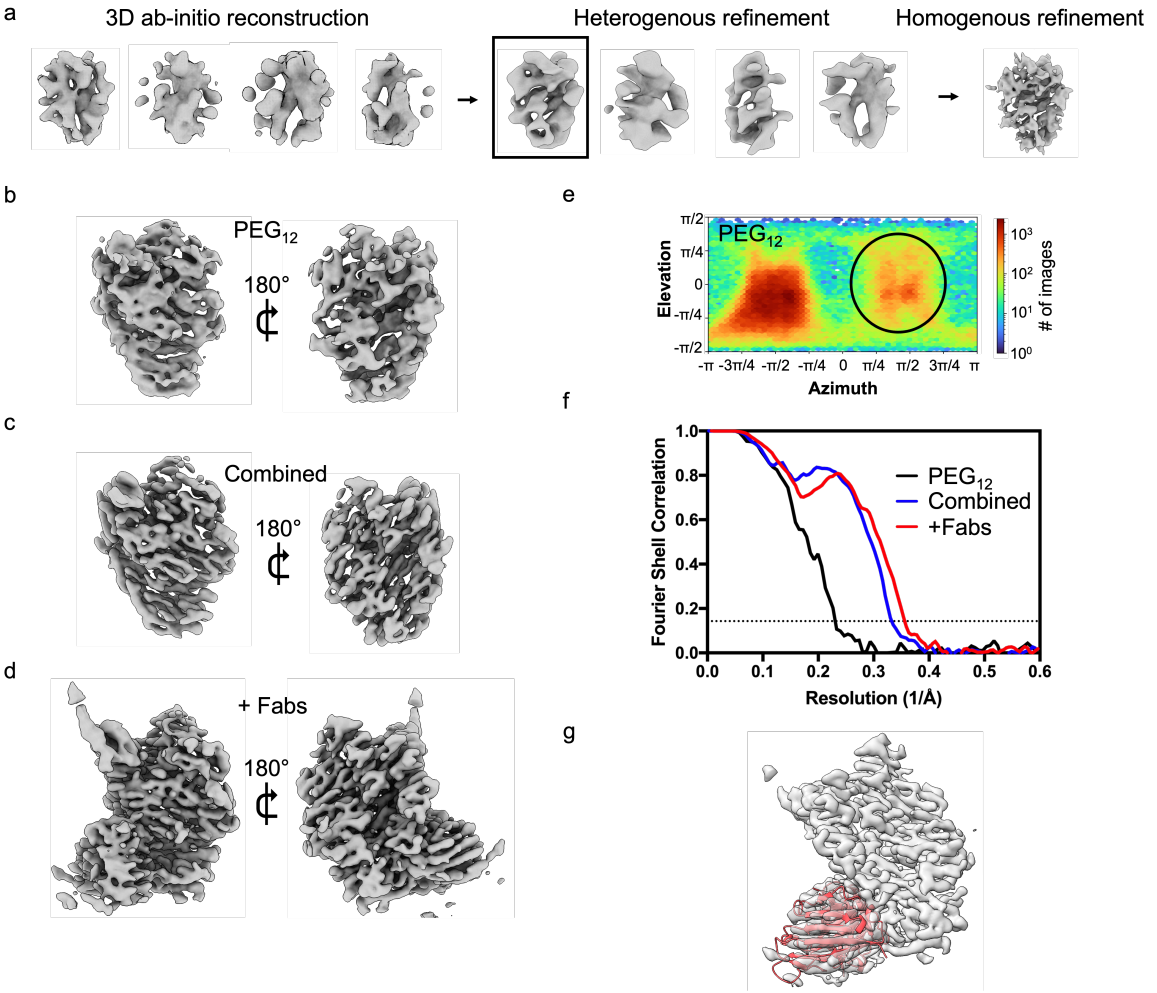


Figure 3.35: Cryo-EM study of IgL_01. a) Workflow for processing of PEGylated IgL_01. b) 3D reconstruction of PEGylated IgL_01, generated after non-uniform refinement, using the map obtained after homogenous refinement (shown in a)) as a reference volume. c) 3D reconstruction of combined particles of IgL_01, generated after non-uniform refinement. d) Electron density of the reconstruction of IgL_01 bound to Fabs, generated after non-uniform refinement. e) Orientation distribution map generated during refinement of PEGylated IgL_01. The unique contribution of PEG_{12} compared to other datasets is circled. f) FSC curves obtained after refinement of a dataset of PEGylated IgL_01, combined datasets of IgL_01, and a dataset of Fab-bound IgL_01. g) 3D reconstruction of IgL_01 bound to Fabs. The model of CP33 scFv is fitted and depicted as a ribbon (in salmon).

3.4.3.7 Functional studies of the 260-kDa antigen

3.4.3.7.1 ELISA binding assay of HgL VWD to collagen

Von Willebrand domains from von Willebrand factor and some integrins have been shown to interact with collagen [379]. In addition, *Entamoeba histolytica* 260-kDa antigen found in cholesterol-rich membrane domains was suggested to work as a collagen receptor, since trophozoite adhesion to collagen was inhibited by addition of Gal [303].

Therefore, binding assays using GST-fused HgL VWD (HgL_21, Figure 3.36.a-c) purified from *E. Coli* were carried out by ELISA against collagens I and III, the main types of collagen found in the intestines, and collagen IV, the primary collagen found in extracellular basement membranes [380]. All of these are also present in the liver. Monomeric pepsin-digested collagens, which are triple helices of collagen rather than fibres, were coated on plates. In these conditions, no binding was detected, whereas GST-tagged integrin α2 I domain bound to collagens in a cation-dependent manner, as shown before [381] (Figure 3.36.d). This suggests a lack of collagen binding by HgL VWD.

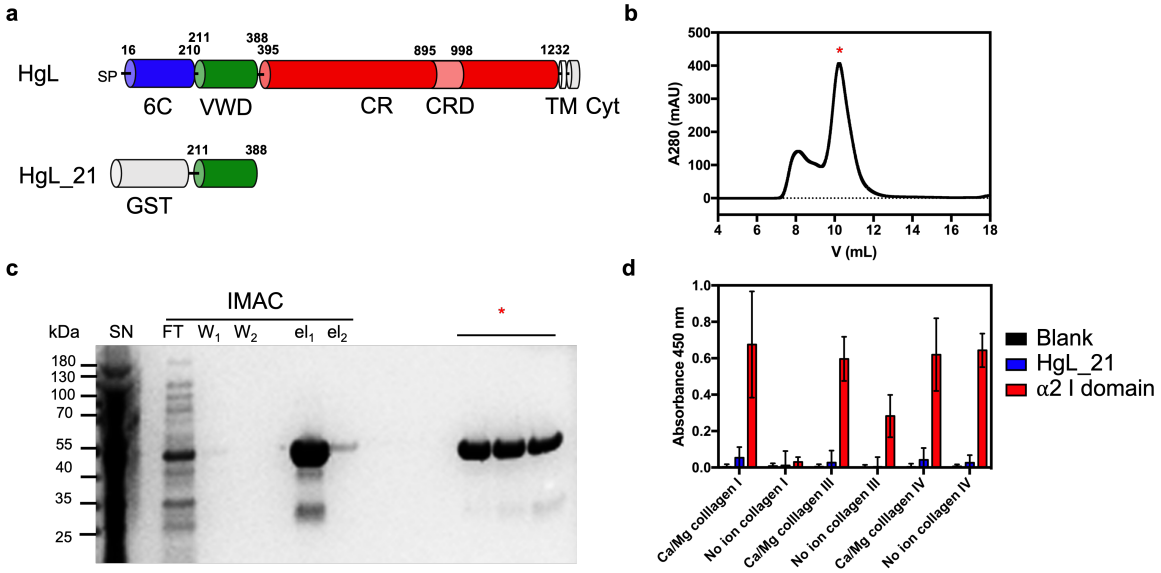


Figure 3.36: ELISA-type binding assay of HgL VWD to collagen. a) Schematic representation of the domains of HgL and the GST-fused VWD construct (HgL_21). b) SEC (S75 Increase 10/300 GL) curve of HgL_21 expressed in *E. Coli* and purified by IMAC. The red star indicates the protein fraction used for collagen binding. c) Reducing SDS-PAGE of each step of HgL_21 purification. *E. Coli* lysate (SN), IMAC fractions (flow through (FT), washes (W_1, W_2 and elutions (el_1, el_2)) and SEC fractions (red star) were loaded. The estimated molecular weight of HgL_21 is 46.1 kDa. d) ELISA-type collagen binding assay. Experiments were carried out in triplicates, with HgL_21 (in blue), the GST-fused I domain of α2 integrin (positive control, in red) and no GST-tagged protein (blank, in black). They were done in the presence or absence of 10 mM $CaCl_2$ and 2 mM $MgCl_2$ (Ca/Mg). Absorbance was detected at a wavelength of 450 nm.

3.4.3.7.2 Interaction of the native antigen with other proteins

Other proposed interacting partners include IgL and complement proteins C8 and C9. Binding to these proteins was investigated by SPR with immobilised native 260-kDa antigen. In these conditions, no binding was observed for C8 nor C9 (Figure 3.37.a). In fact, complement proteins were only reported to interact with the antigen purified

from human serum-adapted parasites. During adaptation, a different set of isoforms may be expressed by *Entamoeba histolytica*, which could explain why binding could not be recapitulated. Alternatively, other proteins could be present to act as a bridge.

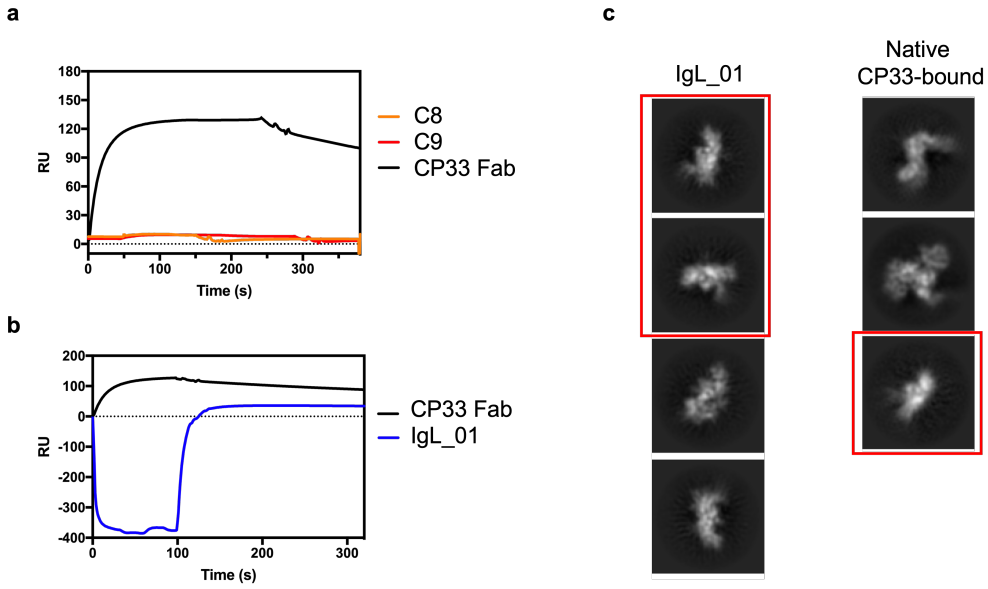


Figure 3.37: Interaction with complement proteins and IgL. a) SPR curves for binding of C8 and C9 at 1 μm to immobilised native 260-kDa antigen. CP33 at 100 nM is used as a positive control. b) SPR curves for binding of IgL_01 at 1 μm to immobilised native 260-kDa antigen. CP33 at 100 nM is used as a positive control. c) Some 2D classes obtained during cryo-EM processing of IgL_01 (left) and Fab-bound naive sample (right). Analogous projections are highlighted in red.

Using IgL_01 as an analyte, strong non-specific binding was measured, even after increase of salt and detergent concentrations in running buffers (Figure 3.37.b). Likewise, IgL_01 did not bind to HgL_05 nor recombinant HgL:LgL heterodimer during SEC.

Nevertheless, both isoforms of IgL were identified as contaminants during MS analysis of HgL from native samples (Appendix Table A.3). Furthermore, a minor 2D class identical to that of IgL_01 was obtained during cryo-EM processing of native 260-kDa antigen, when elution was performed by IMAC (Fab-bound) rather than at low pH (Figure 3.37.c). Therefore, interaction with IgL may occur with low affinity, in an isoform-dependent manner, and could require the membrane environment.

3.4.3.7.3 SPR binding of the native antigen to a panel of mAbs

Binding of LMNG-extracted 260-kDa native antigen to a panel of previously characterised antibodies (1G7, 3F4, H85 and 8A3) was studied by SPR. Of these, 8A3 and

3F4 enhance adherence of trophozoites to CHO cells and human colonic mucins, 1G7 inhibits adherence, and H85 increases parasite susceptibility to lysis. As expected, all of them bound to the purified 260-kDa antigen (Figure 3.38).

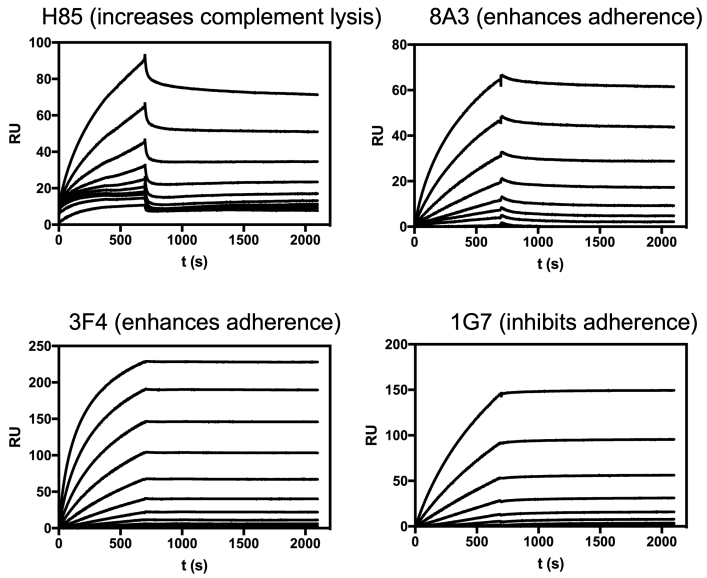


Figure 3.38: SPR of the 260-kDa antigen with a panel of mAbs. SPR curves are shown for four different antibodies (analytes) over immobilised native 260-kDa antigen. Series of dilutions are shown for antibody concentration starting at 100 nM with a dilution factor of 2 between consecutive runs.

3.4.3.7.4 SAXS studies of HgL

Distance distributions, normalised Kratky plots and radius of gyrations were assessed for HgL_05, HgL_18 and HgL_11 (Figure 3.38.a-c). The Kratky curve obtained for scattering of HgL_18 is bell-shaped, with a maximum of 1.18 for a sR_g of 1.96. This is close to the values obtained for compact globular proteins (maximum of 1.18 for a sR_g of 1.96) [382]. In addition, the distance distribution curve has a rather broad peak, with no apparent multiple bumps. The broad peak suggests the presence of at least two domains, while Kratky data indicate that the N-terminal construct is globular.

This is in agreement with cryo-EM data, in which the lipoprotein-like domain and the VWD form a globular core, while the presence of three pseudo domains in HgL_18 may bring some degree of flexibility.

In contrast, Kratky curves of HgL_05 and HgL_11 display broader peaks with tails. This is most apparent for HgL_11, for which bumps suggests multiple domains and

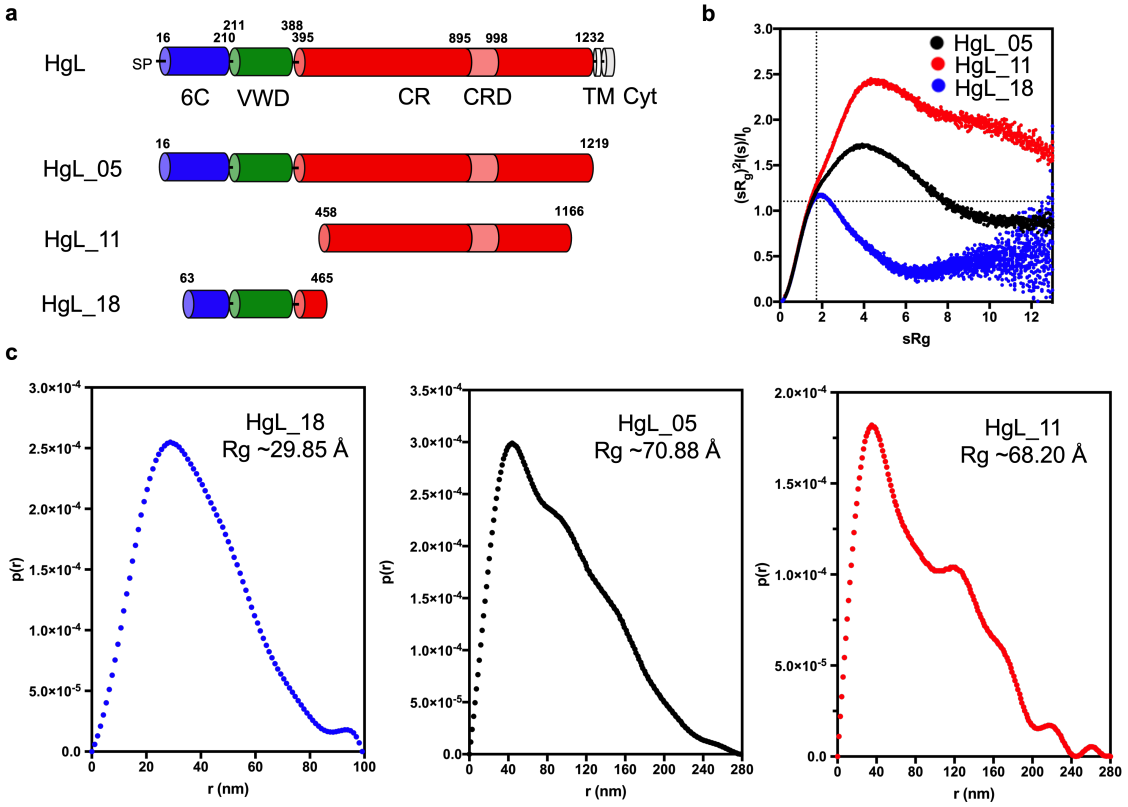


Figure 3.39: SAXS analysis of HgL. a) Schematic representation of the constructs studied by SAXS. b) Normalised Kratky plots obtained for scattering data of HgL_05 (in black), HgL_18 (in blue) and HgL_11 (in red). Dotted lines intersect at 1.104 for $sRg = 1.73$, which corresponds to the maximum value obtained for scattering data of globular proteins. c) Plots describing distance distributions $p(r)$ calculated from scattering data of HgL_05 (in black), HgL_18 (in blue) and HgL_11 (in red). An estimation of Rg is given for each protein fragment.

some degree of flexibility between them. Similarly, multiple bumps are observed in both distance distributions. Intriguingly, estimated Rg are similar for HgL_05 and HgL_11, suggesting that part of the amino-acids included in the C-terminus are involved in interactions with the N-terminus, making the molecule more globular. Indeed, residues 456 to 585 are included in the ring of pseudo domains, which is stabilised through interactions with the N-terminal domains. Without the N-terminus, these residues are likely to adopt an elongated conformation rather than that of a ring.

Altogether, these data highlight the C-terminus as an elongated protein, with flexible linkers between multiple domains. This could explain why protein fragments of this region failed to crystallise and appear unresolved in the cryo-EM reconstruction of native 260-kDa antigen.

3.5 Discussion and perspectives

3.5.1 The 260-kDa antigen has a distinctive architecture

The structure of the 260-kDa antigen is complex and unusual. The N-terminal half of HgL is composed of a lipoprotein-like domain, followed by a VWD and a cysteine-rich ring of pseudo-repeats that forms a pore. The N-terminal half of HgL associates with LgL through hydrophobic interactions and covalent linkage with the first pseudo-repeat. LgL adopts a β-trefoil fold, preceded by an unstructured stretch of about 60 residues.

Overall the partial heterodimer is globular with a distinctive arrangement of domains. For instance, the lipoprotein-like fold has never been found in multi-domain proteins. In addition, the ring conformation, though reminiscent to the architecture of the low density lipoprotein receptor, is unique. There, pseudo-repeats share a conserved novel fold, with cysteines involved in inter and intra-domain covalent bonds. Moreover, the VWD does not contain any MIDAS, which is present in most VWDs for cation-dependent binding to various ligands. Last, positions of β-strands in the β-trefoil domain of LgL are conserved with other structures of this family, however the connecting loops are longer and more complex.

Cryo-EM and SAXS experiments show that HgL C-terminus is elongated, rather flexible and contains multiple domains. This could explain why all attempts to induce crystallisation with constructs of this region failed. This region is important since it elicits antibodies that either inhibit or enhance carbohydrate binding of *Entamoeba histolytica* to target cells.

In particular, the putative CRD contained there mediates carbohydrate binding to Gal and GalNAc moieties presented by colonic mucins during invasion. Intriguingly, the protein shares no structural nor sequence homologies with other types of lectin. In fact, recent tools such as Alphafold and RosettaFOLD only modelled this region, most particularly residues 920 to 988, with very low confidence. Here, various recombinant fragments encompassing the putative CRD were expressed in mammalian cells but failed to crystallise, despite the use of antibody fragments, limited proteolysis or surface modification. Similarly, the restricted CRD was successfully refolded from inclusion

bodies but failed to crystallise. Nevertheless, this sample could be studied by nuclear magnetic resonance, though it appears to contain at least one reactive cysteine.

Preliminary cryo-EM studies of the C-terminal half of HgL bound to CP33 Fab yielded two maps, which both displayed long protrusions. However, these maps suffer from preferential orientation, therefore preventing accurate structure building. In future experiments, sample modification will be performed to alter particle distribution and generate 3D volumes with better angular accuracies.

3.5.2 Robust expression of the amoebic antigen in mammalian cells

Axenic cultures of parasites were successfully established, from which the 260-kDa antigen was extracted by Gal-affinity and immuno-affinity chromatographies. Notably, the purified sample contains a mix of isoforms. To avoid tedious parasite culture, heterodimers of HgL and LgL were co-expressed and purified for the first time in a relevant recombinant system. This is an important achievement since yields obtained are much higher than with native parasites. Furthermore, this system is more accessible as it does not involve culture of amoebae. This will pave the way for further characterisation of the 260-kDa antigen as a heterodimer instead of individual subunits and in experiments that require high amounts of sample.

Co-expression trials suggest that association of HgL with LgL is isoform-dependent. Indeed, isoform of HgL EHI_133900 only co-expressed with one out of three LgL isoforms (EHI_035690). This is reminiscent of integrin heterodimers and their associations between isoforms.

Interestingly, in solution, the full-length native and recombinant heterodimers appear to form higher order oligomers. The relevance of these assemblies remains unclear. However, similar behaviours between native and recombinant samples suggests that dimer formation is not induced by membrane domains, or interaction between heterodimers of different isoforms. Instead, an extracellular domain must induce dimerisation and remains to be characterised. Since full extracellular HgL only acts as a monomer in solution (Appendix Figure A.1), dimerisation appears be LgL-dependent.

Last, recombinant purification of LgL is challenging. Of all expression trials tested, only co-expression with HgL allowed purification of a folded protein. Nonetheless, the cryo-EM structure presented here will be valuable to guide construct design for further characterisation of LgL on its own. For instance, LgL could be purified after mutagenesis of the cysteine and hydrophobic residues that contact HgL [383].

3.5.3 Putative functions of the 260-kDa antigen

The 260-kDa antigen was first identified and is still referred as *Entamoeba histolytica* Gal/GalNAc lectin. However, carbohydrate binding is mostly mediated by relatively small domains, in this case the region defined as a CRD. This leaves approximately 1,500 residues without any attributed function. The structure presented here suggests that the 260-kDa antigen is composed of multiple domains of very different folds, which have been associated with different functions (Figure 3.40).

For instance, the lipoprotein-like domain structure indicates a potential binding site for lipids or hydrophobic molecules. Interestingly, *Entamoeba histolytica* lacks enzymes to synthesise cholesterol and fatty acids [384]. Instead, these are taken from the environment and transported through a partially understood process [385]. The 260-kDa antigen may be involved in this process through its N-terminal lipoprotein-like domain. However, this remains to be shown, and no density for a hydrophobic residue was found inside the cavity of this domain in the cryo-EM reconstruction. In the future, membrane strip assays may lead to identification of ligands for this domain.

The 260-kDa antigen also contains a VWD. Proteins that adopt this fold associate with a variety of extracellular ligands through this domain, such as collagen, platelet glycoprotein Ib-receptor, heparin, fibronectin, intercellular adhesion molecule-1. Among these, *Entamoeba histolytica* trophozoites bind to collagens, therefore binding of HgL VWD to collagen peptides was investigated by ELISA. However, in these conditions, no binding was detected. A screening for potential ligands that associate with this domain will be carried out in the future.

As for the pseudo-repeats of the cysteine-rich domain, it is unclear what their role is and why they arrange in this ring conformation. Nonetheless, this region

may be important for parasite stickiness, since antibodies against residues 436 to 624 there blocked host cell adherence and conferred some degree of protection against development ALA in gerbils [275].

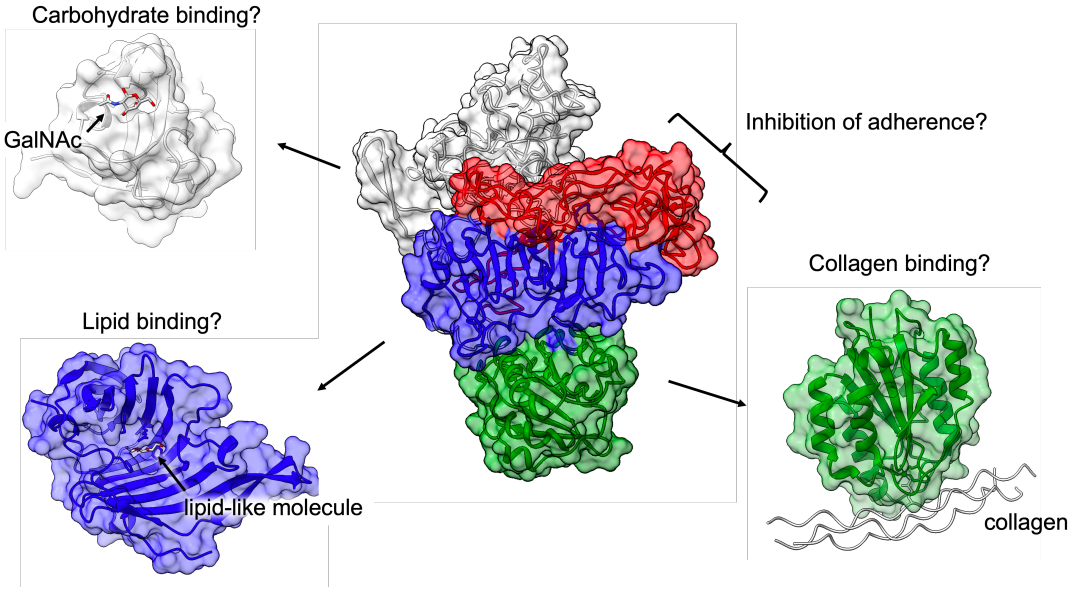


Figure 3.40: Putative functions of the 260-kDa antigen. The possible functions of domains identified by cryo-EM of the native antigen are highlighted. Structures of a β-trefoil domain bound to GalNAc (PDB code 4G9N), a lipoprotein-like domain bound to a lipid-like molecule (PDB code 3BMZ) and a VWD associated with a synthetic collagen triple helix (PDB code 1DZI) are shown.

In addition, the light chain adopts a β-trefoil fold, found in ricin-type lectins. These are known to bind carbohydrates, and many of them have been reported to co-crystallise with Gal and GalNAc. Since the 260-kDa antigen is a Gal/GalNAc lectin, LgL binding to these carbohydrates should be investigated. This has been hampered by challenging purification of LgL on its own. In the future, carbohydrate binding could be measured for recombinant heterodimers that lacks the C-terminal half of HgL. Glycan array could help screen most relevant carbohydrates and heterodimers of LgL bound to the N-terminal half of HgL could be co-crystallised with it. Alternatively, cryo-EM studies could be carried out in the presence of longer sugars.

Although the 15 C-terminal residues of the light chain have been suggested to be modified by a GPI anchor, density for these residues was present in the 3D reconstruction. In fact, these residues form the last β-strand of the well-conserved β-trefoil domain and interact with HgL. In trophozoites, a version of LgL truncated there formed covalently

linked homo-oligomers, failed to associate with HgL, and could not be co-purified by Gal-affinity chromatography [277]. From this observation and because antibodies against LgL did not inhibit adhesion to CHO cells [386], carbohydrate binding of the 260-kDa antigen was exclusively assigned to HgL.

Based on the cryo-EM structure, truncated LgL may have formed homo-oligomers through reactive cysteine 241 normally involved in intra-molecular covalent bond with truncated cysteine 274. This bond may be crucial to maintain β-trefoil integrity and to mediate carbohydrate binding. Furthermore, reported antibodies were raised against GST-fused LgL through an unclear protocol. Here, GST-fused LgL could be purified solubly from *E. Coli*, but cleavage of the GST tag resulted in loss of LgL, which suggests the protein was not folded properly (Appendix Figure A.2). Therefore, antibodies against LgL may not be functionally relevant and its ability to bind carbohydrates remains unclear.

Interestingly, expression of HgL with a truncated version of LgL that lacks the 55 N-terminal residues leads to reduced virulence and lectin clustering [387]. These residues, which exclude cysteines and the β-trefoil domain may therefore be involved in formation of higher order oligomers. This is supported by LgL-dependent oligomerisation described in section 3.4.1.5. However, in preliminary cryo-EM studies of the native 260-kDa antigen, no differences between 2D classes of both oligomeric species obtained after extraction were observed, suggesting oligomerisation occurs with low affinity. In the future, cross-linking experiments could be performed to trap higher order assemblies for imaging.

The C-terminal half of HgL could not be built as it is part of a poorly resolved elongated protrusion that extends away from the globular N-terminal half. 2D classification of the native antigen, together with preliminary cryo-EM and SAXS studies of HgL C-terminus indicate elongation and flexibility in the arrangement of these domains.

Carbohydrate binding has been exclusively attributed to the putative CRD found in the C-terminal half of HgL. CP33 antibody inhibits trophozoite adherence to CHO cells through the CRD. Interestingly, a 27 stretch of residues there (residues 946 to 973) is necessary for CP33 binding to the 260-kDa antigen. However, restricted refolded CRD (residues 895 to 998) does not bind to CP33 on SEC nor by SPR (data not

shown), which suggests the importance of other residues N-terminal to the CRD and that CP33 may bind to a conformational epitope.

Besides, further experiments will dissect interactions of the 260-kDa antigen with its receptors. Though, IgL was co-purified with the 260-kDa antigen during extraction from parasites, no density for IgL in cryo-EM studies of the 260-kDa antigen was observed, and recombinant IgL did not associate with recombinant HgL nor HgL:LgL heterodimer on SEC.

This interaction may be isoform-dependent. Alternatively, IgL could be less abundantly expressed and the interaction may occur with low affinity, which could explain why the complex dissociates on SEC. Moreover, membrane environment might be crucial for this interaction. In the future, parasite lysates will be incubated with IgL_01 or IgL-targeting antibodies, followed by IMAC to assess enrichment of a bigger complex.

Regardless, recombinant IgL is suitable for cryo-EM studies. The protein is cysteine-rich and appears to contain many β-strands which arrange as a flat and rather elongated protein. Future work on IgL will involve de novo structure building and characterisation of ligand binding. In addition, binding of the 260-kDa antigen to other ligands, such as c-Met, TLR-4 and complement proteins will be investigated.

References

[250] OMS. “World Health Report Life in the 21st century A vision for all Report of the Director-General”. In: *51st World Health Assembly* (1998), pp. 241–241.

[251] J. A. Walsh. “Problems in recognition and diagnosis of amebiasis: Estimation of the global magnitude of morbidity and mortality”. In: *Reviews of Infectious Diseases* 8.2 (1986), pp. 228–238.

[252] D. Mirelman. “Effect of culture conditions and bacterial associates on the zymodemes of Entamoeba histolytica”. In: *Parasitology Today* 3.2 (1987), pp. 37–40.

[253] S. L. Stanley and S. L. Reed. “Microbes and microbial toxins: Paradigms for microbial-mucosal interactions. VI. Entamoeba histolytica: Parasite-host interactions”. In: *American Journal of Physiology - Gastrointestinal and Liver Physiology* 280.6 43-6 (2001).

[254] J. C. Carrero et al. “Intestinal amoebiasis: 160 years of its first detection and still remains as a health problem in developing countries”. In: *International Journal of Medical Microbiology* 310.1 (2020), p. 151358.

[255] A. Dance. “Cells nibble one another via the under-appreciated process of trogocytosis”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.36 (2019), pp. 17608–17610.

[256] S. L. Stanley. “Amoebiasis”. In: *Lancet* 361.9362 (2003), pp. 1025–1034.

[257] K. S. Ralston et al. “Trogocytosis by Entamoeba histolytica contributes to cell killing and tissue invasion”. In: *Nature* 508.7497 (2014), pp. 526–530.

[258] S. Gupta, L. Smith, and A. Diakiw. “Amebiasis and Amebic Liver Abscess in Children”. In: *Pediatric Clinics of North America* 69.1 (2022), pp. 79–97.

[259] J. I. Ravdin and R. L. Guerrant. “Role of adherence in cytopathogenic mechanisms of Entamoeba histolytica. Study with mammalian tissue culture cells and human erythrocytes”. In: *Journal of Clinical Investigation* 68.5 (1981), pp. 1305–1313.

[260] J. I. Ravdin et al. “Production of mouse monoclonal antibodies which inhibit in vitro adherence of Entamoeba histolytica trophozoites”. In: *Infection and Immunity* 53.1 (1986), pp. 1–5.

[261] W. A. Petri et al. “Isolation of the galactose-binding lectin that mediates the in vitro adherence of Entamoeba histolytica”. In: *Journal of Clinical Investigation* 80.5 (1987), pp. 1238–1244.

[262] W. A. Petri et al. “Subunit structure of the galactose and N-acetyl-D-galactosamine-inhibitable adherence lectin in Entamoeba histolytica”. In: *Journal of Biological Chemistry* 264.5 (1989), pp. 3007–3012.

[263] G. Ramakrishnan et al. “Physical mapping and expression of gene families encoding the N-acetyl D-galactosamine adherence lectin of Entamoeba histolytica”. In: *Molecular Microbiology* 19.1 (1996), pp. 91–100.

[264] B. J. Mann et al. “Sequence of a cysteine-rich galactose-specific lectin of Entamoeba histolytica.” In: *Proceedings of the National Academy of Sciences of the United States of America* 88.8 (1991), pp. 3248–3252.

[265] E. Tannich, F. Ebert, and R. D. Horstmann. “Primary structure of the 170-kDa surface lectin of pathogenic Entamoeba histolytica.” In: *Proceedings of the National Academy of Sciences* 88.5 (1991), pp. 1849–1853.

[266] J. J. McCoy et al. “Structural analysis of the light subunit of the Entamoeba histolytica galactose-specific adherence lectin”. In: *Journal of Biological Chemistry* 268.32 (1993), pp. 24223–24231.

[267] P. Adler et al. “High affinity binding of the Entamoeba histolytica lectin to polyvalent n-acetylgalactosaminides”. In: *Journal of Biological Chemistry* 270.10 (1995), pp. 5164–5171.

[268] D. Yi et al. “Substructural specificity and polyvalent carbohydrate recognition by the Entamoeba histolytica and rat hepatic N-acetylgalactosamine/galactose lectins”. In: *Glycobiology* 8.10 (1998), pp. 1037–1043.

[269] K. Chadee et al. “Rat and human colonic mucins bind to and inhibit adherence lectin of Entamoeba histolytica”. In: *Journal of Clinical Investigation* 80.5 (1987), pp. 1245–1254.

[270] K. Chadee, C. Ndarathi, and K. Keller. “Binding of proteolytically-degraded human colonic mucin glycoproteins to the Gal/GalNAc adherence lectin of Entamoeba histolytica”. In: *Gut* 31.8 (1990), pp. 890–895.

[271] K. A. Thomsson et al. “Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2”. In: *Glycobiology* 22.8 (2012), pp. 1128–1139.

[272] J. M. H. Larsson et al. “A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn”. In: *Glycobiology* 19.7 (2009), pp. 756–766.

[273] J. M. Dodson et al. “Infection and immunity mediated by the carbohydrate recognition domain of the Entamoeba histolytica Gal/GalNAc lectin”. In: *Journal of Infectious Diseases* 179.2 (1999), pp. 460–466.

[274] R. Yadav et al. “Biophysical studies on calcium and carbohydrate binding to carbohydrate recognition domain of Gal/GalNAc lectin from Entamoeba histolytica: Insights into host cell adhesion”. In: *Journal of Biochemistry* 160.3 (2016), pp. 177–186.

[275] H. Lotter et al. “Identification of an epitope on the Entamoeba histolytica 170-kD lectin conferring antibody-mediated protection against invasive amebiasis”. In: *Journal of Experimental Medicine* 185.10 (1997), pp. 1793–1801.

[276] W. A. Petri et al. “Monoclonal antibodies directed against the galactose-binding lectin of Entamoeba histolytica enhance adherence”. In: *Journal of Immunology* 144.12 (1990), pp. 4803–4809.

[277] G. Ramakrishnan et al. “Entamoeba histolytica: Deletion of the GPI anchor signal sequence on the Gal/GalNAc lectin light subunit prevents its assembly into the lectin heterodimer”. In: *Experimental Parasitology* 96.1 (2000), pp. 57–60.

[278] W. A. Petri, R. Haque, and B. J. Mann. “The Bittersweet Interface of Parasite and Host: Lectin-Carbohydrate Interactions During Human Invasion by the Parasite Entamoeba histolytica”. In: *Annual Review of Microbiology* 56.1 (2002), pp. 39–64.

[279] B. J. Mann et al. “Neutralizing monoclonal antibody epitopes of the Entamoeba histolytica galactose adhesin map to the cysteine-rich extracellular domain of the 170-kilodalton subunit”. In: *Infection and Immunity* 61.5 (1993), pp. 1772–1778.

[280] H. Tachibana et al. “VH3 gene usage in neutralizing human antibodies specific for the Entamoeba histolytica Gal/GalNAc lectin heavy subunit”. In: *Infection and Immunity* 71.8 (2003), pp. 4313–4319.

[281] R. R. Vines et al. “Regulation of adherence and virulence by the Entamoeba histolytica lectin cytoplasmic domain, which contains a $\beta 2$ integrin motif”. In: *Molecular Biology of the Cell* 9.8 (1998), pp. 2069–2079.

[282] M. Shimaoka, J. Takagi, and T. A. Springer. “Conformational regulation of integrin structure and function”. In: *Annual Review of Biophysics and Biomolecular Structure* 31 (2002), pp. 485–516.

[283] E. Li, A. Becker, and S. L. Stanley. “Chinese hamster ovary cells deficient in N-acetylglucosaminyltransferase I activity are resistant to Entamoeba histolytica-mediated cytotoxicity”. In: *Infection and Immunity* 57.1 (1989), pp. 8–12.

[284] L. D. Saffer and W. A. Petri. “Role of the galactose lectin of Entamoeba histolytica in adherence-dependent killing of mammalian cells.” In: *Infection and Immunity* 59.12 (1991), pp. 4681–4683.

[285] A. Bettadapur, H. W. Miller, and K. S. Ralston. “Biting off what can be chewed: Trogocytosis in health, infection, and disease”. In: *Infection and Immunity* 88.7 (2020).

[286] S. L. Reed et al. “Activation of complement by pathogenic and nonpathogenic Entamoeba histolytica”. In: *J Immunol* 136.6 (1986), pp. 2265–2270.

[287] S. L. Reed and I. Gigli. “Lysis of complement-sensitive Entamoeba histolytica by activated terminal complement components. Initiation of complement activation by an extracellular neutral cysteine proteinase”. In: *Journal of Clinical Investigation* 86.6 (1990), pp. 1815–1822.

[288] L. L. Braga et al. “Inhibition of the complement membrane attack complex by the galactose-specific adhesin of entamoeba histolytica”. In: *Journal of Clinical Investigation* 90.3 (1992), pp. 1131–1137.

[289] X. J. Cheng et al. “Identification of the 150-kDa surface antigen of Entamoeba histolytica as a galactose- and N-acetyl-D-galactosamine-inhibitable lectin”. In: *Parasitology Research* 84.8 (1998), pp. 632–639.

[290] H. Tachibana et al. “Differentiation of Entamoeba histolytica from E. dispar facilitated by monoclonal antibodies against a 150-kDa surface antigen”. In: *Parasitology Research* 83.5 (1997), pp. 435–439.

[291] X. J. Cheng et al. “Intermediate subunit of the Gal/GalNAc lectin of Entamoeba histolytica is a member of a gene family containing multiple CXXC sequence motifs”. In: *Infection and Immunity* 69.9 (2001), pp. 5892–5898.

[292] D. L. Beck et al. “Identification and gene expression analysis of a large family of transmembrane kinases related to the Gal/GalNAc lectin in Entamoeba histolytica”. In: *Eukaryotic Cell* 4.4 (2005), pp. 722–732.

[293] V. I. Hernández-Ramírez et al. “Subcellular distribution of the Entamoeba histolytica 140 kDa FN-binding molecule during host-parasite interaction”. In: *Parasitology* 134.2 (2007), pp. 169–177.

[294] X. J. Cheng, H. Tachibana, and Y. Kaneda. “Protection of hamsters from amebic liver abscess formation by a monoclonal antibody to a 150-kDa surface lectin of Entamoeba histolytica”. In: *Parasitology Research* 85.1 (1999), pp. 78–80.

[295] X. J. Cheng and H. Tachibana. “Protection of hamsters from amebic liver abscess formation by immunization with the 150- and 170-kDa surface antigens of Entamoeba histolytica”. In: *Parasitology Research* 87.2 (2001), pp. 126–130.

[296] H. Tachibana et al. “Characterization of Entamoeba histolytica intermediate subunit lectin-specific human monoclonal antibodies generated in transgenic mice expressing human immunoglobulin loci”. In: *Infection and Immunity* 77.1 (2009), pp. 549–556.

[297] K. Kato et al. “Novel hemagglutinating, hemolytic and cytotoxic activities of the intermediate subunit of Entamoeba histolytica lectin”. In: *Scientific Reports* 5 (2015).

[298] X. Min et al. “Evaluation of the C-Terminal Fragment of Entamoeba histolytica Gal/GalNAc Lectin Intermediate Subunit as a Vaccine Candidate against Amebic Liver Abscess”. In: *PLoS Neglected Tropical Diseases* 10.1 (2016), pp. 1–17.

[299] S. J. Kammanadiminti et al. “Regulation of Toll-like receptor-2 expression by the Gal-lectin of Entamoeba histolytica.” In: *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 18.1 (2004), pp. 155–157.

[300] J. M. Galván-Moroyoqui, M. del Carmen Domínguez-Robles, and I. Meza. “Pathogenic bacteria prime the induction of Toll-like receptor signalling in human colonic cells by the Gal/GalNAc lectin Carbohydrate Recognition Domain of Entamoeba histolytica”. In: *International Journal for Parasitology* 41.10 (2011), pp. 1101–1112.

[301] C. G. Huh et al. “Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair”. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.13 (2004), pp. 4477–4482.

[302] M. Borowiak et al. “Met provides essential signals for liver regeneration”. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.29 (2004), pp. 10608–10613.

[303] K. Mittal, B. H. Welter, and L. A. Temesvari. “Entamoeba histolytica: Lipid rafts are involved in adhesion of trophozoites to host extracellular matrix components”. In: *Experimental Parasitology* 120.2 (2008), pp. 127–134.

[304] A. M. Goldston et al. “Exposure to host ligands correlates with colocalization of Gal/GalNAc lectin subunits in lipid rafts and phosphatidylinositol (4,5)-bisphosphate signaling in Entamoeba histolytica”. In: *Eukaryotic Cell* 11.6 (2012), pp. 743–751.

[305] M. D. Abd-Alla, A. M. El-Hawey, and J. I. Ravdin. “Use of an enzyme-linked immunosorbent assay to detect anti-adherence protein antibodies in sera of patients with invasive amebiasis in Cairo, Egypt”. In: *American Journal of Tropical Medicine and Hygiene* 47.6 (1992), pp. 800–804.

[306] J. I. Ravdin et al. "Association of serum antibodies to adherence lectin with invasive amebiasis and asymptomatic infection with pathogenic Entamoeba histolytica". In: *Journal of Infectious Diseases* 162.3 (1990), pp. 768–772.

[307] C. J. G. Soong et al. "A recombinant cysteine-rich section of the entamoeba histolytica galactoseinhibitable lectin is efficacious as a subunit vaccine in the gerbil model of amebic liver abscess". In: *Journal of Infectious Diseases* 171.3 (1995), pp. 645–651.

[308] R. Haque et al. "Amebiasis and Mucosal IgA Antibody against the Entamoeba histolytica Adherence Lectin in Bangladeshi Children". In: *The Journal of Infectious Diseases* 183.12 (2001), pp. 1787–1793.

[309] W. A. Petri and J. A. Ravdin. "Protection of gerbils from amebic liver abscess by immunization with the galactose-specific adherence lectin of Entamoeba histolytica". In: *Infection and Immunity* 59.1 (1991), pp. 97–101.

[310] E. Houpt et al. "Prevention of intestinal amebiasis by vaccination with the Entamoeba histolytica Gal/GalNac lectin". In: *Vaccine* 22.5-6 (2004), pp. 611–617.

[311] C. P. A. Ivory and K. Chadee. "Intranasal immunization with Gal-inhibitable lectin plus an adjuvant of CpG oligodeoxynucleotides protects against Entamoeba histolytica challenge". In: *Infection and Immunity* 75.10 (2007), pp. 4918–4922.

[312] D. Gaucher and K. Chadee. "Construction and immunogenicity of a codon-optimized Entamoeba histolytica Gal-lectin-based DNA vaccine". In: *Vaccine* 20.27-28 (2002), pp. 3244–3253.

[313] H. Lotter et al. "Oral vaccination with recombinant Yersinia enterocolitica expressing hybrid type III proteins protects gerbils from amebic liver abscess". In: *Infection and Immunity* 72.12 (2004), pp. 7318–7321.

[314] B. J. Mann, B. V. Burkholder, and L. A. Lockhart. "Protection in a gerbil model of amebiasis by oral immunization with Salmonella expressing the galactose/N-acetyl D-galactosamine inhibitable lectin of Entamoeba histolytica". In: *Vaccine* 15.6-7 (1997), pp. 659–663.

[315] T. Zhang and S. L. Stanley. "Protection of gerbils from amebic liver abscess by immunization with a recombinant protein derived from the 170-kilodalton surface adhesin of Entamoeba histolytica". In: *Infection and Immunity* 62.6 (1994), pp. 2605–2608.

[316] S. L. Martínez-Hernández et al. "Evaluation of the PEΔIII-LC3-KDEL3 chimeric protein of Entamoeba histolytica-lectin as a vaccine candidate against amebic liver abscess". In: *Journal of Immunology Research* 2021 (2021).

[317] M. M. Abhyankar et al. "Adjuvant composition and delivery route shape immune response quality and protective efficacy of a recombinant vaccine for Entamoeba histolytica". In: *npj Vaccines* 3.1 (2018).

[318] H. Lotter et al. "Protection of gerbils from amebic liver abscess by vaccination with a 25-mer peptide derived from the cysteine-rich region of Entamoeba histolytica galactose-specific adherence lectin". In: *Infection and Immunity* 68.8 (2000), pp. 4416–4421.

[319] M. D. Abd Alla et al. "Efficacy of a Gal-lectin subunit vaccine against experimental Entamoeba histolytica infection and colitis in baboons (Papio sp.)" In: *Vaccine* 30.20 (2012), pp. 3068–3075.

[320] R. Séguin et al. “Identification of the galactose-adherence lectin epitopes of Entamoeba histolytica that stimulate tumor necrosis factor-α production by macrophages”. In: *Proceedings of the National Academy of Sciences of the United States of America* 92.26 (1995), pp. 12175–12179.

[321] D. C. Schain, R. A. Salata, and J. I. Ravdin. “Human T-lymphocyte proliferation, lymphokine production, and amebicidal activity elicited by the galactose-inhibitable adherence protein of Entamoeba histolytica”. In: *Infection and Immunity* 60.5 (1992), pp. 2143–2146.

[322] C. P. A. Ivory and K. Chadee. “Activation of dendritic cells by the Gal-lectin of Entamoeba histolytica drives Th1 responses in vitro and in vivo”. In: *European Journal of Immunology* 37.2 (2007), pp. 385–394.

[323] X. Guo et al. “CD4+ and CD8+ T cell- and IL-17-mediated protection against Entamoeba histolytica induced by a recombinant vaccine”. In: *Vaccine* 29.4 (2011), pp. 772–777.

[324] X. Guo et al. “Protection against intestinal amebiasis by a recombinant vaccine is transferable by T cells and mediated by gamma interferon”. In: *Infection and Immunity* 77.9 (2009), pp. 3909–3918.

[325] M. M. Abhyankar et al. “Nanoformulation of synergistic TLR ligands to enhance vaccination against Entamoeba histolytica”. In: *Vaccine* 35.6 (2017), pp. 916–922.

[326] J. J. Almagro Armenteros et al. “SignalP 5.0 improves signal peptide predictions using deep neural networks”. In: *Nature Biotechnology* 37.4 (2019), pp. 420–423.

[327] S. Möller, M. D. R. Croning, and R. Apweiler. “Evaluation of methods for the prediction of membrane spanning regions”. In: *Bioinformatics* 17.7 (2001), pp. 646–653.

[328] A. Pierleoni, P. Martelli, and R. Casadio. “PredGPI: A GPI-anchor predictor”. In: *BMC Bioinformatics* 9.1 (2008), pp. 1–11.

[329] B. Xue et al. “PONDR-FIT: A meta-predictor of intrinsically disordered amino acids”. In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1804.4 (2010), pp. 996–1010.

[330] R. Gupta and S. Brunak. “Prediction of glycosylation across the human proteome and the correlation to protein function.” In: *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (2002), pp. 310–322.

[331] L. A. Kelley et al. “The Phyre2 web portal for protein modeling, prediction and analysis”. In: *Nature Protocols* 10.6 (2015), pp. 845–858.

[332] J. Yang et al. “The I-TASSER suite: Protein structure and function prediction”. In: *Nature Methods* 12.1 (2014), pp. 7–8.

[333] A. Waterhouse et al. “SWISS-MODEL: Homology modelling of protein structures and complexes”. In: *Nucleic Acids Research* 46.W1 (2018), W296–W303.

[334] Z. Du et al. “The trRosetta server for fast and accurate protein structure prediction”. In: *Nature Protocols* 16.12 (2021), pp. 5634–5651.

[335] M. Baek et al. “Accurate prediction of protein structures and interactions using a three-track neural network”. In: *Science* 373.6557 (2021), pp. 871–876.

[336] J. Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589.

[337] A. R. Aricescu, W. Lu, and E. Y. Jones. "A time- and cost-efficient system for high-level protein production in mammalian cells". In: *Acta Crystallographica Section D: Biological Crystallography* 62.10 (2006), pp. 1243–1250.

[338] H. Tachibana et al. "Bacterial expression of a neutralizing mouse monoclonal antibody Fab fragment to a 150-kilodalton surface antigen of Entamoeba histolytica." In: *The American journal of tropical medicine and hygiene* 60.1 (1999), pp. 35–40.

[339] J. E. Nettleship et al. "A pipeline for the production of antibody fragments for structural studies using transient expression in HEK 293T cells". In: *Protein Expression and Purification* 62.1 (2008), pp. 83–89.

[340] J. Andris-Widhopf et al. "Generation of human scFv antibody libraries: PCR amplification and assembly of light- and heavy-Chain coding sequences". In: *Cold Spring Harbor Protocols* 6.9 (2011), pp. 1139–1150.

[341] A. D. Elbein et al. "Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I". In: *Journal of Biological Chemistry* 265.26 (1990), pp. 15599–15605.

[342] T. S. Walter et al. "Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization". In: *Structure* 14.11 (2006), pp. 1617–1622.

[343] K. Manalastas-Cantos et al. "ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis". In: *Journal of Applied Crystallography* 54.Pt 1 (2021), pp. 343–355.

[344] G. Winter, C. M. C. Lobley, and S. M. Prince. "Decision making in xia2". In: *Acta Crystallographica Section D: Biological Crystallography* 69.7 (2013), pp. 1260–1273.

[345] G. Winter et al. "DIALS: Implementation and evaluation of a new integration package". In: *Acta Crystallographica Section D: Structural Biology* 74 (2018), pp. 85–97.

[346] P. R. Evans and G. N. Murshudov. "How good are my data and what is the resolution?" In: *Acta Crystallographica Section D: Biological Crystallography* 69.7 (2013), pp. 1204–1214.

[347] L. Potterton et al. "CCP4i2: The new graphical user interface to the CCP4 program suite". In: *Acta Crystallographica Section D: Structural Biology* 74 (2018), pp. 68–84.

[348] A. T. Brünger. "Assessment of phase accuracy by cross validation: the free R value. Methods and applications". In: *Acta Crystallographica Section D Biological Crystallography* 49.1 (1993), pp. 24–36.

[349] A. J. McCoy et al. "Phaser crystallographic software". In: *Journal of Applied Crystallography* 40.4 (2007), pp. 658–674.

[350] K. Cowtan. "The Buccaneer software for automated model building. 1. Tracing protein chains". In: *Acta Crystallographica Section D: Biological Crystallography* 62.9 (2006), pp. 1002–1011.

[351] E. F. Pettersen et al. "UCSF Chimera - A visualization system for exploratory research and analysis". In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612.

[352] P. Emsley and K. Cowtan. "Coot: Model-building tools for molecular graphics". In: *Acta Crystallographica Section D: Biological Crystallography* 60.12 I (2004), pp. 2126–2132.

[353] P. V. Afonine et al. “Real-space refinement in PHENIX for cryo-EM and crystallography”. In: *Acta Crystallographica Section D: Structural Biology* 74.6 (2018), pp. 531–544.

[354] C. J. Williams et al. “MolProbity: More and better reference data for improved all-atom structure validation”. In: *Protein Science* 27.1 (2018), pp. 293–315.

[355] J. Caesar et al. “SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time”. In: *Journal of Structural Biology: X* 4 (2020).

[356] J. Zivanov et al. “New tools for automated high-resolution cryo-EM structure determination in RELION-3”. In: *eLife* 7 (2018).

[357] Ruben Sanchez-Garcia et al. “DeepEMhancer: a deep learning solution for cryo-EM volume post-processing”. In: *Communications Biology* 4.1 (2021).

[358] A. Punjani et al. “CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination”. In: *Nature Methods* 14.3 (2017), pp. 290–296.

[359] A. Punjani, H. Zhang, and D. J. Fleet. “Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction”. In: *Nature Methods* 17.12 (2020), pp. 1214–1221.

[360] L. Holm and L. M. Laakso. “Dali server update”. In: *Nucleic acids research* 44.W1 (2016), W351–W355.

[361] J. Pei, M. Tang, and N. V. Grishin. “PROMALS3D web server for accurate multiple protein sequence and structure alignments.” In: *Nucleic acids research* 36.Web Server issue (2008).

[362] H. Singh, S. Singh, and G. P. S. Raghava. “In silico platform for predicting and initiating β-turns in a protein at desired locations”. In: *Proteins: Structure, Function and Bioinformatics* 83.5 (2015), pp. 910–921.

[363] M. A. Wouters et al. “Evolution of distinct EGF domains with specific functions”. In: *Protein Science* 14.4 (2005), pp. 1091–1103.

[364] L. Griffin and A. Lawson. “Antibody fragments as tools in crystallography”. In: *Clinical and Experimental Immunology* 165.3 (2011), pp. 285–291.

[365] A. Colombatti, P. Bonaldo, and R. Doliana. “Type A Modules: Interacting Domains Found in Several Non-Fibrillar Collagens and in Other Extracellular Matrix Proteins”. In: *Matrix* 13.4 (1993), pp. 297–306.

[366] J. J. Dumas et al. “Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Ibα complex reveals conformation differences with a complex bearing von Willebrand disease mutations”. In: *Journal of Biological Chemistry* 279.22 (2004), pp. 23327–23334.

[367] E. G. Huizinga et al. “Structures of glycoprotein Ibα and its complex with von Willebrand factor A1 domain”. In: *Science* 297.5584 (2002), pp. 1176–1179.

[368] C. G. Clark and L. S. Diamond. “Methods for cultivation of luminal parasitic protists of clinical importance”. In: *Clinical Microbiology Reviews* 15.3 (2002), pp. 329–341.

[369] A. Blees et al. “Structure of the human MHC-I peptide-loading complex”. In: *Nature* 551.7681 (2017), pp. 525–528.

[370] R. G. Efremov et al. “Architecture and conformational switch mechanism of the ryanodine receptor”. In: *Nature* 517.7532 (2015), pp. 39–43.

[371] Z. Zhang and J. Chen. "Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator". In: *Cell* 167.6 (2016), 1586–1597.e9.

[372] C. Gatsogiannis et al. "Membrane insertion of a Tc toxin in near-atomic detail". In: *Nature Structural and Molecular Biology* 23.10 (2016), pp. 884–890.

[373] K. Takeda et al. "Crystal structures of bacterial lipoprotein localization factors, LolA and LolB". In: *EMBO Journal* 22.13 (2003), pp. 3199–3209.

[374] Y. Wei et al. "Crystal structures of human lysosomal EPDR1 reveal homology with the superfamily of bacterial lipoprotein transporters". In: *Communications Biology* 2.1 (2019), pp. 1–13.

[375] K. S. Ryan et al. "The violacein biosynthetic enzyme VioE shares a fold with lipoprotein transporter proteins". In: *Journal of Biological Chemistry* 283.10 (2008), pp. 6467–6475.

[376] Y. K. Dong et al. "Crystal structure of RseB and a model of its binding mode to RseA". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.21 (2007), pp. 8779–8784.

[377] G. Sulzenbacher et al. "LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of Mycobacterium tuberculosis". In: *EMBO Journal* 25.7 (2006), pp. 1436–1444.

[378] F. Khan, D. Kurre, and K. Suguna. "Crystal structures of a β-trefoil lectin from Entamoeba histolytica in monomeric and a novel disulfide bond-mediated dimeric forms". In: *Glycobiology* 30.7 (2021), pp. 474–488.

[379] R. A. Romijn et al. "Mapping the collagen-binding site in the von Willebrand factor-A3 domain". In: *Journal of Biological Chemistry* 278.17 (2003), pp. 15035–15039.

[380] M. D. Shoulders and R. T. Raines. "Collagen structure and stability". In: *Annual Review of Biochemistry* 78 (2009), pp. 929–958.

[381] S. W. Hamaia et al. "Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin $\alpha1\beta1$ using collagen toolkits II and III". In: *Journal of Biological Chemistry* 287.31 (2012), pp. 26019–26028.

[382] V. Receveur-Brechot and D. Durand. "How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective". In: *Current Protein & Peptide Science* 13.1 (2012), pp. 55–75.

[383] J. J. Weinstein et al. "PROSS 2: a new server for the design of stable and highly expressed protein variants". In: *Bioinformatics* 37.1 (2021), pp. 123–125.

[384] S. Castellanos-Castro, J. Bolaños, and E. Orozco. "Lipids in Entamoeba histolytica: Host-Dependence and Virulence Factors". In: *Frontiers in Cellular and Infection Microbiology* 10 (2020), p. 75.

[385] J. Bolaños et al. "EhNPC1 and EhNPC2 Proteins Participate in Trafficking of Exogenous Cholesterol in Entamoeba histolytica Trophozoites: Relevance for Phagocytosis". In: *PLoS Pathogens* 12.12 (2016), e1006089.

[386] J. J. McCoy, B. J. Mann, and W. A. Petri. "Adherence and cytotoxicity of Entamoeba histolytica or how lectins let parasites stick around". In: *Infection and Immunity* 62.8 (1994), pp. 3045–3050.

[387] U. Katz et al. “Entamoeba histolytica expressing a dominant negative N-truncated light subunit of its Gal-lectin are less virulent”. In: *Molecular Biology of the Cell* 13.12 (2002), pp. 4256–4265.

General Conclusion

Main findings

Through this thesis, interactions at the host-pathogen interface were structurally characterised.

The first chapter shows that HIV-1 CA assembles into asymmetric and plastic hexamers, and that host factor CypA binds on the capsid through a novel non-canonical interface. Capsomer plasticity contributes to formation of different curvatures that are necessary to form a conical mature capsid and that is sensed by host factors, such as CypA, to modulate viral restriction.

The second chapter describes the structure of the inhibited Sec translocon. Helices of the lateral gate are prised open at the cytosolic side by mycolactone, while the plug helix remains in place. This suggests that the lateral gate of Sec is in a dynamic equilibrium, in which it opens and closes to recruit substrates for protein translocation. However, opening also provides space for allosteric inhibitors such as mycolactone, whose resistance mutations appear to affect Sec dynamics in disfavour of the inhibited conformation.

The last chapter presents the partial structure of the 260-kDa surface antigen of *Entamoeba histolytica* and shows that it contains several domains that may be important in carbohydrate, lipid and protein binding. The architecture of the protein is unique and consists of a globular core that includes LgL and the N-terminal half of HgL, while the C-terminal half of HgL appears to be flexibly linked and poorly resolved.

Progress in the fight against infectious diseases

The infectious diseases studied here affect millions of people worldwide, and their understanding at the molecular level is essential to design targeted therapies.

In the case of HIV-1, the capsid is an interesting target since it is essential for viral replication. Insights into the interactions between the capsid and host proteins that regulate its stability could therfore aid in the design of drugs that impair capsid fitness and/or compete with host factors that favour infectivity. In fact, multiple capsid-targeting antivirals are currently under development for treatment against HIV.

In the case of the Sec translocon, mutations of Sec61α and inhibitors of protein translocation, such as mycolactone, have been linked to disease phenotypes in humans. The structure of the inhibited state of Sec is important, not only because it provides a

deeper understanding of the dynamics of the essential process of protein translocation, but also because it suggests a mechanism of inhibition shared by different inhibitors. Therefore, small molecule modulators of Sec activity could be designed to compete with inhibitors responsible for defects in protein translocation.

As for the 260-kDa surface protein of *Entamoeba histolytica*, this is the most promising vaccine against human intestinal amoebiasis. However, its production is not scalable and the mechanisms at the molecular level by which it triggers an immune response are poorly understood. For unclear reasons, some antibodies against the same region of the protein have opposing effects, with some enhancing adhesion to host cells. Structures of protein fragments bound to such antibodies could help design smaller immunogens that elicit only neutralising antibodies and exceed the efficacy of those raised by the full antigen.

Cryo-EM at the host-pathogen interface

In this thesis, the host-pathogen interface was imaged by cryo-EM, which was powerful at different scales. Indeed, the method allowed imaging of a host cellular protein on large tubular assemblies of HIV-1 capsid protein, that of a small molecule inhibitor bound to a RTC, and that of asymmetric proteins of less than 200 kDa.

Cryo-EM of tubular assemblies allowed visualisation curved CA hexamers and novel interfaces between CypA and CA at near atomic resolution, which could not be achieved by other methods. Helical processing and localised reconstruction were especially powerful to deal with heterogenous tubes of different symmetries.

Cryo-EM was also ideal to study limited quantities of detergent-extracted protein from native membranes, such as RTCs from canine microsomes, and the 260-kDa surface protein from trophozoites of *Entamoeba histolytica*. In the first case, dynamic processes such as translocation could be deciphered from molecules trapped in different states thanks to advances in software development that allow classification of heterogenous particles. In the second case, though the structure is partial, it could still be visualised. Importantly, it highlighted elongated or flexible domains that may prevent crystal formation and study by X-ray crystallography.

Limitations and future advances

Although, in this work, cryo-EM provided new molecular insights into infectious disease processes, some limitations still need to be addressed.

First, as shown in the first two chapters, resolution at the regions of interest can be limited and often requires complementary methods such as molecular dynamics to validate models. As exemplified for the 260-kDa and 150-kDa antigens in the third

chapter, de novo structure building into a cryo-EM reconstruction can be challenging when structure homology is lacking. In this case, crystal or NMR structures of small domains can guide structure building into cryo-EM maps.

In addition, during cryo-EM imaging, samples are not in their native environment. For example, HIV-1 CA form tubular assemblies in high salt *in vitro*. These conditions are different from the native environment, in which HIV-1 capsid is conical. Similarly, although RTCs and the 260-kDa antigen were extracted from native sources, they were imaged in detergent rather than on their original membrane. Several important events can be missed during sample purification. For instance, though the 150-kDa antigen was co-purified with the 260-kDa antigen during extraction from *Entamoeba histolytica*, these were not observed during cryo-EM imaging. In the future, more structural insights may be obtained in native conditions thanks to advances in cryo-ET and image processing, such as subtomogram averaging. For instance, cryo-ET studies of RTCs in cell or bound to microsomal membranes could be performed. Similarly, cryo-ET is suitable to visualise HIV-1 mature capsid, because of its asymmetry. Last, the role of the 260-kDa antigen during host cell binding or during trogocytocis of a host cell could be studied locally using cryo-ET and correlative methods.

Appendices

A

Tables and Figures

Construct	Isoform	Boundaries	Vector	Tag	Forward primer	Reverse primer
				First round of cloning		
HgL_01	EHI_133900	16-190	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGACAAGCTGAACGAGTTCAGCGC	GATGGTGATGGTGGTGCTTGGTACCCAGCAGGCACTTGCTGGTGGTG
HgL_02	EHI_133900	16-210	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGACAAGCTGAACGAGTTCAGCGC	GATGGTGATGGTGGTGCTTGGTACCGTTGTCGTACACCTTCAGC
HgL_03	EHI_133900	16-388	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGACAAGCTGAACGAGTTCAGCGC	GATGGTGATGGTGGTGCTTGGTACCTCTGATGCCTTTGAACAGGGC
HgL_04	EHI_133900	211-388	pHLSec	Histidines	TAAGCAACCGGTAATGTGAACACCCATCTGGC	GATGGTGATGGTGGTGCTTGGTACCTCTGATGCCTTTGAACAGGGC
HgL_05	EHI_133900	16-1219	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGACAAGCTGAACGAGTTCAGCGC	GATGGTGATGGTGGTGCTTGGTACCGATGCACTCGGTGTGGTTG
HgL_06	EHI_133900	211-1219	pHLSec	Histidines	GTTGCGTAGCTGAAACCGGTAATGTGAACACCCATCTGGC	GATGGTGATGGTGGTGCTTGGTACCGATGCACTCGGTGTGGTTG
HgL_07	EHI_133900	395-1219	pHLSec	Histidines	GTTGCGTAGCTGAAACCGGTTGCGACCGGGATAAGTGCAG	GATGGTGATGGTGGTGCTTGGTACCGATGCACTCGGTGTGGTTG
HgL_08	EHI_133900	604-1015	pHLSec	Histidines	GTTGCGTAGCTGAAACCGGTGACAGCACCACCGGAAC	GATGGTGATGGTGGTGCTTGGTACCGCCCTCGATGAAGGCCTTT
HgL_09	EHI_133900	579-1154	pHLSec	Histidines	GTTGCGTAGCTGAAACCGGTATGTGCGAGAACCGGCTGAG	GATGGTGATGGTGGTGCTTGGTACCGGACACGGTGCAGGTTTTCT
HgL_10	EHI_133900	895-998	pet15b	TEV + Histidines	GAGAACCTGTATTTTCAGGGCGGATCCACCAGCTACAAGCTGCATCT	TAGCAGCCGGATCGCATATGTCAGCATGTGGTCTTGGTGGTAT
LgL1_01	EHI_035690	16-274	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTCAGGACGGCAAGGATCAGCTG	GATGGTGATGGTGGTGGTGCTTGGTACCCCGGCTGGCAATGATCTGG
LgL1_02	EHI_035690	16-288	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTCAGGACGGCAAGGATCAGCTG	GATGGTGATGGTGGTGCTTGGTACCAGCGAACACGGGGATGAAGATC
LgL1_03	EHI_035690	76-288	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAGTGATCTTCGCCAGCGAC	GATGGTGATGGTGGTGCTTGGTACCAGCGAACACGGGGATGAAGATC
LgL1_04	EHI_035690	76-288	pet15b	TEV + Histidines	GAGAACCTGTATTTTCAGGGCGGATCCAGAGTGATCTTCGCCAGCGAC	TAGCAGCCGGATCGCATATGTCAAGCGAACACGGGGATGAAGATC
LgL3_01	EHI_148790	17-256	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAAACCAGCTGAGCGAG	GATGGTGATGGTGGTGGTGCTTGGTACCGCTGGCGGCCTCAATCTG
LgL3_02	EHI_148790	17-270	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAAACCAGCTGAGCGAG	GATGGTGATGGTGGTGCTTGGTACCAGCGAACACGGGGATGAAGATC
LgL3_03	EHI_148790	76-270	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAGTGATCTTCGCCAGCGAC	GATGGTGATGGTGGTGCTTGGTACCAGCGAACACGGGGATGAAGATC
LgL3_04	EHI_148790	76-270	pet15b	TEV + Histidines	GAGAACCTGTATTTTCAGGGCGGATCCAGAGTGATCTTCGCCAGCGAC	TAGCAGCCGGATCGCATATGTCAAGCGAACACGGGGATGAAGATC
LgL6_01	EHI_183400	23-248	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAGATCAGCTGAGCCAG	GATGGTGATGGTGGTGGTGCTTGGTACCCACGCTGGCGGTGATCTGG
LgL6_02	EHI_183400	23-262	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAGATCAGCTGAGCCAG	GATGGTGATGGTGGTGCTTGGTACCAGCAAACACGGGGATGAAGATC
LgL6_03	EHI_183400	80-262	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAGAGTGATCTTCGCCAGCGAC	GATGGTGATGGTGGTGCTTGGTACCAGCAAACACGGGGATGAAGATC
LgL6_04	EHI_183400	80-262	pet15b	TEV + Histidines	GAGAACCTGTATTTTCAGGGCGGATCCAGAGTGATCTTCGCCAGCGAC	TAGCAGCCGGATCGCATATGTCAAGCAAACACGGGGATGAAGATC
IgL_01	EHI_006980	14-1082	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGATAAGCTGATCGGCGGC	GATGGTGATGGTGGTGCTTGGTACCGCACTTTCCGTCCTTCAGC
IgL_02	EHI_006980	14-382	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGATAAGCTGATCGGCGGC	GATGGTGATGGTGGTGCTTGGTACCCAGTTTACAAGGTCCGTCTGTGG
IgL_03	EHI_006980	294-753	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTACCGAAGAGAACAAATGCACC	GATGGTGATGGTGGTGCTTGGTACCCAGCCGCTGCTCGGTACAG
IgL_04	EHI_006980	729-967	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGCCAAGTGCAAGACATGCAAG	GATGGTGATGGTGGTGCTTGGTACCGCACTTGGTACACTCTTCCAC
IgL_05	EHI_006980	603-1082	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGAGGGACCTAACGCCGAG	GATGGTGATGGTGGTGCTTGGTACCGCACTTTCCGTCCTTCAGC
IgL_06	EHI_006980	804-1082	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGCCACAGAGTGCTCTGAC	GATGGTGATGGTGGTGCTTGGTACCGCACTTTCCGTCCTTCAGC
				New constructs after limited proteolysis coupled with MS		
HgL_11	EHI_133900	458-1166	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGATCTGGAAGGCTGCGTGG	GATGGTGATGGTGGTGCTTGGTACCCCGGCCCTGATTGGGGGT
HgL_12	EHI_133900	536-1166	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTACCAACAACTGCCAGGAATAC	GATGGTGATGGTGGTGCTTGGTACCCCGGCCCTGATTGGGGGT
HgL_13	EHI_133900	626-973	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTAACCACGAGATCTGCGACTACAC	GATGGTGATGGTGGTGCTTGGTACCTTTGCACTCGATGGGCTGG
HgL_14	EHI_133900	641-946	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGTGTCCAACTGCACCGAG	GATGGTGATGGTGGTGCTTGGTACCCTTTGTTTCCTCGGCGCAC
HgL_15	EHI_133900	664-931	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTACCACCTACTGGGAGAACG	GATGGTGATGGTGGTGCTTGGTACCTGTCTCGATCAGGCTGCAG
HgL_16	EHI_133900	707-1007	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGTGGGCGACGTGTCCAATG	GATGGTGATGGTGGTGCTTGGTACCGTCCTCGATCCGCTTGTTG
HgL_17	EHI_133900	802-946	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGTGCAGCCTACACCAGCC	GATGGTGATGGTGGTGCTTGGTACCCTTTGTTTCCTCGGCGCAC
HgL_18	EHI_133900	63-465	pHLSec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTCACTTTGTGTGGACCACCTG	GATGGTGATGGTGGTGCTTGGTACCGGTGACCACGCAGCCTTC
				Fragments of antibody		
CP33-VH			pOPIN	Cγ1	TGGGTTGCGTAGCTGGTACCGAAGTGAAGCTGATGGAATCTGG	TGGGGGTGTCGTTTTGGCAGAAGAGACTGTCACGGCG
CP33-VL			pOPIN	Cκ + Histidines	ATGGGTTGCGTAGCTGGTACCGACATCGAGATGACACAGAGC	TGCAGCATCAGCCCGAGCTCCCGCTTGATTTCCACCTTGG
XEhI-20-VH			pOPIN	Cγ1	TGGGTTGCGTAGCTGGTACCCAGGTGCAGCTGGTTGAATC	TGGGGGTGTCGTTTTGGCAGAAGAGACTGTCACTGTGGTG
XEhI-20-VL			pOPIN	Cκ + Histidines	ATGGGTTGCGTAGCTGGTACCGAGATCGTGATGACACAGAGC	TGCAGCATCAGCCCGAGCTCCCGCTTGATTTCCACCTTGG
XEhI-H2-VH			pOPIN	Cγ1	TGGGTTGCGTAGCTGGTACCCAGGTTCAGCTGCAGCAG	TGGGGGTGTCGTTTTGGCAGAGCTAACGGTGACCAGTG
XEhI-H2-VL			pOPIN	Cκ + Histidines	ATGGGTTGCGTAGCTGGTACCGACATCGTGATGACACAGAGC	TGCAGCATCAGCCCGAGCTCCCGCTTGAGGGCCACCTT
XEhI-28-VH			pOPIN	Cγ1	TGGGTTGCGTAGCTGGTACCGAGGTGCAGCTGGTTGAATC	TGGGGGTGTCGTTTTGGCAGAAGAAACGGTGACCAGTGTG
XEhI-28-VL			pOPIN	Cκ + Histidines	ATGGGTTGCGTAGCTGGTACCGAGATCGTGCTGACACAGTC	TGCAGCATCAGCCCGAGCTCCCGCTTGATTTCCACCTTGG
EH3015-VH			pOPIN	Cγ1	TGGGTTGCGTAGCTGGTACCGAGATCGTGCTGTCTCAGAGC	TGGGGGTGTCGTTTTGGCCTTGATTTCCAGCTTGGTGCC
EH3015-VL			pOPIN	Cκ + Histidines	ATGGGTTGCGTAGCTGGTACCGAAGTGAAGCTGCTGCAGTC	TGCAGCATCAGCCCGAGCTCAGCAGACACTGTGACCAG
CP33-ScFv_HL			pOPIN	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGAAGTGAAGCTGATGGAATCTGGCGG GTGGTTCCTCTAGATCTTCCTCCTCTGGTGGCGGTGGCTCGGGCGGTGGTG GGGACATCGAGATGACACAGAGC	GGAAGATCTAGAGGAACCACCAGAAGAGACTGTCACGGCGG GTGATGGTGGTGCTTGGTACCCCGCTTGATTTCCACCTTGGTGCCC

(continued on next page)

				Coiled coil constructs		
HgL_19	EHI_133900	16-1228	pHLSec	TEV + ACID	GATGGGTTGCGTAGCTGAAACCGGTGACAAGCTGAACGAGTTCAGCGC	TTGAGCCAACTCTTTCTCCAGTGCCTGCAGCTCCCACTCAAGTTGGGCGTTCTC TTTTTCCAATGCCTGCAGCTCCTTCTCACATTGGGCACTACCCCCACTGCCACC GCTACCACCACTTCCCCCCTGGAAGTAAAGGTTTTCCAAGCCTCCCCCAACGTT GTTGATCTCGCCGGCC
LgL1_05	EHI_035690	16-274	pHLSec	TEV + linker + BASE + Histidines	GATGGGTTGCGTAGCTGAAACCGGTCAGGACGGCAAGGATCAGCTG	TTGAGCGAGTTTCTTCTTAAGTGCTTGGAGCTTCCACTTAAGCTGTGCGTTTTT CTTCTTAAGGGCCTGCAATTTTTTCTTACACTGTGCTGATCCGCCGGACCCCCC ACTTCCACCACTACCTCCCTGAAAATACAAGTTTTCGAGACCGCCCCGGCTGGC AATGATCTGGTAGG
LgL3_05	EHI_148790	17-256	pHLSec	TEV + linker + BASE + Histidines	GATGGGTTGCGTAGCTGAAACCGGTCAGGACGGCAAGGATCAGCTG	TTGAGCGAGTTTCTTCTTAAGTGCTTGGAGCTTCCACTTAAGCTGTGCGTTTTT CTTCTTAAGGGCCTGCAATTTTTTCTTACACTGTGCTGATCCGCCGGACCCCCC ACTTCCACCACTACCTCCCTGAAAATACAAGTTTTCGAGACCGCCGCTGGCGGC CTCAATCTGG
LgL6_05	EHI_183400	23-248	pHLSec	TEV + linker + BASE + Histidines	GATGGGTTGCGTAGCTGAAACCGGTCAGGACGGCAAGGATCAGCTG	TTGAGCGAGTTTCTTCTTAAGTGCTTGGAGCTTCCACTTAAGCTGTGCGTTTTT CTTCTTAAGGGCCTGCAATTTTTTCTTACACTGTGCTGATCCGCCGGACCCCCC ACTTCCACCACTACCTCCCTGAAAATACAAGTTTTCGAGACCGCCCACGCTGGC GGTGATCTG
				Cryo-EM of HgL C-terminus		
HgL_20	EHI_133900	641-1220	pHLsec	Histidines	GATGGGTTGCGTAGCTGAAACCGGTGTGTCCAACTGCACCGAG	GATGGTGATGGTGGTGCTTGGTACCCAGAATACACTCGGTGTGGTTG
				Collagen binding assay		
HgL_21	EHI_133900	211-388	pGEX-2T	Thrombin + GST	CCAAAATCGGATCTGGTTCCGCGTGGATCCAATGTGAACACCCATCTGGCC	CAGATCGTCAGTCAGTCACGATGAATTCTCATCTGATGCCTTTGAACAGGGCTG

Table A.1: List of primers.

Construct	Boundaries	Conditions	Concentration (mg/mL)	Commercial screens
HgL_01	16-190	D, M	D: 10, 40 M: 10	D, M: Morpheus I, MIDAS, JCSG+, Memgold I, Memgold II
HgL_02	16-210	D, M	D: 10, 35; M: 10	D, M: Morpheus I, JCSG+, Memgold I, Memgold II
HgL_03	16-388	D, M	D: 10, 15; M: 15	D,M: Morpheus I, MIDAS, JCSG+, PACT, Memgold I, Memgold II
HgL_04	211-388	G, D	G: 10; D: 15	G,D: Morpheus I, MIDAS, JCSG+, PACT, Memgold I, Memgold II
HgL_05	16-1219	G, D, M, D+F	G: 10; D: 10, 20; M: 20; D+F: 10	G,D,M: Morpheus I, MIDAS, JCSG+, PACT, Memgold I, Memgold II D+F: Morpheus I, MIDAS, JCSG+
HgL_06	211-1219	D	D: 10	D: Morpheus I, MIDAS, JCSG+
HgL_07	395-1219	None	None	None
HgL_08	604-1015	G, D, M, D+F, D+S, M+S, M+F	G: 10; D: 10, 25; M: 15; M+S 9; M+F 11	G, D, M: Morpheus I, MIDAS, JCSG+, PACT, Memgold I, Memgold II M+S, M+F: Morpheus I, JCSG+, Memgold I
HgL_09	579-1154	G, D, M, M+S, M+F	G: 10; D: 10, 30; M: 15	G, D, M: Morpheus I, MIDAS, JCSG+, PACT, Memgold I, Memgold II M+S, M+F: Morpheus I, JCSG+, Memgold I, Memgold II
HgL_10	895-998	D (*E. Coli*), D+GalNAc	D: 5, 10; D+GalNAc: 9	D, D+GalNAc: Morpheus I, MIDAS, JCSGI, JCSGII, JCSGIII, JCSG IV, PACT, Memgold I, Memgold II
HgL_11	458-1166	D, D+F	D: 10; D+F: 15	D: Morpheus I, MIDAS, JCSG+, Memgold I, Memgold II D+F: Morpheus I, JCSG+, Memgold I
HgL_12	536-1166	D, D+F	D: 10; D+F: 10	D: Morpheus I, MIDAS, JCSG+, Memgold I, Memgold II D+F: Morpheus I, JCSG+
HgL_13	626-963	D, D+F, D+S, M+F, M+S, F+LP, S+LP	D: 10, 20; D+F: 15, M+F: 8, M+S: 12, F+LP: 10, S+LP 15	D: Morpheus I, MIDAS, JCSG+, Memgold I, Memgold II D+F: Morpheus I, JCSG+, Memgold I, Memgold II M+F, M+S: Morpheus I, JCSG+, Memgold I F+LP, S+LP: Morpheus I, JCSG+
HgL_14	641-946	D	D: 15	D: Morpheus I, MIDAS, JCSG+, Memgold I
HgL_15	664-931	D	D: 17	D: Morpheus I, MIDAS, JCSG+, Memgold I
HgL_16	707-1007	D, D+F, D+S, M+F, M+S, F+LP, S+LP	D: 15; D+F: 10, M+F: 10, M+S: 15, F+LP: 7, S+LP: 5	D: Morpheus I, MIDAS, JCSG+, Memgold I, Memgold II D+F: Morpheus I, JCSG+, Memgold I, Memgold II M+F, M+S: Morpheus I, JCSG+, Memgold I, Memgold II F+LP, S+LP: Morpheus I, JCSG+
HgL_17	802-946	D	D: 10; 20	D: Morpheus I, MIDAS, JCSG+, Memgold I
HgL_18	63-465	D	D: 15	D: Morpheus I, MIDAS, JCSG+, Memgold I
IgL_01	14-1082	D, M, D+F, M+F	D: 10, 30; M: 50, D+F:10, M+F: 15	D: Morpheus I, MIDAS, JCSG+, Memgold I, Memgold II M: JCSG+, Memgold I D+F, M+F: Morpheus I, JCSG+, Memgold I
IgL_02	14-382	D	D: 10	D: Morpheus I, MIDAS, JCSG+
IgL_03	294-753	None	None	None
IgL_04	729-967	D	D: 10	D: Morpheus I, MIDAS, JCSG+
IgL_05	603-1082	D	D: 20	D: Morpheus I, MIDAS, JCSG+
IgL_06	804-1082	D	D: 10	D: Morpheus I, MIDAS, JCSG+
CP33 Fab	NA	D	D: 5, 9, 18	D: Morpheus I, MIDAS, JCSG+, PACT, PEGions
CP33 scFv	NA	D	D: 5, 10	D: Morpheus I, MIDAS, JCSG+, PACT, PEGions
CP33 scFv	NA	D	D: 5, 10	D: Morpheus I, MIDAS, JCSG+, PACT, PEGions
HgL_19:LgL1_05	16-1228:16-274	D	D: 9	D: Morpheus I, JCSG+, Memgold I

Table A.2: Crystallisation trials. G, D, M, F, S and LP stand for glycosylated, deglycosylated, methylated, bound to Fab, bound to scFv and obtained after limited proteolysis, respectively.

IDs	MW (kDa)	Sequence length	Peptides	Unique peptides	Sequence coverage (%)	Unique sequence coverage (%)	Intensity
P32022, EHI_133900, HgL3	143.58	1278	78	18	58.3	14.6	6.68×10^{11}
C4LTM0, EHI_012270, HgL2	144.33	1286	72	35	55.6	29.9	2.24×10^{11}
C4M8N3, EHI_077500, HgL1	145.09	1291	72	4	52	2.2	4.61×10^{10}
B1N559, EHI_042370, HgL5	142.58	1270	68	5	51	5.7	1.38×10^{10}
C4M6L0, EHI_065330, IgL2	120.39	1105	26	19	33.8	25.9	5.61×10^{9}
C4M9B6, EHI_003930	114.98	995	1	1	0.7	0.7	2.93×10^{8}
Q964D2, EHI_006980, IgL1	119.51	1101	12	5	12.9	5	2.75×10^{8}
C4M7A1, EHI_046650, HgL4	134.78	1207	10	10	9.4	9.4	2.26×10^{8}
C4M0Z5, EHI_135090	41.97	367	3	3	9.5	9.5	1.65×10^{8}
C4M1M6, EHI_133970	40.112	339	1	1	2.9	2.9	1.62×10^{7}

Table A.3: Mass spectrometry of in-gel tryptic digest of native HgL.

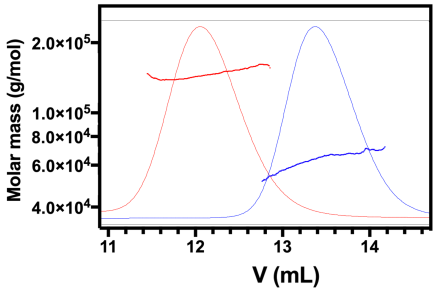


Figure A.1: SEC-MALS of HgL_05 and HgL_09. Molecular weights were estimated at 145.9 kDa and 63.4 kDa for HgL_05 and HgL_09, respectively.

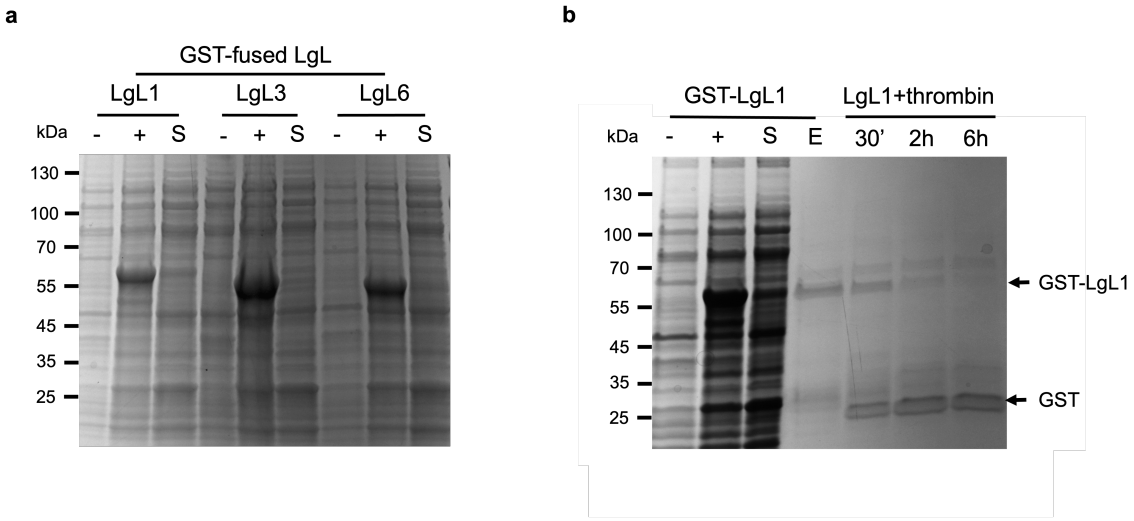


Figure A.2: Purification trials of GST-fused LgL. Three isoforms of full LgL were expressed in *E. Coli*: LgL1 (EHI_035690), LgL3 (EHI_147980) and LgL6 (EHI_183400). a) SDS-PAGE showing protein content of bacterial cultures before and after induction with IPTG (- and +, respectively), and of the soluble fraction purified after sonication (S). b) SDS-PAGE of the same steps and elution after IMAC (E) followed by cleavage with thrombin over time. Only GST is retrieved, while LgL is unstable without the soluble tag.

B
Publications

nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-020-0467-8

Check for updates

Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A

Tao Ni[1], Samuel Gerard[1], Gongpu Zhao[2], Kyle Dent[3], Jiying Ning[2], Jing Zhou[4], Jiong Shi[4], Jordan Anderson-Daniels[4], Wen Li[5,6], Sooin Jang[5,6], Alan N. Engelman[5,6], Christopher Aiken[4] and Peijun Zhang[1,2,3] ✉

The mature retrovirus capsid consists of a variably curved lattice of capsid protein (CA) hexamers and pentamers. High-resolution structures of the curved assembly, or in complex with host factors, have not been available. By devising cryo-EM methodologies for exceedingly flexible and pleomorphic assemblies, we have determined cryo-EM structures of apo-CA hexamers and in complex with cyclophilin A (CypA) at near-atomic resolutions. The CA hexamers are intrinsically curved, flexible and asymmetric, revealing the capsomere and not the previously touted dimer or trimer interfaces as the key contributor to capsid curvature. CypA recognizes specific geometries of the curved lattice, simultaneously interacting with three CA protomers from adjacent hexamers via two noncanonical interfaces, thus stabilizing the capsid. By determining multiple structures from various helical symmetries, we further revealed the essential plasticity of the CA molecule, which allows formation of continuously curved conical capsids and the mechanism of capsid pattern sensing by CypA.

The HIV-1 capsid is a conical assembly of approximately 200 hexamer and 12 pentamer capsomeres of the CA. CA comprises two α-helical domains, an N-terminal domain (CA_{NTD}) and a C-terminal domain (CA_{CTD}), connected by a flexible linker. CA is the most conserved among HIV-1 proteins and is essential for key functions, including virus assembly, reverse transcription, nuclear entry and integration[1]. In fact, it is the curved capsid lattice, rather than individual CA protomers, that serves as a primary binding platform for multiple host proteins that promote or inhibit HIV-1 infection[1–3].

The HIV-1 capsid has been the target of intensive study owing to its multiple roles in HIV-1 replication and its potential as a therapeutic target. Numerous efforts have focused on understanding how CA assembles into a mature fullerene cone. High-resolution crystal structures of hexameric and pentameric capsomeres have been determined alone[4–6] and in complex with various ligands (PF74 (refs. [4,7]), GS-CA1 (ref. [8]), IP6 (refs. [9,10])), but all in planar form. In addition, intermediate-resolution structures of curved assemblies have been resolved by cryo-EM[11] or cryo-electron tomography (cryo-ET) and subtomogram averaging[12], revealing a new trimer interface critical to capsid assembly and rigid-body tilting and twisting between capsomeres to produce the curved lattice. However, a precise and reliable atomic model of the curved capsid assembly has not been tractable, which has impeded our understanding of virus–host interactions that exceed individual capsomeres. Many host cell defense proteins, including the restriction factors Trim5α[13], TrimCyp[2,14] and MxB[3], target the HIV-1 capsid and inhibit virus replication through pattern recognition of the curved assembly, but not the individual CA protomers or capsomeres.

Mature capsid assemblies are pleomorphic and highly flexible and thus represent a challenge for high-resolution structure determination. We developed a data collection method, ArbitrEM, to automatically collect, from selected capsid assemblies, the large number of cryo-EM images that are required to achieve high resolution, thus vastly enhancing the effectiveness and throughput of data acquisition. We further devised an image processing and reconstruction workflow, including symmetry expansion, localized three-dimensional (3D) classification and reconstruction, for exceedingly flexible and pleomorphic assemblies. Here, we present cryo-EM structures of a mature capsid assembly at 3.6-Å resolution and in complex with the host protein CypA at 4.0-Å resolution, from which we built atomic models of the hexameric capsomere and the CypA–CA lattice complex. While previous published structures of the wild-type (WT) CA hexamer[4] or cross-linked CA hexamer and pentamer[5,6] are six (or five)-fold symmetric and in planar form due to crystalline lattice constrains, the cryo-EM structure of the CA hexamer from the curved assembly is, by contrast, intrinsically curved and asymmetric, distinguishing it from all previous atomic structures. The curvature of CA hexamers varies depending on the helical symmetry, whereas the interhexamer interactions remain constant, resulting in variably curved CA surface lattices in assemblies. CypA senses the curved CA lattice, simultaneously interacting with three CA protomers from adjacent hexamers via the canonical CypA loop and additional noncanonical new interfaces, thus stabilizing the capsid. We further determined multiple capsomere and CypA–CA structures from several helical symmetries, which provided a basis for dissecting the mechanisms of continuous curvature formation and how CypA embraces the curved CA surface lattice.

[1]Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [2]Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [3]Electron Bio-Imaging Centre, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, UK. [4]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA. [5]Department of Medicine, Harvard Medical School, Boston, MA, USA. [6]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA. ✉e-mail: peijun@strubi.ox.ac.uk

Results

Cryo-EM structure of HIV-1 CA tubular assemblies. HIV-1 CA forms tubular assemblies of diverse diameters and helical symmetries[11,15]. CA tubes are flexible and fragile, reflecting the metastable nature of the capsid that protects the genome, while at the same time permitting controlled disassembly (uncoating) during infection[16]. These features of the capsid make it a very challenging target for atomic-resolution structure determination by cryo-EM, which would typically require manually recording thousands of images. To facilitate this task, we developed ArbitrEM (Extended Data Fig. 1), a data collection script and workflow built on the SerialEM platform[17], to automatically record micrographs of selected targets (in this case, tubes that are suitable for high-resolution analysis) (Extended Data Fig. 1 and Methods). The major advantages of ArbitrEM include: (1) targeting is carried out on-the-fly automatically; (2) multiple targets can be positioned with minimal stage movement, therefore improving throughput; and (3) additional new targets can be added offline during data collection. Once tubes of interest were identified, ArbitrEM automatically collected ~100 micrographs per hour with 30-nm positioning accuracy, a great improvement over manual data collection. Subsequently, the tube images were cropped (Extended Data Fig. 2a) and sorted on the basis of their helical symmetries: initially by their diameters (Extended Data Fig. 2b) and then by manual inspection of their fast Fourier transform (FFT) patterns (Extended Data Fig. 2c). More than 12 helical symmetries were identified within a 3-nm narrow range of tube diameters (Extended Data Fig. 2b, red box), illustrating the challenge in image processing.

Assemblies carrying the A92E substitution in CA, which prevents tube aggregation[15], were analyzed, initially by real-space helical reconstruction[18] (Extended Data Fig. 2d,e and Supplementary Table 1) and then, because CA assemblies are flexible, using a localized reconstruction method[19] (Extended Data Fig. 2f; see Methods). With this approach, structures of A92E assemblies from three abundant helical symmetries, (−12,11), (−13,11) and (−13,12), were determined at resolutions of 4.4, 4.7 and 4.9 Å, respectively (Extended Data Fig. 3a,b and Supplementary Table 1). Combining data from the three helical symmetries only marginally improved the map (4.2-Å resolution), despite nearly doubling the amount of data. For this reason, we altered our approach and used CypA to stabilize WT CA assembly[20–22], and determined the structure of the CypA-stabilized assembly at 3.6-Å resolution (Extended Data Fig. 3 and Table 1). From this structure, we built an atomic model of the hexameric capsomere in a curved lattice (Fig. 1a and Table 1). Comparison of capsomere structures with the same helical symmetry (−13, 12) in the presence or absence of CypA revealed that the CypA-stabilized capsomere is essentially the same as the A92E capsomere (Extended Data Figs. 3a and 4).

The best-resolved region in the 3.6-Å resolution cryo-EM map is at the inner core of the hexamer, comprising CA helices 1–3 (Fig. 1a, arrow), whereas the CA_{CTD} and CypA-binding loop are less well resolved (Fig. 1a). The interdomain linker is also clearly resolved (Fig. 1a, arrow), while the N-terminal β-hairpin density is relatively weak compared to the inner core. However, at low iso-surface levels, the main chain of the β-hairpin can be unambiguously placed. The overall quality of the atomic model is indicated by the Fourier shell correlation (FSC) between the model and the density map (Extended Data Fig. 3c) and exemplified in the overlay of the model and the density mesh, where side chain densities are clear and can be modeled unambiguously (Extended Data Fig. 3d).

The resulting atomic model revealed a highly dynamic outer surface, including the CypA-binding loop, followed by the β-hairpin and loop between H4 and H5 (Extended Data Fig. 5a,b), in addition to a dynamic region at the inner surface of the assembly, in particular the trimer interface comprising CA_{CTD} helix H10 of three adjacent hexamers (Extended Data Fig. 5c–e). These regions are targets for host cell protein binding, such as CypA[20,23], Trim5α[13,24], TrimCyp[14] and MxB[3,25]; therefore, structural dynamics could play important roles in such interactions.

Table 1 | Cryo-EM data collection, refinement and validation statistics

	Wild-type CA hexamer (EMD-10226, PDB 6SKK)	CA–CypA complex (EMD-10739, PDB 6Y9W)
Data collection and processing		
Magnification	47,196	47,196
Voltage (kV)	300	300
Electron exposure (e^-/Å^2)	40	40
Defocus range (μm)	0.8–2.0	0.8–2.0
Pixel size (Å)	1.06	1.06
Symmetry imposed	*C2*	*C1*
Initial particle images (no.)	207,264	207,264
Final particle images (no.)	207,264	161,690
Map resolution (Å)	3.6	4.1
FSC threshold	0.143	0.143
Map resolution range (Å)	3.5–4.3	3.8–5.4
Refinement		
Initial model used	PDB 4XFX	PDB 4XFX PDB 1AK4
Model resolution (Å)		
FSC threshold	0.5	0.5
Model resolution range (Å)	3.8	4.7
Map sharpening *B* factor (Å^2)	−217	−50
Model composition		
Non-hydrogen atoms	0	0
Protein residues	1,320	2,364
Ligands	0	0
B factors (Å^2)		
Protein	74.12	157.9
Ligand	0	0
R.m.s. deviations		
Bond lengths (Å)	0.003	0.012
Bond angles (°)	0.596	1.065
Validation		
MolProbity score	1.41	1.52
Clashscore	7.50	9.87
Poor rotamers (%)	0.18	0.45
Ramachandran plot		
Favored (%)	98.32	97.99
Allowed (%)	1.68	1.97
Disallowed (%)	0	0.04

The CA hexamer is intrinsically curved and asymmetric. A long-standing question regarding HIV-1 capsid assembly is focused on the physical properties that allow a continuously curved non-symmetrical shell to be produced from ~1,200 copies of a single protomer. Our data reveal that the CA hexamer is not symmetrical and is curved by approximately 20° (Fig. 1b), which is distinct from all existing high-resolution structures[4,5]. This asymmetry is evident in all structures from five helical symmetries. Superposition of the

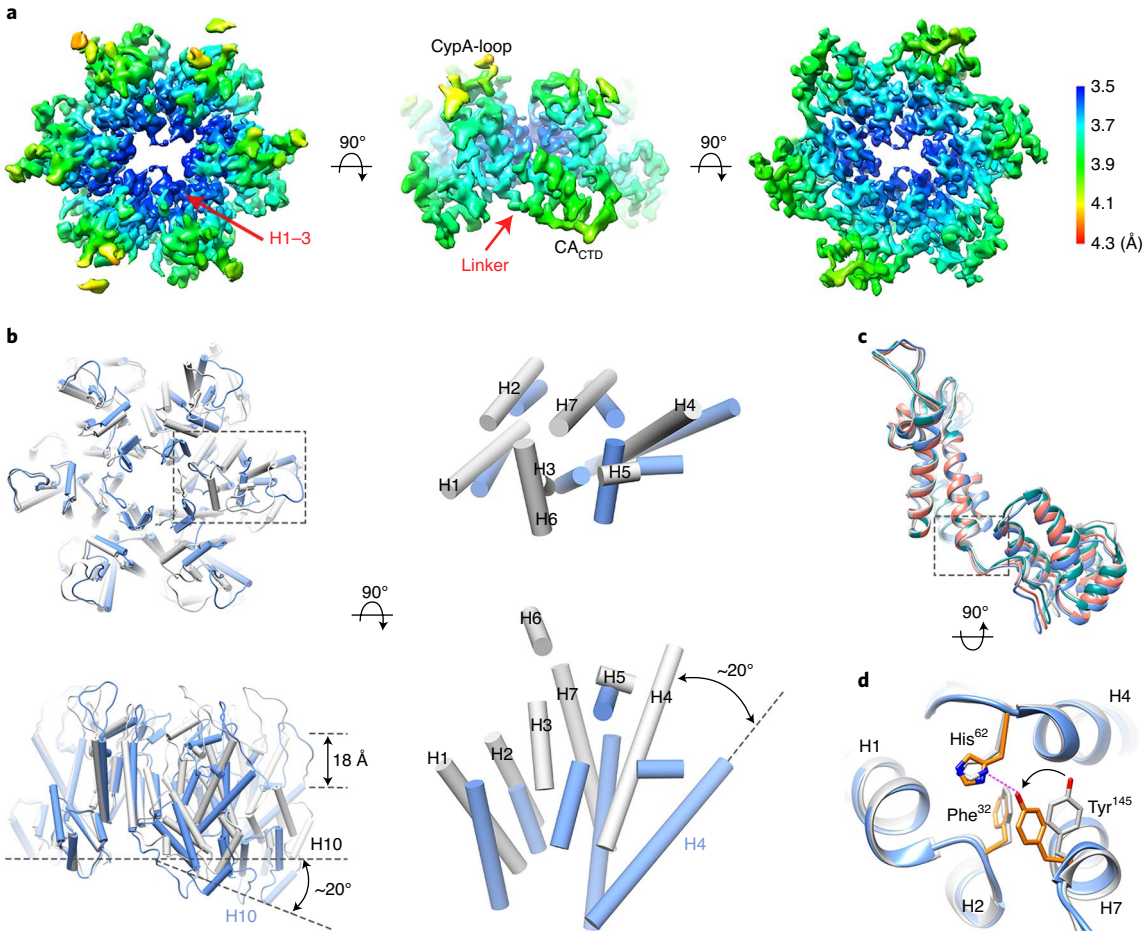


Fig. 1 | The mature CA hexamer is intrinsically curved and asymmetric. a, Localized real-space helical reconstruction of in vitro-assembled WT CA tubes with (−13, 8) helical symmetry at 3.6-Å resolution. Three orthogonal views of the CA hexamer density, from top, side and bottom, are shown. The density map is colored according to the local resolution. CA helices H1–3, CypA loop, CA_{NTD}–CA_{CTD} linker and CA_{CTD} are labeled. **b**, Superposition of WT CA hexamers from a planar crystal (gray, PDB 4XFX) and tubular assembly (blue), which displays ~20° curvature. Close-up views of CA_{NTD} helices are shown on the right. **c**, Flexible domain connection among three asymmetric CA monomers, colored in blue, orange and cyan, overlaid with PDB 4XFX monomer (gray). **d**, Close-up view of the boxed region in **c**, displaying a new hydrogen bond between His 62 and Tyr 145 in a CA protomer along the most curved direction.

WT CA hexamer from a planar crystal (PDB 4XFX) and the tubular assembly reveals a large discordance (r.m.s. deviation (r.m.s.d.) of 3.21 Å of all Cα in the hexamer) (Fig. 1b). The distances between diametrically opposed CA_{CTDs} within the hexamer (between N195 Cα) vary from 77.2 Å to 83.6 Å (Extended Data Fig. 6a), compared to 85.3 Å in planar crystals. In fact, CA_{CTD} distances correlate well with the local curvature of hexamer arrays, measured from the cryo-EM structures of the five helical assemblies determined in this study (Extended Data Fig. 6b), suggesting that the CA hexamer itself is naturally flexible to meet the needs of the local environment. The flexibility of the CA hexamer does not reside in the individual CA_{NTD} and CA_{CTD} domains (r.m.s.d. values of 0.55 Å and 0.84 Å, respectively) but originates from variable orientations between the CA_{NTD} and CA_{CTD} through the flexible, yet well-ordered, linker (Fig. 1c). The highly conserved linker residues are critical for HIV-1 infection[26]. Intriguingly, Tyr 145 forms a new intramolecular hydrogen bond with His 62 within the most curved CA subunit (Fig. 1d), in contrast to its intermolecular contact with Arg 162 at the CA_{NTD} and CA_{CTD} interface uniformly observed in planar crystal hexamers (PDB 3H4E and PDB 4XFX, for example). Breaking this intramolecular hydrogen bond by substitution of either His 62 (H62F, H62W)[27] or Tyr 145 (Y145A and Y145F, Y145A)[26,28] impaired assembly in vitro and virus infection in tissue culture.

Asymmetry is also present at the hexamer center, in particular at the β-hairpins (Extended Data Fig. 6c) and residue Arg 18 (Extended Data Fig. 6d). Intriguingly, clearly resolved Arg 18 side chains adopt two alternative conformations along the curved directions (Extended Data Fig. 6d, dashed box, and Extended Data Fig. 6e) that differ from those observed in crystal structures with and without the pocket cofactor IP6 (ref. [9]) (Extended Data Fig. 6f,g). It is tempting to speculate that Arg 18 could bind and release IP6 in exchange for cytosolic nucleotides through such dual configurations. Further structure-guided molecular dynamics simulations may shed light on the complex interplay between Arg 18, IP6 and nucleotides.

The flexible and asymmetric nature of the hexameric capsomere provides insight into a long-puzzled observation that we and others had regarding the small-molecule capsid inhibitor, PF74. It had been difficult to reconcile concentration-dependent effects of PF74 on CA tube assembly: while low concentrations of PF74 elicited the formation of long CA tubes, higher PF74 concentrations templated the formation of very short and rigid tubes (Extended Data Fig. 7a–f). These differing effects on CA assembly are also consistent with the two distinct antiviral mechanisms exerted by the compound at low and high concentrations[29]. These opposing effects can now be explained in light of the curved hexamer structure (Extended Data Fig. 7g–i).

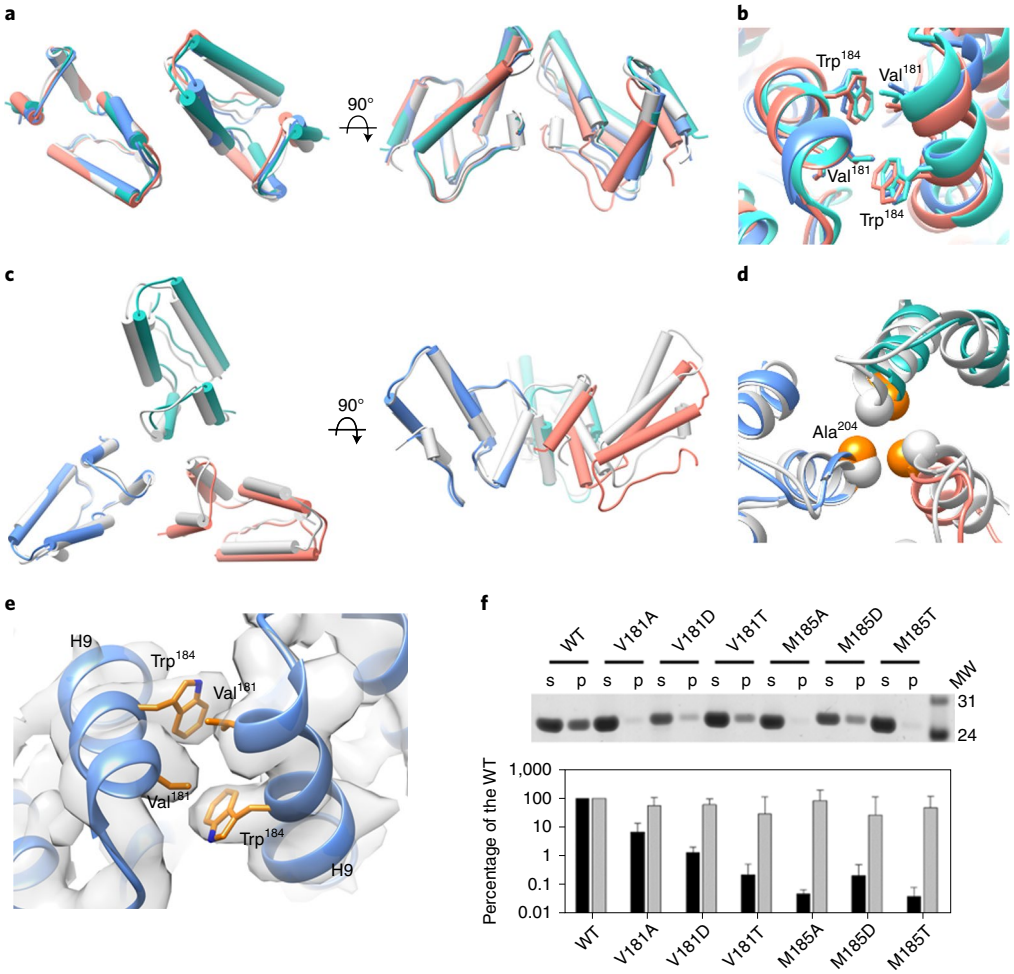


Fig. 2 | Plasticity of interhexamer interfaces in mature capsid. a, Overlay of three dimer interfaces from the (−13, 8) helical assembly (color) superimposed with the same interface from a planar crystal (gray, PDB 4XFX). **b**, Enlarged view of the superimposed dimer interfaces in **a** in ribbon and stick representation. Val 181 and Trp 184 side chains are shown. **c**, Superposition of the trimer interfaces from the (−13, 8) helical assembly (color) and PDB 4XFX (gray). **d**, Enlarged view of the trimer interface in ribbon representation. Ala 204 side chains are shown as spheres in the cryo-EM model (orange) and planar crystal structure (gray). **e**, Close-up view of the dimer interface, displaying the interaction between Val 181 and Trp 184. **f**, Effects of Val 181 and Met 185 mutations on in vitro capsid assembly (top; s, supernatant; p, pellet), viral infectivity (bottom; black) and p24 in culture supernatant (bottom; gray); data are mean and s.d. of $n = 3$, measured from independent experiments. The uncropped gel image is shown in the Source data.

As PF74 interacts with two adjoining CA molecules at the CA_{NTD} and CA_{CTD} interface in the crystal structure, it stabilizes the hexamer and promotes capsid assembly. However, the six CA_{NTD} and CA_{CTD} interfaces in the curved hexamer are not equal. At low PF74 occupancy, PF74 promotes long tube formation by retaining the asymmetry and flexibility of the CA hexamer; at full PF74 occupancy, binding of PF74 causes induced fit, resulting in homogenization of six CA_{NTD} and CA_{CTD} interfaces, and therefore losing flexibility and persistence and generating shorter, more rigid tubes. This is supported by observations that PF74 increases the stiffness of isolated cores[30] and accelerates capsid opening but stabilizes the remaining lattice in vitro[31].

Plasticity of interhexamer interfaces. The CA dimer interface has been extensively studied by X-ray crystallography[4,32] and NMR[28,33], but how the interaction contributes to curvature formation is unknown. A recent cryo-ET study found that tilts (−1° to 29°) and twists (−12° to 12°) between neighboring hexamers could account for lattice curvature[12]. We observed a similar range of tilts (0° to 31°) and twists (−8° to 8°) among five different symmetries, but only if we artificially treated the cryo-EM hexamers as rigid and flat. Yet, the three dimer interfaces in the curved assembly appear to be similar to the planar crystal dimer (Fig. 2a,b). The deviations among three dimers and among five helical assemblies were small, with r.m.s.d. values of 0.67–1.18 Å and 0.19–1.39 Å, respectively. The trimer interface was also preserved (Fig. 2c,d). Therefore, the variable curvature of the capsid (0° to 31°) is contributed largely by the asymmetrically curved hexamer itself without substantially altering dimer and trimer interactions, whereas minor adjustments of hydrophobic interactions at the dimer and trimer interfaces provide some plasticity.

Our cryo-EM structure delineates the interaction between Val 181 and Trp 184 at the dimer interface (Fig. 2e), in addition to the well-characterized Trp 184 and Met 185 interaction[32]. Although the Val 181–Trp 184 interaction was previously suggested by NMR experiments[28], its contribution to the overall capsid structure can now be appreciated. Substitutions of either Val 181 or Met 185 greatly impaired capsid assembly in vitro and reduced HIV-1 infectivity by ~10 to >1,000-fold (Fig. 2f). In addition, CA mutant virions exhibited a high frequency of defective particles (Extended Data Fig. 8a,b). Intriguingly, changes in Val 181 resulted in drastic morphological changes in vitro, from

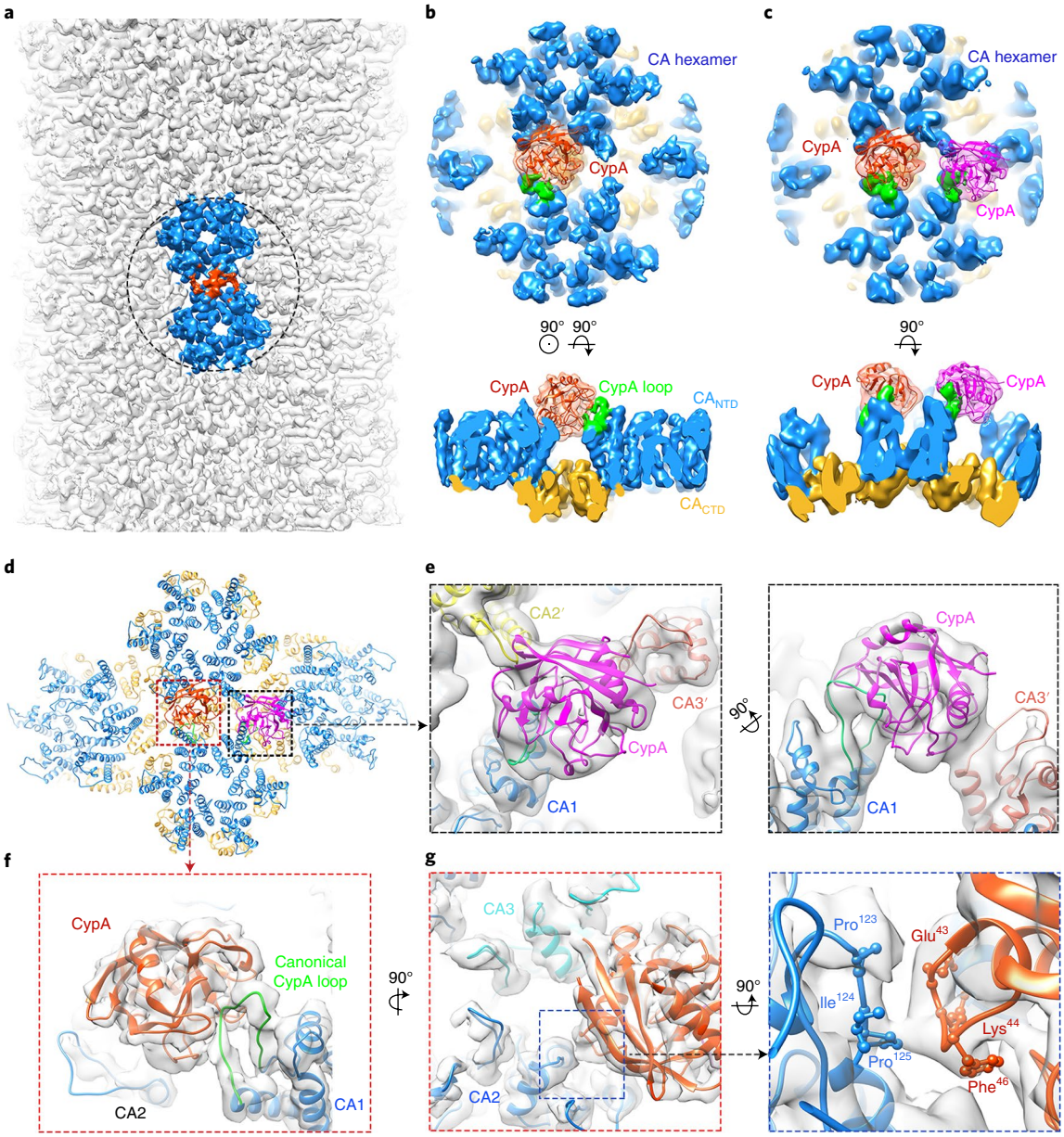


Fig. 3 | Cryo-EM structure of CypA bound to mature CA lattice. a, Surface representation of a CypA-decorated CA tube with (−13, 8) helical symmetry. The circled area includes two CA hexamers (blue) with CypA density (red) above the dimer interface that was extracted for symmetry expansion, localized reconstruction and classification. **b**, Cryo-EM structure of CypA/CA in Mode 1: one CypA interacting with two CA hexamers above the dimer interface. The canonical CypA-binding loop is colored green. **c**, Cryo-EM structure of CypA/CA in Mode 2: two CypA molecules on the CA lattice surface. One CypA (orange) interacts with two CA hexamers above the dimer interface as in **b**, whereas the other CypA (magenta) interacts with three CA hexamers above the trimer interface. **d**, Atomic model of a four-hexamer CA lattice with CypA. **e**, Enlarged views of the boxed region in **d** (dashed black) at the trimer interface. The three interacting CA molecules are labeled as CA1, CA2′ and CA3′, where CA1 harbors the canonical interface (shown in green). **f**, Enlarged views of boxed region in **d** (dashed red) at the dimer interface. The three interacting CA molecules are labeled CA1, CA2 and CA3. The CypA-binding loop on CA1 is shown in green. **g**, Interacting residues between CA2 and CypA (boxed region in **f**) are shown in ball and stick representation.

a completely flat lattice with V181C, to an extremely curved assembly with V181D (17-nm diameter compared to 50 nm in WT) (Extended Data Fig. 8c–i), reinforcing the critical role of the dimer interface in capsid assembly.

Cryo-EM structure of CypA–CA tubular assemblies. CypA is important for HIV-1 infection of human cells[34] and disruption of the CA–CypA interaction renders HIV-1 susceptible to potent restriction by human TRIM5α[35]. We previously proposed a second noncanonical interface between CypA and CA on the basis of a low-resolution cryo-EM map and molecular dynamic simulations[20]. To overcome the challenges of variable CypA-binding modes, partial occupancy and capsid asymmetry, we devised a data processing and 3D reconstruction strategy using symmetry expansion, localized 3D classification, alignment and reconstruction (Extended Data Fig. 2g–j), which yielded structures of CypA complexed with CA tubular assemblies from six helical symmetries to high resolution (Fig. 3, Supplementary Table 1 and Extended Data Fig. 9a,b). Localized 3D classification revealed two CypA-binding modes: Mode 1 with one CypA located above the CA dimer interface (Fig. 3b,

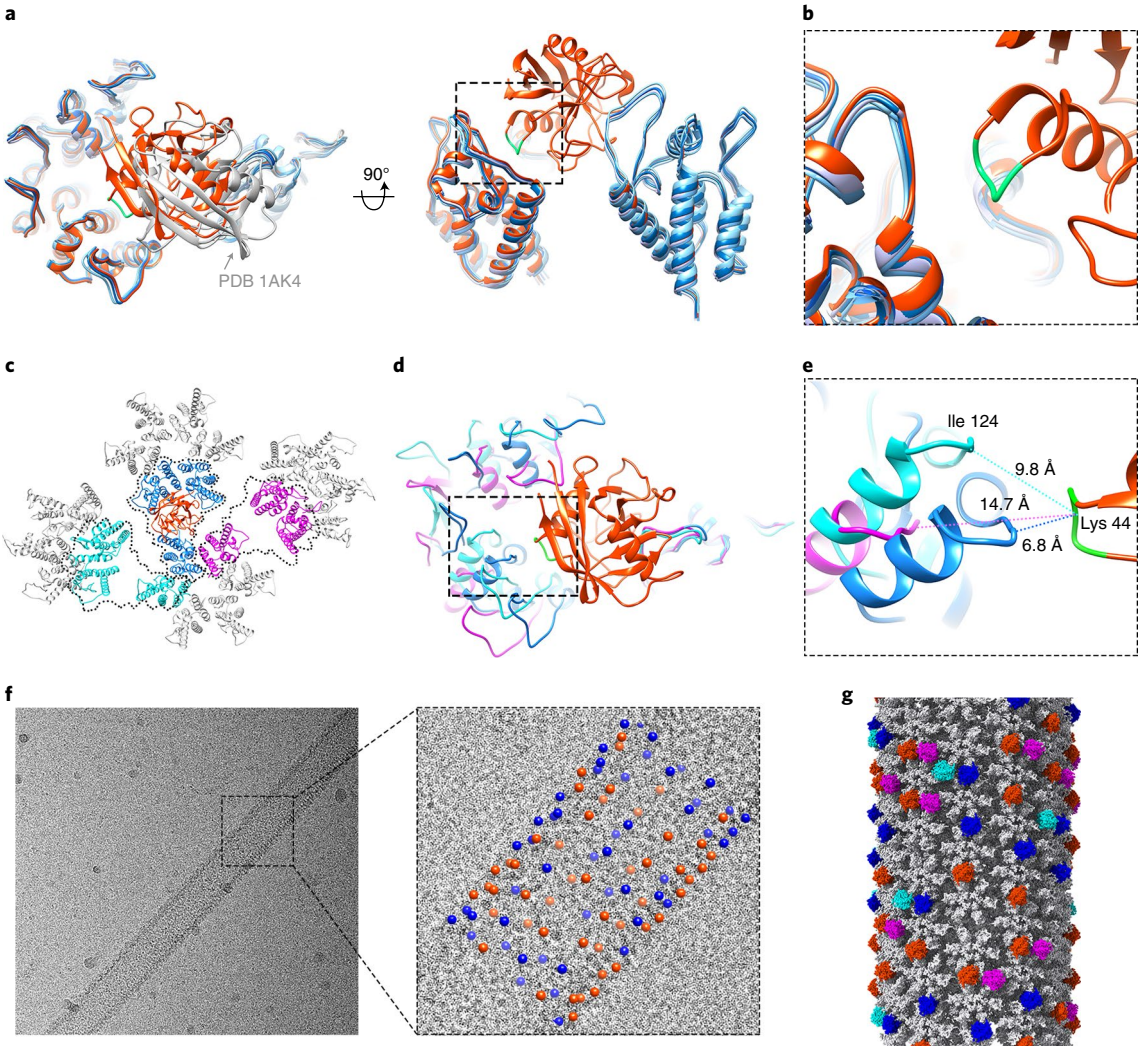


Fig. 4 | The noncanonical CypA/CA interaction is conserved and defined by CA lattice geometry. a, Alignment of CypA molecules above the dimer interface from six different helical symmetries (−8, 13), (−7, 13), (−13, 7), (−13, 8), (−13, 9) and (−13, 10), shown in two orthogonal views. The crystal structure of CypA/CA_{NTD} complex (PDB 1AK4) is colored in gray (left panel). The noncanonical interface is marked with a dashed box. **b**, An enlarged view of the noncanonical interface. **c**, Schematic representation of three possible CypA positions (dashed contours) above the dimer interfaces in the (−13, 8) CA lattice. **d**, Alignment of three CA molecules (cyan, blue and pink) from the CA lattice in **c**, along with a bound CypA (orange). **e**, Enlarged view of the boxed region in **d**, showing the difference in distance between Cα of Lys 44 (CypA) and Cα of Ile 124 of three CA protomers. **f**, A cryo-EM image of CA tubes decorated with CypA with zoomed-in view of the boxed region, overlaid with experimental CypA positions (only Mode 1 for clarity) projected onto the tube image. **g**, 3D representation of the tube segment shown in **f**, where the actual position of each CypA is mapped onto the tube surface according to its 3D class (see Extended Data Fig. 2). Orange and blue CypAs are above the dimer interface; magenta and cyan CypAs are above the trimer interface. The orange and magenta CypAs bind in an upward-pointing conformation, whereas blue and cyan CypAs bind pointing downward.

orange) and Mode 2 with an additional CypA above the CA trimer interface (Fig. 3c, magenta). In both binding modes, a single CypA simultaneously interacts with three different CA molecules, one via the canonical CypA-binding loop (Fig. 3b,c, green, and Fig. 3e,f, CA1) and the other two from neighboring hexamers via noncanonical interactions (Fig. 3e–g, CA2, CA3 and CA2′, CA3′). In particular, the second binding interface of CypA in Mode 1 overlaps with the site that we previously described[20] and involves residues Glu 43-Lys 44-Gly 45-Phe 46 and Pro 123-Ile 124-Pro 125 from CA2 (Fig. 3g). This interface appears to be weaker than the canonical interface, enclosing a 320.7-Å^2 interaction area, compared to 496.5 Å^2 for the canonical interface. Interestingly, in binding Mode 2, the two CypA molecules always orient the same way, with the canonical binding sites on adjacent protomers within a hexameric capsomere (Fig. 3c).

CypA senses specific capsid surface geometry. We have analyzed CypA–CA tubes from different helical symmetries and observed a preferential CypA-binding pattern on CA lattices (Extended Data Fig. 9c). To understand the mechanism of capsid pattern sensing by CypA, we aligned the CypA–CA_{NTD} complexes from six helical symmetries (Fig. 4a). All six symmetries with different curvatures present the same binding mode, which is distinct from the crystal structure[22] (Fig. 4a, gray, and Supplementary Video 1). More interestingly, the noncanonical interaction site is highly conserved among the six symmetries (Fig. 4a,b). Therefore, despite variably curved surface lattices, CypA recognizes a specific capsid surface geometry for its interaction. Since three adjacent CA molecules within a hexamer possess different geometry and curvature owing to asymmetry, it is reassuring that among three potential CypA sites above the dimer interface (Fig. 4c, dashed contours), only the

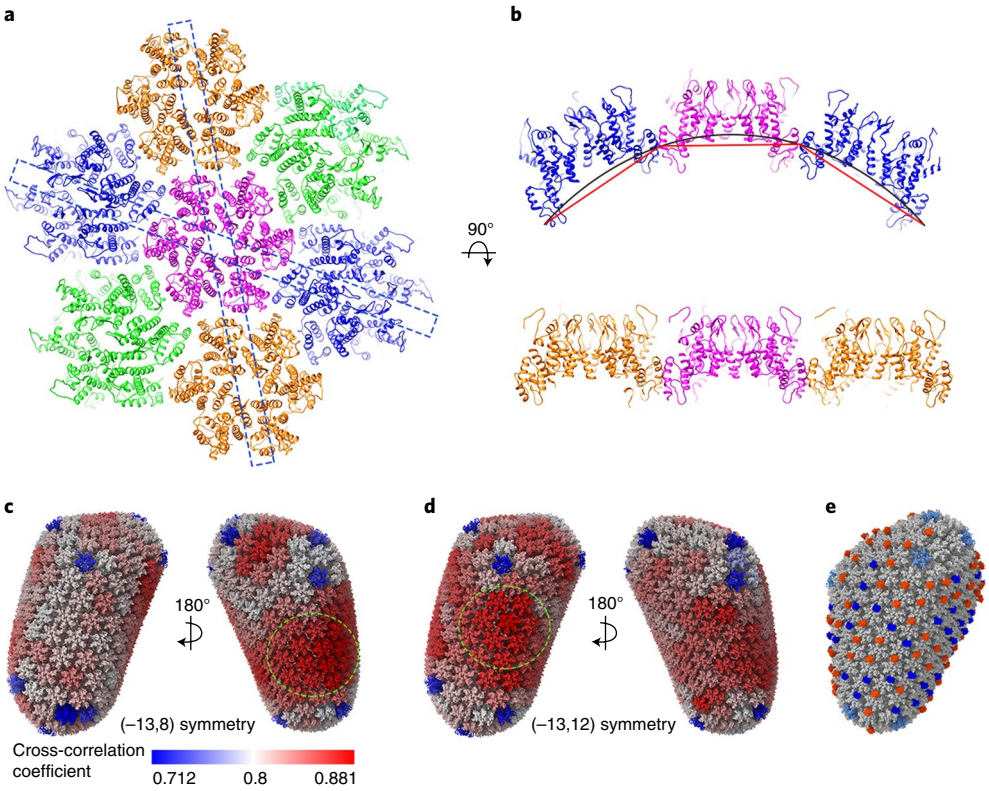


Fig. 5 | Continuous curvature formation in a conical capsid assembly and its role in CypA binding. a, Seven-hexamer local curvature unit, comprising a central curved, asymmetric hexamer (magenta) and three pairs of neighboring curved hexamers along three helical directions, with hexamers in gold nearly parallel to the tube axis (least curved) and hexamers in blue along the most curved direction. **b**, Clipped views with a 90° rotation from the boxed regions in **a**, enclosing blue–magenta hexamers (top) and gold–magenta hexamers (bottom). The top panel illustrates that the curvature is primarily derived from curved hexamers (black line), instead of from tilting and twisting between flat hexamers (red line). The bottom panel reveals the relatively flat gold-magenta-gold aspect the of local seven-hexamer lattice. **c**,**d**, Surface presentations of curvature variation in conical capsid assembly, colored according to the cross-correlation score between a seven-hexamer lattice from a conical core (PDB 3J3Q) and those from helical symmetry (−13, 8) (**c**) or (−13, 12) (**d**). The green circles mark the area of highest correlation value. **e**, Surface presentations of CypA (red and blue) binding to the conical capsid (PDB 3J3Q). The positions of CypA (Mode 1 only) on the capsid surface lattice were determined by matching the underlying seven-hexamer units of CypA/CA tubes (Fig. 4f) to the conical capsid model (PDB 3J3Q).

site with matching geometry is occupied (Fig. 4c–e, blue). We can further map every individual CypA (which has been classified into Mode 1 or Mode 2) back onto the initial tube from which it originated (Fig. 4f,g). The result clearly illustrates how CypA decorates the CA tubular surface lattice with a specific pattern of binding preference (Fig. 4g and Supplementary Video 2). There are significant overlaps between the binding sites for CypA and putative binding sites (four out of six) for TRIM5α[36], which explains how CypA can block human TRIM5α binding to preassembled cross-linked CA tubes in vitro[37], and protects HIV-1 from potent restriction by human TRIM5α in cell culture[35]. CypA is present at a reasonably high concentration (~58–109 μM) in activated $CD4^+$ T cells (Extended Data Fig. 10), indicating that incoming virions are likely exposed to saturating levels of the host factor.

Continuous curvature formation and CypA binding in the conical capsid. It has been proposed that the conical mature capsid spreads its surface curvature by continuously changing dimer and trimer interfaces between CA hexamers[6,12,38]. Cryo-EM structures of curved assemblies argue that these interfaces, on the contrary, are relatively stable. Our results indicate that the key contributor to the curvature is the intrinsically and variably curved hexameric capsomere (Fig. 5a,b), which originates from the highly conserved but flexible linker between the CA_{NTD} and CA_{CTD}. To investigate whether and how various curvatures observed in helical assemblies represent the local curvature in the native conical capsid, we cross-correlated a conical capsid model (PDB 3J3Q)[11] and the local curvature unit of seven CA hexamers from five different symmetries. As exemplified in Fig. 5 for (−13, 8) and (−13, 12) symmetries, the best-matched curvature units mapped to distinct regions of the conical capsid (Fig. 5c,d). As expected, we observed that the regions with the lowest cross-correlations correspond to locations of pentamers (Fig. 5c,d, blue). The result suggests that a continuously curved surface can be achieved through variations in local curvatures manifested in helical assemblies. Changes in local curvatures result in exposed surfaces of different sizes and/or geometries, which may alter the accessibility of different host cell factors, such as CypA[20], MxB[3,25] and Trim proteins, via specific capsid pattern recognition[2,3,20]. We illustrated this by mapping CypA binding to the conical capsid surface through cross-correlation-based match of underlying local curvature units from conical capsid and CypA–CA tubular assemblies (Fig. 5e and Supplementary Video 3). We suggest that the capacity for curvature adaptation is a conserved feature across the broad morphologies of different types of retroviral capsids.

Discussion

Infections by retroviruses, such as HIV-1, critically depend on their capsids. The capsid is the primary interface between the

virus and host cell during viral ingress. Many host cell defense proteins, including the restriction factors Trim5α, TrimCyp and MxB, target the HIV-1 capsid during the early stages of infection and potently inhibit virus replication. These restriction factors appear to function through capsid pattern sensing that specifically recognizes the assembled capsid as opposed to individual CA molecules. While numerous atomic structures of HIV-1 CA domains, monomers, hexamers and pentamers have been determined by NMR and X-ray crystallography methods[4–6,28,32], none of these was derived from curved capsid assemblies, limiting our understanding of how the variable capsid curvature is achieved and how host factors specifically recognize the curved lattice pattern to exert their functions. The cryo-EM structures of the curved CA hexamer presented here imply a different mechanism underlying capsid formation. Hexamer flexibility suggests that the CA subunits are more easily accommodated as the capsid lattice is built, rather than each CA subunit having to adopt a very precise conformation to fit into a rigid symmetrical hexamer first. Consequently, the binding of host factor CypA to the capsid surface lattice involves additional new interfaces that are absent in the crystal structure of the binary complex. Information derived from our study will potentially allow designs of more robust therapeutic agents that exploit the intrinsic curvature of the capsid to block HIV-1 replication.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-020-0467-8.

Received: 7 March 2020; Accepted: 19 June 2020;
Published online: 3 August 2020

References

1. James, L. C. & Jacques, D. A. The human immunodeficiency virus capsid is more than just a genome package. *Annu. Rev. Virol.* **5**, 209–225 (2018).
2. Li, Y.-L. et al. Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids. *Elife* **5**, 582 (2016).
3. Smaga, S. S. et al. MxB restricts HIV-1 by targeting the tri-hexamer interface of the viral capsid. *Structure* **27**, 1234–1245 (2019).
4. Gres, A. T. et al. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349**, 99–103 (2015).
5. Pornillos, O. et al. X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282–1292 (2009).
6. Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469**, 424–427 (2011).
7. Price, A. J. et al. Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLoS Pathog.* **10**, e1004459 (2014).
8. Yant, S. R. et al. A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model. *Nat. Med.* **25**, 1377–1384 (2019).
9. Mallery, D. L. et al. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *Elife* **7**, e1002762 (2018).
10. Dick, R. A. et al. Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509–512 (2018).
11. Zhao, G. et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).
12. Mattei, S., Glass, B., Hagen, W. J. H., Kräusslich, H.-G. & Briggs, J. A. G. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434–1437 (2016).
13. Pertel, T. et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361–365 (2011).
14. Sayah, D. M., Sokolskaja, E., Berthoux, L. & Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569–573 (2004).
15. Li, S., Hill, C. P., Sundquist, W. I. & Finch, J. T. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409–413 (2000).
16. Perilla, J. R. & Schulten, K. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun.* **8**, 15959 (2017).
17. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
18. Zhang, P., Meng, X. & Zhao, G. in *Electron Crystallography of Soluble and Membrane Proteins* (eds Schmidt-Krey, I. & Cheng, Y.) 381–399 (Humana Press, 2012).
19. Ilca, S. L. et al. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015).
20. Liu, C. et al. Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.* **7**, 10714 (2016).
21. Shah, V. B. et al. The host proteins transportin SR2/TNPO3 and cyclophilin A exert opposing effects on HIV-1 uncoating. *J. Virol.* **87**, 422–432 (2013).
22. Gamble, T. R. et al. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285–1294 (1996).
23. Schaller, T. et al. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.* **7**, e1002439 (2011).
24. Morger, D. et al. The three-fold axis of the HIV-1 capsid lattice is the species-specific binding interface for TRIM5α. *J. Virol.* **92**, e01541-17 (2018).
25. Fricke, T. et al. MxB binds to the HIV-1 core and prevents the uncoating process of HIV-1. *Retrovirology* **11**, 68 (2014).
26. Jiang, J. et al. The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology* **421**, 253–265 (2011).
27. Noviello, C. M. et al. Second-site compensatory mutations of HIV-1 capsid mutations. *J. Virol.* **85**, 4730–4738 (2011).
28. Byeon, I.-J. L. et al. Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780–790 (2009).
29. Saito, A. et al. Roles of capsid-interacting host factors in multimodal inhibition of HIV-1 by PF74. *J. Virol.* **90**, 5808–5823 (2016).
30. Rankovic, S., Ramalho, R., Aiken, C., Rousso, I. & Kirchhoff, F. PF74 reinforces the HIV-1 capsid to impair reverse transcription-induced uncoating. *J. Virol.* **92**, e00845-18 (2018).
31. Márquez, C. L. et al. Kinetics of HIV-1 capsid uncoating revealed by single-molecule analysis. *Elife* **7**, e1002762 (2018).
32. Momany, C. et al. Crystal structure of dimeric HIV-1 capsid protein. *Nat. Struct. Mol. Biol.* **3**, 763–770 (1996).
33. Bayro, M. J. & Tycko, R. Structure of the dimerization interface in the mature HIV-1 capsid protein lattice from solid state NMR of tubular assemblies. *J. Am. Chem. Soc.* **138**, 8538–8546 (2016).
34. Braaten, D., Franke, E. K. & Luban, J. Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J. Virol.* **70**, 3551–3560 (1996).
35. Kim, K. et al. Cyclophilin A protects HIV-1 from restriction by human TRIM5α. *Nat. Microbiol.* **4**, 2044–2051 (2019).
36. Skorupka, K. A. et al. Hierarchical assembly governs TRIM5α recognition of HIV-1 and retroviral capsids. *Sci. Adv.* **5**, eaaw3631 (2019).
37. Selyutina, A. et al. Cyclophilin A prevents HIV-1 restriction in lymphocytes by blocking human TRIM5α binding to the viral core. *Cell Rep.* **30**, 3766–3777.e6 (2020).
38. Obal, G. et al. Conformational plasticity of a native retroviral capsid revealed by X-ray crystallography. *Science* **349**, 95–98 (2015).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Protein expression and purification. The CA proteins (WT and A92E mutant) and CypA were expressed in *Escherichia coli* and purified by ion exchange chromatography and size-exclusion chromatography, as described previously[11,20].

Cryo-EM specimen preparation. Full-length CA A92E (12 mg ml^{-1}) was diluted to 2 mg ml^{-1} in high salt buffer (1 M NaCl, 50 mM Tris pH 8.0) and incubated at 37 °C for 1 h for tubular assembly. Sample (2.5 µl) was applied to the carbon side of a glow-discharged perforated Quantifoil grid (Quantifoil Micro Tools), followed by application of 3 µl of low salt buffer (100 mM NaCl, 50 mM Tris pH 8.0) on the back side of the grid and blotting from the back side with a filter paper, before plunge-freezing in liquid ethane using a manual gravity plunger.

CypA-stabilized WT CA tubes were prepared according to the procedure described previously[20] with addition of CypA to the preassembled WT CA at a CA:CypA molar ratio of 6:2.5. Sample (2 µl) was applied to the carbon side of a glow-discharged Quantifoil grid (R2/2), followed by an application of 2 µl of low salt buffer (500 mM NaCl, 50 mM Tris pH 8.0) on the back side of the grid, before blotting from the back side and plunge-freezing.

Cryo-electron microscopy data collection. For the HIV CA A92E tubes, data were manually collected on a Tecnai Polara microscope (FEI) operated at 300 kV at liquid nitrogen temperature. Images were recorded on a Gatan K2 summit direct electron detector at a nominal magnification of ×31,000, in super-resolution counting mode. The dose rate was set to ~8 electrons per physical pixel per second. The total accumulative electron dose was ~41 e^{-}/Å^{2} fractioned over 30 subframes with a total exposure time of 6 s. The defocus values were from 0.7 to 2.2 µm (Table 1).

For the CypA-stabilized WT CA tubes, data were collected on a ThermoFisher Krios G3 with a K2 Summit direct electron detector in counting mode using an in-house automated data collection software, ArbitrEM, developed on the SerialEM platform[17] (Extended Data Fig. 1). Using the SerialEM navigator, a series of 'holes' were selected and 'view-map' images acquired automatically using a script AcquireHoleViewMaps at a rate of ~100 holes per hour (×11,000–15,000, ~5 µm^{2} field-of-view, at 75 µm under-focus using an exposure time of 1–2 s). For each map, tubes of interest were identified and the acquisition targets marked by adding navigator points within SerialEM. A Python script (processNavigator.py) was run to compile this information for efficient use by the ArbitrEM script. Data collection using the ArbitrEM script was then started, during which each foil hole was revisited to within an accuracy of ±75 nm. For each hole, the stage *XY* offset of the revisited position from that of the original hole view-map position was measured and the targeting beam-image shifts (x and y) calculated according to: $(\Delta x,\Delta y)_n = AP_{\mathrm{stage}(X,Y)n} - \mathrm{Oc}_{(x,y)} - \mathrm{Vr}_{\mathrm{stage}(X,Y)} - \mathrm{Vt}_{\mathrm{stage}(X,Y)}$; where, for n acquisition points (AP) for a particular targeting view-map (Vt), the stage coordinates, stage(X,Y), are considered in relation to those of the revisited view-map image (Vr). Oc(x,y) is a convenience shift offset that allows for any systematic errors in alignment between view and acquisition magnifications to be corrected on-the-fly. For each view-map autofocusing was performed, and after appropriate delays for stage and image shift, movies were acquired. The electron dose before data acquisition was limited to 1.2 e^{-}/Å^{2}. The total accumulative electron dose was ~40 e^{-}/Å^{2} fractioned over 40 subframes with a total exposure time of 8 s (Table 1). Scripts, further details and example imaging conditions are provided on GitHub: https://github.com/kyledent/ArbitrEM.

A similar approach, py-EM, was recently published[39]. Both methods utilize the 'Realign to' function within SerialEM, but in different ways. Specifically, py-EM uses 'Realign to' as the sole mechanism of targeting, relying on multiple iterations of image acquisition and refinement of stage position for a single acquisition point (slower), whereas ArbitrEM uses SerialEM's 'Report Shift' in combination with 'Realign To' to calculate the displacement of the specimen from the view-map, then taking this into account before calculating a custom beam-image shift to achieve the imaging position specified by the user (faster). As a result, with ArbitrEM, custom beam-image shifts can be calculated on-the-fly (that is, contextually) to achieve high accuracy, allowing multiple sites to be targeted while minimizing the number of stage movements and therefore improving throughput. While the py-EM method greatly improves the ease of selection for 'large particles' observable within relatively low magnification, ArbitrEM was developed to allow targeting of 'finer objects', with a throughput close to standard single-particle acquisition.

Image processing for CA assemblies. Motion correction was performed using MotionCor2 (ref. [40]) and dose-weighted micrographs were further processed. Each tube was boxed out from the micrograph using the e2helixboxer.py program in EMAN2 (ref. [41]) and initially sorted into different groups by diameter (Extended Data Fig. 2). From each diameter group, the tubes were further sorted through manual inspection of their power spectra. All the tubes in the same symmetry group were reconstructed using an in-house script based on the iterative helical real-space reconstruction (IHRSR)[42] method to search for helical parameters: rise (Δz) and turn ($\Delta\phi$). Using this method, we processed the three most abundant helical symmetries in the CA A92E dataset, (−12,11), (−13,11) and (−13,12), and two helical symmetries, (−13,8) and (−13,12), in the WT CypA–CA dataset.

After being sorted on the basis of helical symmetry, the tubes belonging to each helical symmetry were processed separately using RELION[43]. Tubes were segmented using a box size of 720 pixels with an interbox distance of six subunits (6× helical rise). Two rounds of reference-free two-dimensional (2D) classification were performed to clean the dataset, and good particles were subjected to 3D classification and refinement in RELION using the initial model and helical parameters generated through the IHRSR method. To improve the resolution of hexamers, a localized reconstruction method was used. The refined helical segments were first subjected to particle symmetry expansion using relion_particle_symmetry_expand with the refined helical parameters, generating a six times larger dataset. The expanded dataset was imported into Scipion[44] to perform localized reconstruction[19]. Density subtraction was performed for each segment, leaving only one hexamer of interest in each segment. These hexamers were extracted at a box size of 120 pixels (CA A92E dataset) or 144 pixels (WT CA dataset) in Scipion and further refined applying *C*2 symmetry in RELION, with only local angular search and in-plane translation searches (Supplementary Table 1).

Image processing for CypA–CA assemblies. To classify the CypA-binding mode on CA tubes, the refined helical segments were first subject to particle symmetry expansion, as described above. Density subtraction was performed for each segment, leaving an area that included only one CypA density and local region of CA_{NTD} within each segment (boxed region in Extended Data Fig. 2e). The density-subtracted segment dataset was then imported to Scipion to perform localized reconstruction. The subparticles were extracted at a box of 144 pixels in Scipion and 3D classification without alignment was performed in RELION, with a regularization factor $T = 10$, and a resolution limit of 8 Å in the expectation step. Depending on the location of CypA, different binding modes were identified from 3D classification (Extended Data Fig. 2h). The classification and alignment results were applied to raw particles without density subtraction and the raw particles were further refined with only local angular and translational search. Two binding modes above the dimer and trimer interfaces were pooled, aligned and summed together to obtain the final density (Extended Data Fig. 2i,j).

Model building and refinement. The crystal structure of CA monomer (PDB 4XFX) was first manually rigid-body fitted into the density map in Chimera[45] to generate the initial hexamer. CA_{NTD} and CA_{CTD} were rigid-body fitted into the density map in Coot[46]. The β-hairpin (residues Pro 1–Glu 13) was modeled using that from CA monomer (PDB 3H4E) and further refined manually in Coot. The CypA-binding loop was not well resolved in the density map and the corresponding region from the X-ray crystallography monomer (PDB 4XFX) was used. The model was then subject to real-space refinement using phenix.real_space_refinement[47], without crystallographic symmetry restraint and secondary structure restraint. For CypA–CA complexes, the crystal structure of the CypA–CA_{NTD} complex (PDB 1AK4) and of the refined CA hexamer from tubular assemblies (this study) were used as initial models. CypA and CA_{NTD} were rigid-body fitted into the density, and were then aligned and combined with the CA hexamer model. The structure alignment was performed with Chimera and the surface contact analysis with PDBePISA[48].

Mapping the local curvature of a seven-hexamer unit from helical symmetries to a conical capsid. We initially intended to map the local curvature of a seven-hexamer unit from various helical symmetries to capsid models derived from the cryo-ET and subtomogram averaging study of Mattei et al.[12] (kind gift from J. Briggs). However, we found that the hexamers (and pentamers) in these capsid models were not connected, and the surface arrays were not continuous. We then used the all-atom MD simulation model (PDB 3J3Q) based on a tomography HIV-1 core[11,15]. The all-atom MD model was first converted into a density map using e2proc3d.py, with a resolution of 10 Å and pixel a size of 3 Å, given that the r.m.s.d. of hexamers in PDB 3J3Q is 3.5 ± 1.1 Å. The seven-hexamer local curvature unit from each helical symmetry was first aligned with each hexamer in PDB 3J3Q in six directions using the 'matchmaker' command in Chimera[45]. Each aligned seven-hexamer unit was fitted into the 10-Å resolution PDB 3J3Q converted map using the 'fitmap' command in Chimera. Six cross-correlation coefficients were obtained at each hexamer position in PDB 3J3Q with six potential orientations, and the highest one was used as the local cross-correlation value between the curvature unit in helical symmetry and that in PDB 3J3Q. The cross-correlation coefficient (from blue to red) in these hexamer positions was then mapped onto the PDB 3J3Q model and presented in ChimeraX[49].

Mapping CypA to the CA surface of the all-atom MD simulation model. To map CypA molecules to a helical segment, the positions of CypA on the 2D raw micrographs were first calculated using the localized reconstruction method. CypA positions were derived from the 3D classification in RELION (Extended Data Fig. 2h). The corresponding CypA-binding modes were then mapped to the 3D helical model (Fig. 4g).

A CypA-bound capsid model was generated according to the binding preference of CypA to CA tubes observed in various helical symmetries and their local geometries (Fig. 4f and Extended Data Fig. 9). For the same reason as above, we used the all-atom MD simulation model (PDB 3J3Q). The six cross-correlation coefficients between the curvature unit in the tubular assemblies and that in PDB 3J3Q at each hexamer position were used to determine which local geometry satisfies CypA binding. For each seven-hexamer local curvature unit in the conical

capsid model, two CypA molecules were initially placed above the dimer interfaces among the central least-curved three-hexamer array in the same manner as in the helical assembly. The clashing CypA molecules were then randomly removed to generate the final model.

HIV-1 CA assembly assay. HIV-1 WT CA and mutant proteins were diluted to 2 mg ml^{-1} in high salt buffer (1 M NaCl, 50 mM Tris pH 8.0) and incubated at 37 °C for 1 h for tubular assembly. Samples (6 µl) were withdrawn from assembly reaction mixtures for negative staining on transmission electron microscopy (TEM) grids with 2% uranyl formate. The morphologies of variant of CA assemblies were characterized by TEM using a Tecnai T12 transmission electron microscope at 120 kV.

The remaining samples were pelleted at 20,000*g* with an Eppendorf centrifuge 5417R for 15 min, and supernatants (s) and pellets (p, resuspended in the same volume) were mixed with 4× LDS loading buffer (Invitrogen) for gel. Supernatant and pellet samples, without boiling, were loaded on 10% SDS–PAGE and stained with Coomassie Blue.

Cells, viruses and cores. The 293T, HeLa and TZM-bl cells were cultured in DMEM (Cellgro), supplemented with 10% FBS, penicillin (50 IU ml^{-1}) and streptomycin (50 µg ml^{-1}), in a humidified 37 °C incubator containing 5% CO_2. Point mutations in the CA region of the R9 HIV-1 proviral plasmid were introduced by PCR mutagenesis. The mutant viral fragments generated by PCR were transferred into R9 using Spe1 and ApaI restriction enzymes. The corresponding regions of the mutant inserts within the viral plasmids were sequenced to confirm the presence of these mutations and the absence of extraneous mutations.

Virus stocks were produced by calcium phosphate transient transfection of 293T cells. Plasmid DNA (20 µg) was used per 2.0×10^6 cells in 10-cm dishes, as previously described[50]. The supernatant was removed from the plate 36–48 h after transfection, and the monolayer was gently rinsed with sterile PBS before replenishing the media. The culture supernatants were collected 24 h later and filtered through 0.45-µm pore-size filters. Viral aliquots were frozen at −80 °C and virus production quantified by p24 ELISA, as previously described[51].

Single-cycle assay of HIV-1 infectivity. Viral infectivity was assessed by titration on TZM-bl cells: HIV-1 stocks were serially diluted in a culture medium containing polybrene (8 µg ml^{-1}), and 100-µl volumes were added to TZM-bl target cells seeded the day before (15,000 cells per well in 96-well plates). The cultures were maintained for an additional 48 h before lysis and detection with a luciferase substrate. To determine the extent of infection, the luminescence reading in relative light units (RLU) was quantified in a Molecular Devices MaxL instrument. The corrected RLU values were normalized by room temperature activity to calculate the specific infectivity of the viruses in RLU ng^{-1} p24, using luciferase values from infections performed with virus dilutions that fell within the linear range of the assay.

Transmission electron microscopy analysis of HIV-1 particles. Cultured 293T cells (10^7 in 15-cm dishes) were transfected with 30 µg of plasmid using PolyJet In Vitro DNA Transfection Reagent for lentivirus production (SignaGen Laboratories) according to the manufacturer's protocol. At 48 h post-transfection, culture medium was collected and centrifuged to remove cell debris. Clarified supernatant was filtered using a 0.45-µm filter, directly into a Beckman polyallomer ultracentrifuge tube 13 mm × 51 mm (final volume ~38 ml). Particles were pelleted by ultracentrifugation using a Beckman SW32-Ti rotor at 26,000 r.p.m. for 2.5 h at 4 °C. Supernatants were aspirated and pellets fixed by addition of 1 ml of 1× fixative solution (1.25% paraformaldehyde, 2.5% glutaraldehyde, 0.03% picric acid in 0.1 M sodium cacodylate buffer, pH 7.4) directly into the ultracentrifuge tube and incubation at 4 °C for 24 h. Pellets were sectioned and stained by the Harvard Electron Microscopy Facility and particle morphologies quantified as previously described[52].

Semi-quantitation of CypA level in primary CD4$^+$ T cells. HIV- and HCV-negative human blood or buffy coats were purchased from Research Blood Components. Primary CD4$^+$ T cells were isolated and cultured as described[53]. Briefly, peripheral blood mononuclear cells were isolated by Ficoll gradient centrifugation and CD4$^+$ T cells were enriched by negative selection using EasySep T cell isolation kit (StemCell Technologies). As assessed by immunostaining with anti-CD4-PE (Miltenyi Biotec) and flow cytometry, CD4$^+$ T-cell purity was >96%. CD4$^+$ T cells (2.5×10^6 per ml) were cultured in RPMI 1640 medium containing 10% FBS, 100 IU ml^{-1} penicillin, 0.1 mg ml^{-1} streptomycin, 5 mM sodium pyruvate and 5 mM HEPES. Cells were activated in the presence of 5 µg ml^{-1} of phytohemagglutinin plus 20 U ml^{-1} interleukin-2 for 2–3 d.

CypA levels from two different blood donors were calculated by semi-quantitative western blotting using purified recombinant CypA protein (My BioSource (MBS2012356) reconstituted in 20 mM Tris–HCl, 150 mM NaCl (pH 8.0)) as a standard curve. Cells (25,000) washed twice with PBS were lysed by sonication in 2× Laemmli sample buffer, followed by heating at 95 °C for 10 min. Following fractionation by SDS–PAGE, cell proteins were transferred to PVDF membranes, which were reacted with antibodies against CypA (Abcam, ab41684) or lamin B1 (Abcam, ab16048). Membranes were developed using ECL prime reagent (Amersham Biosciences) and imaged with a ChemiDoc MP imager (Bio-Rad). Quantification of western blot signal intensities was done by Image Lab v.5.2.1 software (Bio-Rad). Based on the molecular mass of 18.0 kDa, we estimated approximately 7.6×10^6 molecules of CypA per cell. Given that cells contain approximately $7–13 \times 10^{10}$ total proteins at the concentration of 1 M and approximate lymphocyte volume of 110–200 µm^3 (refs. [54–56]), we estimate the concentration of CypA in activated CD4$^+$ T cells at ≈58–109 µM.

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

Data availability

Cryo-EM density maps for CA A92E hexamers are deposited at the EMDB under accession codes EMD-10240 for (−13, 11), EMD-10228 for (−13, 12) and EMD-10239 for (−12, 11). Cryo-EM density maps for CypA-stabilized CA WT hexamers are deposited at the EMDB under accession codes EMD-10226 for (−13, 8), EMD-10246 for (−13, 12) and EMD-10229 for (−13,8) tube map. Cryo-EM density maps for the CypA/CA complexes are deposited at the EMDB under accession codes EMD-10740 (−13, 7), EMD-10739 (−13, 8), EMD-10742 (−13, 9), EMD-11176 (−13, 10), EMD-10738 (−8, 13) and EMD-10741 (−7, 13). Real-space refined atomic models of the CA hexamers are deposited at the Protein Data Bank under accession codes PDB 6SLU (−13, 11), PDB 6SKM (−13, 12), PDB 6SLQ (−12, 11), PDB 6SKK (−13, 8), PDB 6SMU (−13, 12) and PDB 6SKN (−13,8) tube. Real-space refined atomic models of the CypA/CA complexes are deposited at the Protein Data Bank under accession codes PDB 6Y9X (−13, 7), PDB 6Y9W (−13, 8), PDB 6Y9Z (−13, 9), PDB 6ZDJ (−13, 10), PDB 6Y9V (−8, 13) and PDB 6Y9Y (−7, 13). Source data are provided with this paper.

Code availability

The software is freely available from GitHub: https://github.com/kyledent/ArbitrEM.git.

References

39. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).
40. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
41. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
42. Egelman, E. H. Reconstruction of helical filaments and tubes. *Meth. Enzymol.* **482**, 167–183 (2010).
43. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
44. de la Rosa-Trevín, J. M. et al. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
45. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
46. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
47. Afonine, P. V. et al. Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).
48. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
49. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
50. Wang, W. et al. Inhibition of HIV-1 maturation via small-molecule targeting of the amino-terminal domain in the viral capsid protein. *J. Virol.* **91**, e02155-16 (2017).
51. Aiken, C. Pseudotyping human immunodeficiency virus type 1 (HIV-1) by the glycoprotein of vesicular stomatitis virus targets HIV-1 entry to an endocytic pathway and suppresses both the requirement for Nef and the sensitivity to cyclosporin A. *J. Virol.* **71**, 5871–5877 (1997).
52. Jurado, K. A. et al. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. *Proc. Natl Acad. Sci. USA* **110**, 8690–8695 (2013).
53. Hultquist, J. F. et al. CRISPR–Cas9 genome engineering of primary CD4+ T cells for the interrogation of HIV-host factor interactions. *Nat. Protoc.* **14**, 1–27 (2019).
54. Downey, G. P. et al. Retention of leukocytes in capillaries: role of cell size and deformability. *J. Appl. Physiol.* **69**, 1767–1778 (1990).
55. Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **35**, 1050–1055 (2013).
56. Schmid-Schönbein, G. W., Shih, Y. Y. & Chien, S. Morphometry of human leukocytes. *Blood* **56**, 866–875 (1980).

Acknowledgements

We thank C. Liu and J. Liu for help with initial manual data collection, Y. Cheng for access to the UCSF Polara and K2 detector, J. Dong and R. Esnouf for computer system support, J. Perilla for scientific discussion and T. Brosenitsch for critical reading of the manuscript. We thank D. Mastronarde for discussions regarding ArbitrEM development. This work was supported by the National Institutes of Health (P50AI150481, P.Z., C.A., A.N.E.), the UK Wellcome Trust Investigator Award 206422/Z/17/Z (P.Z.) and the UK Biotechnology and Biological Sciences Research Council grant BB/S003339/1 (P.Z.). S.G. is funded by a Wellcome Trust PhD Studentship. We acknowledge Diamond for access and support of the Cryo-EM facilities at the UK Electron Bio-Imaging Centre (eBIC, proposal EM14856 and NT21004, P.Z.), funded by the Wellcome Trust, MRC and BBSRC.

Author contributions

P.Z. conceived the research and, with A.N.E. and C.A., designed the experiments. K.D. developed ArbitrEM for automated targeted data collection. T.N. and J.N. prepared samples for cryo-EM. T.N., K.D. and G.Z. collected cryo-EM data. T.N., S.G. and G.Z. analyzed cryo-EM data. T.N. refined the atomic models. J.N. performed biochemical analysis and negatively stained EM. J.Z., J.S., J.A.-D. and W.L. performed virus infectivity assays and sectioning and imaging particle production. S.J. prepared primary $CD4^+$ T cells and determined the intracellular concentration of CypA. T.N. and P.Z. analyzed atomic models and wrote the paper with support from all the authors.

Competing interests

A.N.E. over the past 12 months has received fees from ViiV Healthcare, Co. for work unrelated to this topic. No other authors declare the potential for competing financial interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-020-0467-8.

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-020-0467-8.

Correspondence and requests for materials should be addressed to P.Z.

Peer review information Peer reviewer reports are available. Inês Chen was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

CellPress
OPEN ACCESS

Molecular Cell

Article

Structure of the Inhibited State of the Sec Translocon

Samuel F. Gérard,[1] Belinda S. Hall,[2] Afroditi M. Zaki,[1] Katherine A. Corfield,[2] Peter U. Mayerhofer,[2] Catia Costa,[3] Daniel K. Whelligan,[4] Philip C. Biggin,[1] Rachel E. Simmonds,[2,5,*] and Matthew K. Higgins[1,5,6,*]
[1]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
[2]School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK
[3]Surrey Ion Beam Centre, University of Surrey, Surrey, UK
[4]Department of Chemistry, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: rachel.simmonds@surrey.ac.uk (R.E.S.), matthew.higgins@bioch.ox.ac.uk (M.K.H.)
https://doi.org/10.1016/j.molcel.2020.06.013

SUMMARY

Protein secretion in eukaryotes and prokaryotes involves a universally conserved protein translocation channel formed by the Sec61 complex. Unrelated small-molecule natural products and synthetic compounds inhibit Sec61 with differential effects for different substrates or for Sec61 from different organisms, making this a promising target for therapeutic intervention. To understand the mode of inhibition and provide insight into the molecular mechanism of this dynamic translocon, we determined the structure of mammalian Sec61 inhibited by the *Mycobacterium ulcerans* exotoxin mycolactone via electron cryo-microscopy. Unexpectedly, the conformation of inhibited Sec61 is optimal for substrate engagement, with mycolactone wedging open the cytosolic side of the lateral gate. The inability of mycolactone-inhibited Sec61 to effectively transport substrate proteins implies that signal peptides and transmembrane domains pass through the site occupied by mycolactone. This provides a foundation for understanding the molecular mechanism of Sec61 inhibitors and reveals novel features of translocon function and dynamics.

INTRODUCTION

Biogenesis of eukaryotic secretory or transmembrane proteins often occurs at the surface of the endoplasmic reticulum through the process of co-translational translocation. Here proteins targeted to the membrane by a signal sequence are translocated during translation through a membrane-spanning protein conduit formed by the Sec61$\alpha\beta\gamma$ complex. An analogous system, SecYEG, is found in the plasma membrane of bacteria, where it also mediates protein secretion. The structure of the heterotrimeric Sec complex is conserved throughout evolution and contains the core channel-forming subunit Sec61α/SecY together with the smaller subunits Sec61β/SecG and Sec61γ/SecE (Van den Berg et al., 2004; Voorhees et al., 2014). The translocon forms a gated channel that maintains membrane integrity while selectively opening to allow passage of the unfolded polypeptide chain. In addition, the translocon must open laterally to allow the transmembrane helices of membrane proteins to pass into the membrane environment as they are translated.

Structural studies of the translocon have provided snapshots of how this dynamic complex operates. Sec61α/SecY is formed from 10 transmembrane helices that are arranged in a "clamshell" architecture (Van den Berg et al., 2004). In the resting state, this forms a transmembrane channel, which is occluded by a central constriction of hydrophobic residues, known as the pore ring, and a short plug helix, which blocks the lumenal exit. A second potential opening, the lateral gate, lies at the interface between helices H2/H3 and H7/H8. This gate is stabilized by a polar cluster and is proposed to open to provide a route for translocating transmembrane helices to pass into the membrane (Van den Berg et al., 2004). During protein translocation, the plug helix is displaced to open the channel, whereas the dynamic opening and closing of the lateral gate is thought to allow regions of the elongating polypeptide to partition into the endoplasmic reticulum lumen or the membrane environment, depending on their hydrophobicity.

More recently, cryoelectron microscopy studies have allowed visualization of ribosome-translocon complexes in different stages of translocation. These show that ribosome binding primes the translocon, loosening the lateral gate and breaking the polar cluster (Voorhees et al., 2014). During co-translational translocation, the N-terminal signal peptide interacts with a hydrophobic patch of residues near the lateral gate, occupying the position previously taken by helix H2 of Sec61α (Voorhees and Hegde, 2016a). In this state, the plug helix is displaced, opening the channel for translocation. The translocon can also operate post-translationally, with the allosteric activators

 © 2020 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

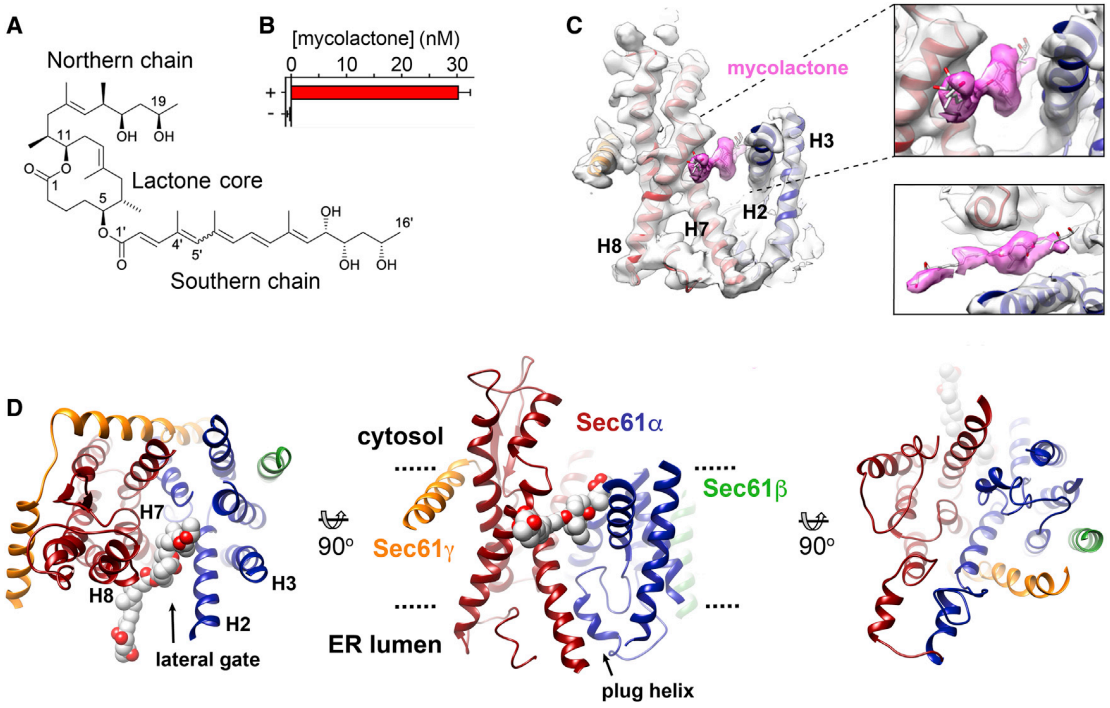


Figure 1. The Structure of the Sec61 Translocon Inhibited by Mycolactone

(A) The structure of mycolactone A/B (743 Da). The 12-membered lactone core is indicated, as are the two polyketide side chains, commonly referred to as the northern and southern chains. Mycolactone A/B is a 3:2 rapidly equilibrating mixture of *Z*-Δ4′,5′ and *E*-Δ4′,5′ geometric isomers at the second double bond in the southern fatty acid tail.

(B) Mycolactone concentrations in purified ribosome-translocon complexes extracted from membranes treated (+) or not treated (–) with mycolactone, determined by high-resolution LC-MS with the extracted ion chromatogram peaks integrated to obtain the $[mycolactone+Na]^+$ (m/z 765.4721–765.5103) ion eluting at 2.38–2.98 relative to the calibration curve.

(C) Electron density for Sec61α in the presence of mycolactone. Electron density maps were low-pass-filtered to 5 Å; the density feature corresponding to mycolactone is colored pink. The insets show two close-up views centered on the mycolactone density.

(D) The structure of the Sec translocon with Sec61α colored, with helices H1–H5 in blue, H6–H10 in red, Sec61β in green, and Sec61γ in orange. Mycolactone is shown as spheres, with carbon in white and oxygen in red.

See also Figures S1–S5.

Sec62 and Sec63 stabilizing an activated conformation of Sec61α in which the cytoplasmic side of the lateral gate is open (Itskanov and Park, 2019; Wu et al., 2019) but the plug helix remains in place (Wu et al., 2019). This prepares the translocon to bind and be gated by weak signal peptides.

These studies reveal that the translocon is dynamic, with structural transitions modulated by associations with the signal peptide (Gouridis et al., 2009) or allosteric activators (Itskanov and Park, 2019; Wu et al., 2019). Open states are stabilized by binding partners that hold helices H2 and H3 in a conformation in which the lateral gate is open and the plug helix is displaced (Gemmer and Förster, 2020; Lang et al., 2017; Rapoport et al., 2017; Voorhees and Hegde, 2016b). There are also a number of inhibitors that act on Sec61α but no detailed molecular insight into how they modulate the structure or conformational dynamics of the translocon.

Two studies showed that selective inhibitors of protein secretion can block co-translational translocation by acting directly on Sec61 (Besemer et al., 2005; Garrison et al., 2005). Inhibitors of Sec61-dependent translocation that act in this way include mycolactone, apratoxin A, cotransin, and ipomoeassin F (Cross et al., 2009; Hall et al., 2014; Haßdenteufel et al., 2018; Junne et al., 2015; Mackinnon et al., 2014; Paatero et al., 2016; Zong et al., 2019). These molecules have diverse structures but compete when binding to Sec61α, suggesting that they bind to overlapping sites (Baron et al., 2016; Paatero et al., 2016; Zong et al., 2019). Indeed, mutations of Sec61α that cause resistance to one inhibitor often also cause resistance to the others, suggesting a shared mode of action. Despite this, different inhibitors show variable specificity for translocons from different species and block translocation of different classes of substrates, raising the possibility of developing therapeutically useful selective translocon modulators. Of the current inhibitors, mycolactone (Figure 1A) is the most potent (Hall et al., 2014). This diffusible lipid-like exotoxin is synthesized by the Buruli ulcer pathogen *Mycobacterium ulcerans* (Demangel and High, 2018; Yotsu et al., 2018) and forms a stable complex with Sec61α (Baron et al., 2016). It prevents co-translational translocation of secretory proteins, including inflammatory mediators and cytokines, at nanomolar concentrations (Baron et al., 2016; Hall et al., 2014; McKenna et al., 2016) and blocks Sec61-dependent insertion of many transmembrane proteins (Baron et al., 2016; McKenna et al., 2017). Mycolactone inhibits translocation at a stage after ribosome engagement with the

Table 1. Refinement and Model Statistics

-	Ribosome: Sec61 + Mycolactone	Ribosome: Sec61
Data Collection and Processing		
Magnification	165,000	165,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	49	49
Defocus range (μm)	−1.0 to −2.5	−1.0 to −2.5
Pixel size (Å)	0.822	0.822
Symmetry imposed	C1	C1
Initial particle images (no.)	108,129	49,733
Final particle images (no.)	45,733	24,852
Map resolution (Å)	2.63	2.85
Fourier Shell Correlation threshold	0.143	0.143
Refinement		
Map pixel size (Å)	1.033	
Map sharpening *B* factor (Å^2)	5.7	
Map lowpass filter (Å)	4.0	
Model resolution (Å)	4.0	
Model Composition		
Non-hydrogen atoms	3,608	
Protein residues	441	
Ligands	MYC: 1	
Average B Factors (Å2)		
Protein	65.98	
Ligands	71.85	
Root-Mean-Square Deviations (RMSDs)		
Bond lengths (Å)	0.006	
Bond angles (°)	1.047	
Validation		
MolProbity score	1.79	
Clashscore	5.40	
Poor rotamers (%)	0.53	
Ramachandran Plot		
Favored (%)	91.61	
Allowed (%)	8.39	
Disallowed (%)	0.00	

translocon and affects the interaction of signal peptides (McKenna et al., 2016). The availability of this potent and stably bound inhibitor provided us with the opportunity to visualize the translocon trapped in the inhibited state.

RESULTS

Cryoelectron Microscopy of Mycolactone-Inhibited Ribosome-Translocon Complexes

Although there is, to date, no high-resolution crystal structure of a eukaryotic Sec61 homolog, cryoelectron microscopy of ribosome-translocon complexes has allowed visualization of translocons with different binding partners or captured in different states. We therefore purified ribosome-translocon complexes (RTCs) from canine microsomal membranes that had been incubated with mycolactone at a concentration that completely prevented prepro-α factor translocation (Figure S1). Liquid chromatography-mass spectrometry (LC-MS) was used to confirm the presence of mycolactone in these complexes (Figure 1B; Figure S2). In addition, to ensure that the observed changes were due to the presence of mycolactone rather than due to details of our preparation or imaging protocols, we prepared a control sample of ribosome-translocon complexes in the absence of inhibitor.

We next prepared grids to allow visualization of these samples by cryoelectron microscopy. Although samples prepared in standard holey grids or grids overlaid with a 2-nm layer of carbon showed preferred orientations, complexes vitrified on graphene oxide-coated grids allowed imaging of multiple views and generation of three-dimensional reconstructions (Figures S3–S5; Table 1). Ribosomes were predominantly translocon bound, and particle subtraction and focused classification on mycolactone-bound translocons yielded a map with an overall resolution of 2.6 Å and local resolutions for Sec61α from 2.4–8.7 Å (Figure 1; Figure S3). The resolution is highest for the C-terminal ribosome-bound half of Sec61α (H6–H10), whereas the N-terminal half (H1–H5) is more dynamic and less clearly defined (Figure S3). A model was built containing all transmembrane helices of Sec61α, the majority of Sec61γ, and a poly-alanine helix for Sec61β. The control dataset generated an electron density map at an overall resolution of 2.8 Å in which the positions of the helices were essentially indistinguishable from the previously reported structure of a primed translocon (Voorhees et al., 2014), indicating that the changes observed in the mycolactone-bound translocon were due to the presence of mycolactone (Figures S5). An elongated density within the cytosolic side of the lateral gate of the translocon was the only difference between mycolactone-bound and mycolactone-free translocons that was not due to helix movement and was attributed to mycolactone (Figure 1C). A comparison of mycolactone-bound and mycolactone-free ribosome-translocon complexes revealed no differences in the conformation of the ribosome, the occupancy of tRNA-binding sites, or the presence of associated proteins because of the presence of mycolactone (Figures S1 and S3–S5), supporting the finding that mycolactone does not directly affect translation elongation (Hall et al., 2014).

Mycolactone Stabilizes a Partially Activated Conformation of Sec61

A comparison of the conformations of the mycolactone-bound and mycolactone-free translocons (Voorhees et al., 2014) identified no differences in the ribosome-bound C-terminal half of Sec61α. However, in the presence of mycolactone, a major structural change was observed within the more flexible N-terminal half, with an ~9-Å movement of the cytoplasmic ends of helices H2 and H3 and smaller tilting of helix H4. This movement of H2 and H3 away from H7 and H8 results in opening of the cytoplasmic side of the lateral gate (Figure 2). Toward the lumenal side of the translocon lies the plug helix. This plug is present in the primed conformation of the translocon (Voorhees et al., 2014) but is displaced in the signal peptide-bound conformation, opening the channel (Voorhees and Hegde, 2016a). In the presence of mycolactone, the plug helix is clearly visible in the

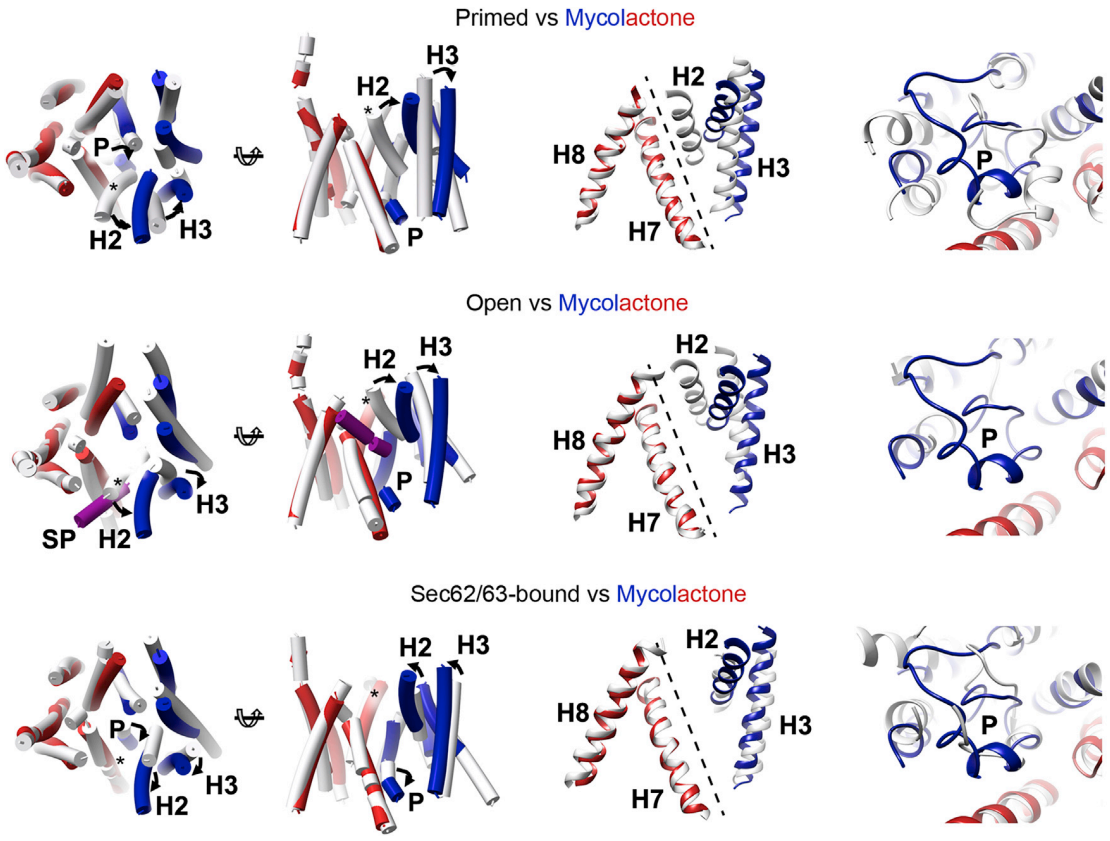


Figure 2. The Mycolactone-Stabilized Conformation of the Sec Translocon

Structural overlays of the mycolactone-bound conformation of Sec61α, with helices H1–H5 in blue and helices H6–H10 in red. This has been overlaid with the primed conformation (top panel; PDB: 3J7Q), the open signal peptide-bound conformation (center; PDB: 3JC2), and the Sec62/63-bound conformation (bottom; PDB: 6ND1), all in white. The left panel shows the helices as cylinders. The center panel shows helices H2, H3, H7, and H8 in cartoon representation, with the lateral gate shown as a dotted line. The right panel shows a close up centered on the plug helix (P).

electron density maps, retaining its α-helical character but showing a displacement of ~7 Å from its location in the primed state. However, this displacement is not sufficient to allow channel opening, with the plug helix still occluding the channel (Figure 2).

Comparison of the structure of the mycolactone-bound translocon with those of eukaryotic primed (Voorhees et al., 2014), signal peptide-bound (Voorhees and Hegde, 2016a), and Sec62/63-bound states (Itskanov and Park, 2019; Wu et al., 2019; Figure 2) reveals mycolactone to stabilize a conformation of Sec61α most similar to that observed in the presence of a complex of Sec62 and Sec63. This conformation of Sec61α is associated with post-translational translocation, in which the translocon is poised to accept substrates and to be gated by inefficient signal peptides. Indeed, stabilization of this state by Sec62/63 is thought to facilitate translocation, making it surprising that this is the state stabilized by a translocation inhibitor.

A Proposed Binding Site for Mycolactone at the Cytoplasmic Entrance of the Channel

Comparison of the electron density maps for mycolactone-bound and mycolactone-free translocons (Voorhees et al., 2014) reveals a single density feature that was not attributed to helix movement (Figure 1C). Because mycolactone must be present in the translocon to account for stabilization of this conformation, we docked mycolactone into this density. This binding pocket lies between helices H2 and H8 of Sec61α, within a hydrophobic groove created by opening of the lumenal side of the lateral gate (Figure 3A). Residues that line this groove and contact mycolactone include hydrophobic side chains that form part of the pore ring (V85 and I179) and hydrophobic patch (L89 and I179) of Sec61α and are thought to play important roles in channel closure and signal peptide binding (Figures 3B and 3C). The electron density attributed to mycolactone consists of a wider central region into which we docked the lactone core of mycolactone. On either side of this are two narrower regions of density likely to correspond to the polyketide side chains. These are not sufficiently well defined to precisely place mycolactone because of the ~5-Å resolution of this region of the map (Figure S3) coupled with the potential for mycolactone isomerization (George et al., 1999) and the inherent flexibility of its polyketide chains.

To further assess the pose of mycolactone within this binding site and to determine whether it remains stably bound during

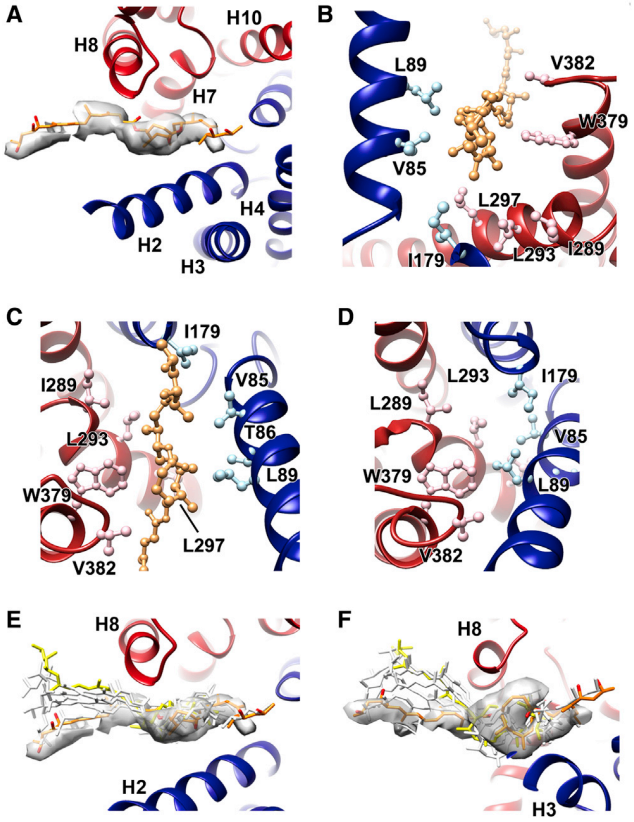


Figure 3. The Mycolactone Binding Site and the Location of Residues Whose Mutation Leads to Mycolactone Resistance

(A) Structure of Sec61α, with helices H1–H5 in blue and H6–H10 in red. The electron density for mycolactone is shown as a white surface.

(B) A close up of the mycolactone binding site, highlighting hydrophobic residues that lie in this pocket.

(C) A second view of the mycolactone binding site from the cytoplasmic face, viewed approximately perpendicular to the membrane.

(D) A view of the primed conformation of Sec61α, viewed from the same direction as in (C).

(E and F) Two views of the outcome of molecular dynamics simulations in which mycolactone was allowed to move within a restrained translocon. The original position of mycolactone, determined by fitting into the electron density, is shown as orange sticks, whereas gray sticks show variation in the position of mycolactone through a 100-ns trajectory, and yellow sticks show the position of mycolactone at the end of the trajectory. The electron density for mycolactone is shown as a white surface.

See also Figure S6.

simulation, we used molecular dynamics (Figures 3E and 3F; Figure S6). Two possible orientations of mycolactone could fit within the electron density, with the longer southern chain protruding into the hydrophobic environment of the lipid bilayer or into the core of the translocon. To assess the likelihood of each possible orientation, we modeled the *E*-Δ4′,5′ isomer of mycolactone, the presumptive active form (Gehringer et al., 2019), into this density in either possible orientation, and molecular dynamics simulations were conducted to assess their stability. In both cases, mycolactone remained stably bound over the 100-ns time course of the simulation. However, the conformation in which the southern chain protrudes into the lipid bilayer showed less variation during the simulation and was better accommodated within the binding pocket as well as within the electron density, suggesting this to be the most likely binding mode. Here the lactone core forms the majority of the interaction with the translocon, with the northern chain also interacting, whereas the long southern chain is predominantly flexible and disordered.

Mycolactone Resistance Mutations Reduce Mycolactone Binding, Most Likely by Modulating Translocon Dynamics

The structure of the mycolactone-bound translocon allowed us to revisit how Sec61α mutations cause resistance to inhibitors. A number of studies have identified mutations that overcome the cytotoxic effects of mycolactone. Indeed, the same mutations often affect multiple different inhibitors, supporting the idea of a common binding mode (Baron et al., 2016; Junne et al., 2015; Mackinnon et al., 2014; Ogbechi et al., 2018; Paatero et al., 2016; Zong et al., 2019). We now conducted an additional forward genetic screen that identified nine substitutions in six codons that gave mycolactone resistance, five of which overlapped with those identified previously (Figure 4A; Figure S7; Tables S1 and S2). Although not identified in our screen, mutation of T86 is also reported to confer resistance to mycolactone (Zong et al., 2019). Unexpectedly, of the residues associated with resistance mutations, only T86 directly contacts mycolactone. Instead, the majority lie on or around the plug helix, on the lumenal side of the translocon, away from the mycolactone-binding pocket (Figure 4B). Indeed, no extra electron density is observed in the mycolactone-bound translocon, which directly contacts residues associated with resistance mutations.

Although we cannot be sure why resistance mutations were not found in the binding pocket, this may be due to the essential role of residues that contact mycolactone. These include hydrophobic residues that play functionally important roles in the structure and dynamics of Sec61α, including roles in the pore ring (V85 and I179) and the hydrophobic patch (L89 and I179). It is therefore possible that mutations in these residues that cause sufficiently large changes in side-chain chemistry to reduce mycolactone binding also make the translocon non-functional, precluding their appearance in a forward genetic screen. Indeed, the naturally occurring mutation V85D, found in hypogammaglobulinemia, is defective in co-translational translocation (Schubert et al., 2018).

With the surprising finding that resistance mutations are not found in residues that contact mycolactone, we used LC-MS and molecular dynamics to assess whether these mutations affect binding of mycolactone to the translocon. To directly measure mycolactone binding, we prepared microsomes from unaltered TRex-293 cells as well as from cells stably overexpressing Sec61α with resistance mutations at positions S71, G80, and S82. RTCs were prepared in the presence of mycolactone, and LC-MS was used to quantify mycolactone in these complexes. In wild-type cells, we estimate that each translocon bound an average of 0.87 ± 0.04 mycolactone molecules. When performed with cells overexpressing mutant Sec61α variants, in each case, the average number of mycolactone molecules per translocon decreased, with reductions of 70%–80% when normalization

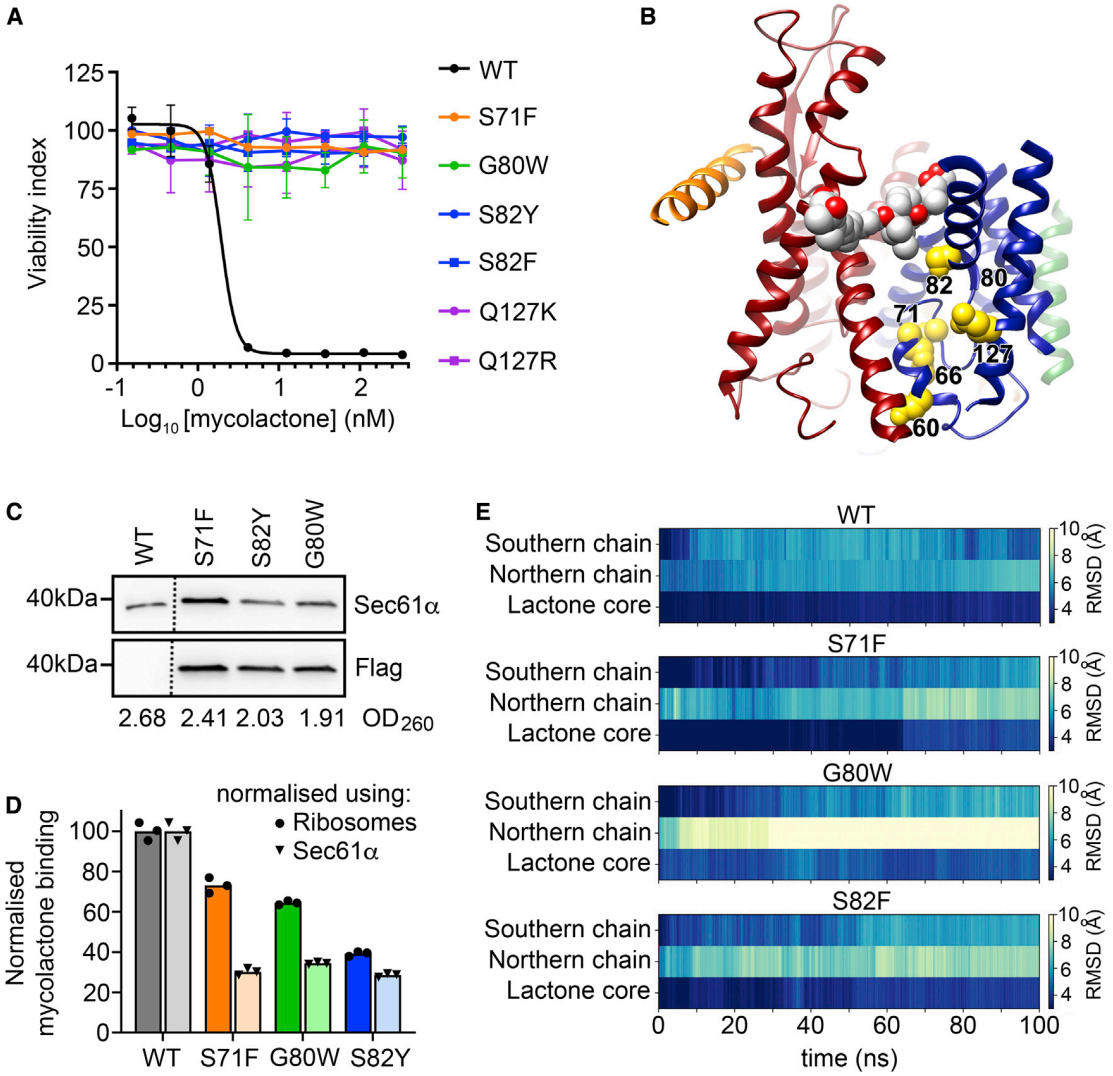


Figure 4. The Location of Mycolactone Resistance Mutations and Their Effect on Mycolactone Binding

(A) Parental HCT-116 cells and representative clones with different amino acid substitutions were tested for their sensitivity to mycolactone A/B. Data are expressed as a normalized viability index of cells treated with inhibitor for 5 days, after which metabolic activity was assessed with Resazurin dye (alamarBlue assay), and values were normalized to a DMSO control. The IC_{50} of wild-type cells for mycolactone A/B was 1.94 nM (1.44 ng/mL). Data are mean ± SEM of n = 3.

(B) A representation of the Sec translocon with Sec61α colored, with helices H1–H5 in blue and H6–H10 in red, Sec61β in green, and Sec61γ in orange. Mycolactone is shown as spheres, with carbon in white and oxygen in red. Residues whose mutation leads to mycolactone resistance are represented as yellow spheres.

(C and D) Microsomes were prepared from TRex-293 cells or those stably transfected with C-terminal FLAG-tagged mutant Sec61A1 constructs. Mycolactone-exposed RTCs were prepared, and the peak fractions were subjected to LC-MS.

(C) Immunoblotting of peak fractions. Migration relative to known molecular weight markers is shown, as is optical density 260 (OD_{260}). The location of an excised lane is indicated by a dotted line.

(D) Relative mycolactone abundance in the RTCs was estimated using high-resolution LC-MS by integrating the extracted ion chromatogram peaks for the $[mycolactone+Na]^+$ (m/z 765.4721–765.5103) ion eluting at 2.38–2.98 relative to a calibration curve. To compare between preparations, the data were standardized on ribosome concentration (calculated from OD_{260}) or Sec61α abundance (pixel intensity of the immunoblot band). Triplicate analysis from n = 1.

(E) Assessment of the dynamics of mycolactone bound to the translocon during 100-ns simulations for wild-type and mutant translocons. Helices H6–H10 were restrained, and movement was allowed for mycolactone and the remainder of the translocon. In each case, the root-mean-square deviation from the starting position is shown at each nanosecond for the southern chain, northern chain, and lactone core. Three independent simulations were conducted, and these plots show the average.

See also Figures S7–S9 and Tables S1 and S2.

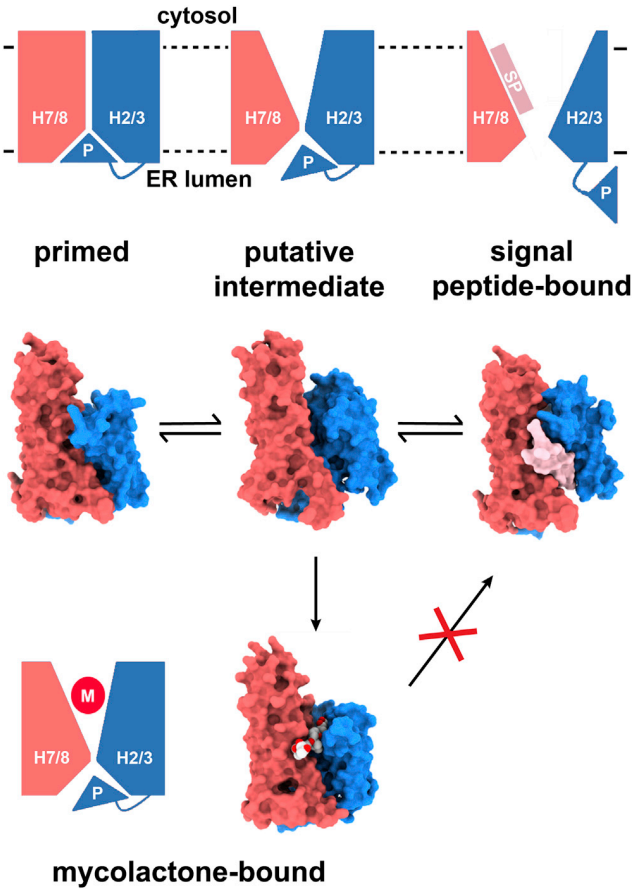


Figure 5. A Model of the Mode of Action of Mycolactone
In this model, the ribosome-bound primed conformation of the Sec translocon is hypothesized to be in equilibrium with a second state (the putative intermediate), in which the lateral gate is open while the plug helix blocks the channel. This is similar in structure to the state stabilized by Sec62/63. Mycolactone can enter this intermediate state and stabilize its conformation while hindering access to the signal peptide binding site.

was conducted using a Sec61α immunoblot (Figures 4C and 4D; Figure S8). Because endogenous Sec61α remains in these preparations, this shows that all three mutations cause major reductions in the binding of mycolactone to the translocon.

In parallel, we performed molecular dynamics to assess the stability of mycolactone in the binding pocket. Long simulations of the entire ribosome-translocon system to observe Sec and mycolactone conformational dynamics were computationally prohibitive. However, short MD simulations can successfully be used to provide an assessment of binding stability (Liu et al., 2017). Because helices H6–H10 of the translocon were equivalent in conformation between mycolactone-bound and -free conformations (Figure 2), we fixed these in position and allowed free movement of mycolactone and helices H1–H5. In three repeats of a 100-ns simulation, mycolactone remained stably bound. However, introduction of the S71F, G80W, and S82F mutations each reduced the stability of mycolactone binding, in particular of the northern chain (Figure 4E; Figure S9). Together with MS analysis, these studies show that resistance mutations decrease mycolactone binding despite being in residues that do not make direct contact with mycolactone.

Therefore, the most probable hypothesis for how resistance mutations function is that they modulate translocon dynamics, reducing the likelihood of formation of the mycolactone binding pocket. To assess this, we measured the change in separation of three pairs of residues that contact mycolactone, V85-W379, L89-V382, and L89-W379, during the molecular dynamics simulations of mycolactone-bound S71F, G80W, and S82F described above (Figure S9). Intriguingly, each mutation changed the separation of these residues in different ways. S82F seems to favor an opening of the lateral gate in the region of the mycolactone binding pocket, perhaps because of introduction of a bulky side chain in a site buried in the mycolactone-bound conformation. In contrast, G80W favored closing of the lateral gate. A conclusive analysis would require extensive simulations of complete RTCs and detailed biochemical analyses of the mutant translocons. However, our analysis suggests that favoring the open or closed state relative to the mycolactone-bound state can be associated with resistance. Indeed, a similar allosteric mechanism has been proposed to explain the role of the *prl* phenotypes that allow export of proteins with defective or absent signal peptides (Junne et al., 2007, 2015; Smith et al., 2005; Trueman et al., 2012). Most of the residues in which we identified resistance mutations are also associated with *prl* phenotypes (Table S1), supporting the idea that their effects are the result of changes in translocon dynamics. Our findings therefore confirm that resistance mutations reduce mycolactone binding without being in residues that directly contact the inhibitor and suggest that this is achieved by changes in conformational dynamics that disfavor the mycolactone-bound conformation.

DISCUSSION

This study reveals the mode of action of Sec61α inhibitors while also highlighting unexpected features of the mechanism of the translocon. First we show that mycolactone wedges open the cytoplasmic side of the lateral gate of Sec61α. This stabilizes a conformation very similar to that stabilized by Sec62/63, which is poised and permissive for post-translational translocation. Therefore, unexpectedly, mycolactone does not stabilize the inactive translocon but traps the helices of Sec61α in a partly activated conformation.

The density attributed to mycolactone lies directly across the cytoplasmic side of the channel. Although the route the signal peptide takes during co-translational translocation is not fully resolved, the location of the engaged signal peptide has been mapped to a site within the lateral gate in translocating ribosome-translocon complexes (Voorhees and Hegde, 2016a). The locations occupied by mycolactone and the signal peptide overlap and will be mutually exclusive in occupation. This suggests that the signal peptide reaches its binding site by passing through the region occupied by mycolactone and that the inhibitor directly prevents signal peptide-mediated opening of the channel and subsequent removal of the plug.

These findings also have consequences for our understanding of the dynamics of the translocon (Figure 5). The binding site for

mycolactone will only be formed when the cytoplasmic side of the lateral gate opens through the tilting of helices H2 and H3 away from H7 and H8. It is hard to envisage mycolactone occupying this binding pocket when Sec61α is in the process of translocation because the translocating translocon will be open, and the nascent polypeptide chain is expected to occlude the mycolactone binding site. Indeed, this is supported by biochemical assays showing that mycolactone does not affect cross-linking of a substrate to Sec61α when added after translocation is initiated by microsome addition (McKenna et al., 2016). Instead, mycolactone will bind while the translocon is idle. This suggests that non-translocating Sec61α is in a dynamic equilibrium, transitioning between the closed state and a transient intermediate that resembles the inhibitor-bound state. This conformational mobility will be advantageous for the translocon because the plug helix remains in place, keeping the channel of the translocon closed to minimize ion leak, but movements in the lateral gate create an opportunity for signal peptide binding and gating. The finding that Sec62/63 stabilizes this intermediate to facilitate post-translational translocation supports this view. However, this intermediate is also susceptible to inhibition, with transient opening of the lateral gate creating the inhibitor binding pocket, allowing mycolactone to wedge open the lateral gate and block access to the signal peptide binding site.

Translocation inhibitors with profoundly different chemical structures show cross-competition and are affected by the same resistance mutations (Van Puyenbroeck and Vermeire, 2018), suggesting that this mechanism is not restricted to mycolactone but common and widely exploitable. Indeed, our findings may also explain why the same resistance mutations affect the efficacy of inhibitors as diverse in structure as cyclic peptides, glycolipids, and polyketide lactones because they modulate formation of their shared binding site. Translocation inhibitors have been suggested as potential anti-cancer and anti-inflammatory agents (Van Puyenbroeck and Vermeire, 2018) and would be expected to have powerful activity against enveloped viruses. This structure not only yields new insight into the mechanism of the translocon but also provides invaluable insight to guide future rational drug design.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAIL
  - Cell lines
- METHOD DETAILS
  - Mycolactone
  - Purification of ribosome-translocon complexes
  - *In vitro* translation/translocation assays
  - Liquid chromatography mass spectrometry
  - SDS-PAGE and immunoblotting
  - Grid preparation and data acquisition
  - Image processing
  - Homology modeling and refinement of atomic models
  - Molecular dynamics simulations
  - Forward genetic screen and viability assays
  - Generation of mycolactone resistant cell lines
  - Immunofluorescence
  - Microsome preparation
  - Figure preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.06.013.

ACKNOWLEDGMENTS

We thank Adam Costin, Susan Lea, Errin Johnson, Justin Deme, and Joseph Caesar for support with data collection and data analysis through the Central Oxford Structural Molecular Imaging Centre (COSMIC) electron microscopy facility. We thank Ramanujan Hegde for valuable comments on the manuscript. We are grateful to Yoshito Kishi (Harvard University) and A.E. Johnson for providing the synthetic mycolactone A/B and rough microsomal membranes (CPMM) used in this study. We thank Melanie Bailey (University of Surrey) for providing access to the Thermo Orbitrap Q-Exactive Plus mass spectrometer, which was partly funded by the Engineering and Physical Sciences Research Council (EP/P001440/1 and EP/R031118/1). We are grateful to the Dobberstein lab, Stephen High (University of Manchester), Nicolas Blanchard (French National Centre for Scientific Research), Wei Shi (Ball State University), and Jack Taunton (University of California, San Francisco) for providing the Sec61β antibody, the PPAF plasmid, mycolactone epimers, ipomoeassin F, and cotransin, respectively. We acknowledge the contribution of the undergraduate students who assisted with sequencing of mycolactone-resistant clones: Jade Bishop, Sarah Cameron, Eve Currall, Tessa Fowler, Marina Kalli, and Bradley Revell. We thank Aloysius Loglo and Louise Hsieh for assistance with cloning and maintaining overexpressing cell lines. This study was conceived at the West African Centre for Cell Biology of Infectious Pathogens (WACCBIP) research conference 2018. M.K.H. and R.E.S. hold Wellcome Trust Investigator Awards in Science (202843/Z/16/Z and 101020/Z/13/Z). S.F.G. holds a Wellcome Trust-funded Studentship in Cellular Structural Biology. This project made use of computation time on the Joint Academic Data Science Endeavour (JADE; EP/P020275/1) via the High-end Computing Biomolecular Simulation Consortium (HECBioSim; http://www.hecbiosim.ac.uk), supported by the Engineering and Physical Sciences Research Council (EPSRC; grant EP/R029407/1). P.C.B. and A.M.Z. thank the Medical Research Council (MRC) for funding (BB/S001247/1).

AUTHOR CONTRIBUTIONS

B.S.H. established protocols for sample production, prepared samples for LC-MS analysis, generated stable cell lines overexpressing mutant Sec61α, and analyzed fractionated samples by SDS-PAGE and western blot. S.F.G. and M.K.H. conducted electron microscopy sample preparation, data acquisition, and image processing. R.E.S. performed the forward genetic screen. B.S.H. and K.A.C. characterized and analyzed the mycolactone resistance mutations and isolated cell-derived microsomes. P.U.M. prepared the canine pancreatic microsomal membranes. C.C. and D.K.W. performed LC-MS. A.M.Z. and P.C.B. planned and designed MD simulations. A.M.Z. performed and analyzed all MD simulations. R.E.S. and M.K.H. devised and supervised the study. S.F.G., R.E.S., and M.K.H. drafted the manuscript. All authors discussed the results, contributed figures and text, and commented on the manuscript.

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

Received: December 6, 2019
Revised: April 30, 2020
Accepted: June 5, 2020
Published: July 20, 2020

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Baron, L., Paatero, A.O., Morel, J.D., Impens, F., Guenin-Macé, L., Saint-Auret, S., Blanchard, N., Dillmann, R., Niang, F., Pellegrini, S., et al. (2016). Mycolactone subverts immunity by selectively blocking the Sec61 translocon. J. Exp. Med. *213*, 2885–2896.

Bayley, C.I., Cieplak, P., Cornell, W., and Kollman, P.A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. *97*, 10269–10280.

Besemer, J., Harant, H., Wang, S., Oberhauser, B., Marquardt, K., Foster, C.A., Schreiner, E.P., de Vries, J.E., Dascher-Nadel, C., and Lindley, I.J. (2005). Selective inhibition of cotranslational translocation of vascular cell adhesion molecule 1. Nature *436*, 290–293.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

Cao, S., Norris, A., Wisse, J.H., Miller, J.S., Evans, R., and Kingston, D.G. (2007). Ipomoeassin F, a new cytotoxic macrocyclic glycoresin from the leaves of Ipomoea squamosa from the Suriname rainforest. Nat. Prod. Res. *21*, 872–876.

Chitwood, P.J., Juszkiewicz, S., Guna, A., Shao, S., and Hegde, R.S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. Cell *175*, 1507–1519.e16.

Collins, P.G., and Gilmore, R. (1991). Ribosome binding to the endoplasmic reticulum: a 180-kD protein identified by crosslinking to membrane-bound ribosomes is not required for ribosome binding activity. J. Cell Biol. *114*, 639–649.

Cross, B.C., McKibbin, C., Callan, A.C., Roboti, P., Piacenti, M., Rabu, C., Wilson, C.M., Whitehead, R., Flitsch, S.L., Pool, M.R., et al. (2009). Eeyarestatin I inhibits Sec61-mediated protein translocation at the endoplasmic reticulum. J. Cell Sci. *122*, 4393–4400.

Demangel, C., and High, S. (2018). Sec61 blockade by mycolactone: A central mechanism in Buruli ulcer disease. Biol. Cell *110*, 237–248.

Elmlund, D., and Elmlund, H. (2012). SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. J. Struct. Biol. *180*, 420–427.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Garrison, J.L., Kunkel, E.J., Hegde, R.S., and Taunton, J. (2005). A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum. Nature *436*, 285–289.

Gehringer, M., Mäder, P., Gersbach, P., Pfeiffer, B., Scherr, N., Dangy, J.P., Pluschke, G., and Altmann, K.H. (2019). Configurationally Stabilized Analogs of *M. ulcerans* Exotoxins Mycolactones A and B Reveal the Importance of Side Chain Geometry for Mycolactone Virulence. Org. Lett. *21*, 5853–5857.

Gemmer, M., and Förster, F. (2020). A clearer picture of the ER translocon complex. J. Cell Sci. *133*, jcs231340.

George, K.M., Chatterjee, D., Gunawardana, G., Welty, D., Hayman, J., Lee, R., and Small, P.L. (1999). Mycolactone: a polyketide toxin from Mycobacterium ulcerans required for virulence. Science *283*, 854–857.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. J. Struct. Biol. *157*, 281–287.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25.

Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C.G., and Economou, A. (2009). Signal peptides are allosteric activators of the protein translocase. Nature *462*, 363–367.

Hall, B.S., Hill, K., McKenna, M., Ogbechi, J., High, S., Willis, A.E., and Simmonds, R.E. (2014). The pathogenic mechanism of the Mycobacterium ulcerans virulence factor, mycolactone, depends on blockade of protein translocation into the ER. PLoS Pathog. *10*, e1004061.

Haßdenteufel, S., Johnson, N., Paton, A.W., Paton, J.C., High, S., and Zimmermann, R. (2018). Chaperone-Mediated Sec61 Channel Gating during ER Import of Small Precursor Proteins Overcomes Sec61 Inhibitor-Reinforced Energy Barrier. Cell Rep. *23*, 1373–1386.

Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J. Chem. Theory Comput. *4*, 116–122.

Hong, H., Gates, P.J., Staunton, J., Stinear, T., Cole, S.T., Leadlay, P.F., and Spencer, J.B. (2003). Identification using LC-MSn of co-metabolites in the biosynthesis of the polyketide toxin mycolactone by a clinical isolate of Mycobacterium ulcerans. Chem. Commun. (Camb.) (22), 2822–2823.

Hoover, W.G. (1985). Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A Gen. Physiol. *31*, 1695–1697.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–38.

Itskanov, S., and Park, E. (2019). Structure of the posttranslational Sec protein-translocation channel complex from yeast. Science *363*, 84–87.

Jämbeck, J.P., and Lyubartsev, A.P. (2012a). Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. J. Phys. Chem. B *116*, 3164–3179.

Jämbeck, J.P., and Lyubartsev, A.P. (2012b). An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. J. Chem. Theory Comput. *8*, 2938–2948.

Jämbeck, J.P., and Lyubartsev, A.P. (2013). Another Piece of the Membrane Puzzle: Extending Slipids Further. J. Chem. Theory Comput. *9*, 774–784.

Jorgensen, W.L., Chandrasekhar, J., and Madura, J.D. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926.

Junne, T., Schwede, T., Goder, V., and Spiess, M. (2007). Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. J. Biol. Chem. *282*, 33201–33209.

Junne, T., Wong, J., Studer, C., Aust, T., Bauer, B.W., Beibel, M., Bhullar, B., Bruccoleri, R., Eichenberger, J., Estoppey, D., et al. (2015). Decatransin, a new natural product inhibiting protein translocation at the Sec61/SecYEG translocon. J. Cell Sci. *128*, 1217–1229.

Kandt, C., Ash, W.L., and Tieleman, D.P. (2007). Setting up and running molecular dynamics simulations of membrane proteins. Methods *41*, 475–488.

Kelkar, A., and Dobberstein, B. (2009). Sec61beta, a subunit of the Sec61 protein translocation channel at the endoplasmic reticulum, is involved in the transport of Gurken to the plasma membrane. BMC Cell Biol. *10*, 11.

Lang, S., Pfeffer, S., Lee, P.H., Cavalié, A., Helms, V., Förster, F., and Zimmermann, R. (2017). An Update on Sec61 Channel Functions, Mechanisms, and Related Diseases. Front. Physiol. *8*, 887.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins *78*, 1950–1958.

Liu, K., Watanabe, E., and Kokubo, H. (2017). Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. J. Comput. Aided Mol. Des. *31*, 201–211.

López, C.A., Unkefer, C.J., Swanson, B.I., Swanson, J.M.J., and Gnanakaran, S. (2018). Membrane perturbing properties of toxin mycolactone from Mycobacterium ulcerans. PLoS Comput. Biol. *14*, e1005972.

MacKinnon, A.L., Garrison, J.L., Hegde, R.S., and Taunton, J. (2007). Photoleucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. J. Am. Chem. Soc. *129*, 14560–14561.

Mackinnon, A.L., Paavilainen, V.O., Sharma, A., Hegde, R.S., and Taunton, J. (2014). An allosteric Sec61 inhibitor traps nascent transmembrane helices at the lateral gate. eLife *3*, e01483.

McKenna, M., Simmonds, R.E., and High, S. (2016). Mechanistic insights into the inhibition of Sec61-dependent co- and post-translational translocation by mycolactone. J. Cell Sci. *129*, 1404–1415.

McKenna, M., Simmonds, R.E., and High, S. (2017). Mycolactone reveals the substrate-driven complexity of Sec61-dependent transmembrane protein biogenesis. J. Cell Sci. *130*, 1307–1320.

Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. *32*, 2319–2327.

Nose, S. (1983a). A molecular dynamics method for simulations in the canonical ensemble. Mol. Phys. *52*, 255–268.

Nose, S. (1988). A unified formulation of the constant temperature molecular dynamics methods. J. Chem. Phys. *81*, 511.

Nose, S., and Klein, M.L. (1983). A study of solid and liquid carbon tetrafluoride using the constant pressure molecular dynamics technique. J. Chem. Phys. *78*, 6928.

Ogbechi, J., Hall, B.S., Sbarrato, T., Taunton, J., Willis, A.E., Wek, R.C., and Simmonds, R.E. (2018). Inhibition of Sec61-dependent translocation by mycolactone uncouples the integrated stress response from ER stress, driving cytotoxicity via translational activation of ATF4. Cell Death Dis. *9*, 397.

Paatero, A.O., Kellosalo, J., Dunyak, B.M., Almaliti, J., Gestwicki, J.E., Gerwick, W.H., Taunton, J., and Paavilainen, V.O. (2016). Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel. Cell Chem. Biol. *23*, 561–566.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. *52*, 7182.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Rapoport, T.A., Li, L., and Park, E. (2017). Structural and Mechanistic Insights into Protein Translocation. Annu. Rev. Cell Dev. Biol. *33*, 369–390.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Saint-Auret, S., Abdelkafi, H., Le Nouen, D., Bisseret, P., and Blanchard, N. (2017a). Synthetic strategies towards mycolactone A/B, an exotoxin secreted by Mycobacterium ulcerans. Org. Chem. Front. *4*, 2380–2386.

Saint-Auret, S., Abdelkafi, H., Le Nouen, D., Guenin-Macé, L., Demangel, C., Bisseret, P., and Blanchard, N. (2017b). Modular total syntheses of mycolactone A/B and its [^{2}H]-isotopologue. Org. Biomol. Chem. *15*, 7518–7522.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Scheres, S.H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nat. Methods *9*, 853–854.

Schubert, D., Klein, M.C., Hassdenteufel, S., Caballero-Oteyza, A., Yang, L., Proietti, M., Bulashevska, A., Kemming, J., Kühn, J., Winzer, S., et al. (2018). Plasma cell deficiency in human subjects with heterozygous mutations in Sec61 translocon alpha 1 subunit (SEC61A1). J. Allergy Clin. Immunol. *141*, 1427–1438.

Smith, M.A., Clemons, W.M., Jr., DeMars, C.J., and Flower, A.M. (2005). Modeling the effects of prl mutations on the Escherichia coli SecY complex. J. Bacteriol. *187*, 6454–6465.

Song, F., Fidanze, S., Benowitz, A.B., and Kishi, Y. (2002). Total synthesis of the mycolactones. Org. Lett. *4*, 647–650.

Swint-Kruse, L., and Brown, C.S. (2005). Resmap: automated representation of macromolecular interfaces as two-dimensional networks. Bioinformatics *21*, 3327–3328.

Tieleman, D.P., Forrest, L.R., Sansom, M.S., and Berendsen, H.J. (1998). Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: molecular dynamics simulations. Biochemistry *37*, 17554–17561.

Trueman, S.F., Mandon, E.C., and Gilmore, R. (2012). A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. J. Cell Biol. *199*, 907–918.

Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36–44.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. (2005). GROMACS: fast, flexible, and free. J. Comput. Chem. *26*, 1701–1718.

Van Puyenbroeck, V., and Vermeire, K. (2018). Inhibitors of protein translocation across membranes of the secretory pathway: novel antimicrobial and anticancer agents. Cell. Mol. Life Sci. *75*, 1541–1558.

Voorhees, R.M., and Hegde, R.S. (2016a). Structure of the Sec61 channel opened by a signal sequence. Science *351*, 88–91.

Voorhees, R.M., and Hegde, R.S. (2016b). Toward a structural understanding of co-translational protein translocation. Curr. Opin. Cell Biol. *41*, 91–99.

Voorhees, R.M., Fernández, I.S., Scheres, S.H., and Hegde, R.S. (2014). Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. Cell *157*, 1632–1643.

Walter, P., and Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. Methods Enzymol. *96*, 84–93.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. J. Comput. Chem. *25*, 1157–1174.

Wu, X., Cabanos, C., and Rapoport, T.A. (2019). Structure of the post-translational protein translocation machinery of the ER membrane. Nature *566*, 136–139.

Yotsu, R.R., Suzuki, K., Simmonds, R.E., Bedimo, R., Ablordey, A., Yeboah-Manu, D., Phillips, R., and Asiedu, K. (2018). Buruli Ulcer: a Review of the Current Knowledge. Curr. Trop. Med. Rep. *5*, 247–256.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, e42166.

Zong, G., Hu, Z., O'Keefe, S., Tranter, D., Iannotti, M.J., Baron, L., Hall, B., Corfield, K., Paatero, A.O., Henderson, M.J., et al. (2019). Ipomoeassin F Binds Sec61α to Inhibit Protein Translocation. J. Am. Chem. Soc. *141*, 8450–8461.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-Sec61α	Santa Cruz Biotechnology	Cat#sc-393182, RRID: AB_2301616
Rabbit polyclonal anti-Sec61β	Kelkar and Dobberstein, 2009	N/A
Rabbit polyclonal anti-RS6	Cell Signaling Technology	Cat#2217, RRID: AB_331355
Mouse monoclonal anti-Ribophorin 2	Santa Cruz Biotechnology	Cat#sc-166421 RRID: AB_2238716
DYKDDDDK Tag Rabbit anti-Tag, Polyclonal	Invitrogen	Cat# PA1984B RRID: AB_347227
ECL Mouse IgG, HRP-linked	GE Healthcare	Cat#NXA931 RRID: AB_772209
ECL Rabbit IgG, HRP-linked	GE Healthcare	Cat#NA934 RRID: AB_772206
Goat Anti-Rabbit IgG (H+L) Antibody, Alexa Fluor 488 Conjugated	Invitrogen	Cat# A-11008 RRID: AB_143165
Chemicals, Peptides, and Recombinant Proteins		
Mycolactone A/B	Song et al., 2002	CAS:222050-77-3
L-[S35]-Methionine	Hartmann Analytic	Cat#SCM01
Digitonin	Sigma-Aldrich	Cat#D141, CAS:11024-24-1
Cotransin	MacKinnon et al. 2007; Mackinnon et al. 2014	CAS: 1000770-96-6
Ipomoeassin F	Cao et al. 2007; Zong et al., 2019	CAS:915392-44-8
Resazurin sodium salt	Sigma Aldrich	Cat#R7017, CAS:62758-13-8
Mycolactone A/B, mixture of epimers at C12	Saint-Auret et al., 2017a, 2017b	N/A
Paraformaldehyde	Sigma Aldrich	Cat#P6148
Deposited Data		
Canine ribosome:Sec61 complex cryo-EM map in the presence of mycolactone	This paper	EMDB: 11064
Canine Sec61 channel bound to mycolactone model coordinates	This paper	PDB: 6Z3T
Canine ribosome:Sec61 complex cryo-EM map in the absence of mycolactone	This paper	EMDB: 11064
Original images and immunoblots, Sanger sequencing, viability assays and LC-MS data	This paper	https://data.mendeley.com/datasets/cc92fyz9sv/1
Experimental Models: Cell Lines		
HCT-116 cell line	ATCC	CCL-247
TRex-293 cell line	Thermo Fisher	R71007
TRex-293 Sec61α S71F C-terminal Flag	This paper	N/A
TRex-293 Sec61α G80W C-terminal Flag	This paper	N/A
TRex-293 Sec61α S82Y C-terminal Flag	This paper	N/A
Oligonucleotides		
Sec61A1_frag1F TAGCACTGACGTGTCTCTCG	Sigma Aldrich	N/A
Sec61A1_frag1R TCCCCATACATCCCGGTCAT	Sigma Aldrich	N/A
Sec61A1_frag2F CTTCAACGGAGCCCAAAAGT	Sigma Aldrich	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sec61A1_frag2R GTGTTGTACTGGCCACGGTAG	Sigma Aldrich	N/A
Sec61A1_frag3F TCATCGCCACCATCTTTGTCTT	Sigma Aldrich	N/A
Sec61A1_frag3R GGACCATGGAGGTCTCTCGG	Sigma Aldrich	N/A
Sec61A1_frag4F TATACATAGTGTTCATGCTGGGCT	Sigma Aldrich	N/A
Sec61A1_frag4R ACACAGTGGAATGAAAGAATACGA	Sigma Aldrich	N/A
Sec61A1_frag5F TAGTGTGCTGCCAGATTCCC	Sigma Aldrich	N/A
Sec61A1_frag5R TCAAATTCCATTCCTCGGCCA	Sigma Aldrich	N/A
Recombinant DNA		
pMXSec61S71F	ThermoFisher Scientific GeneArt	N/A
pMXSec61G80W	ThermoFisher Scientific GeneArt	N/A
pMXSec61S82Y	ThermoFisher Scientific GeneArt	N/A
pNLF1-C (CMV/Hygro)	Promega	Cat#N1361
pNSec61S71F	This paper	N/A
pNSec61G80W	This paper	N/A
pNSec61S82Y	This paper	N/A
Software and Algorithms		
Coot	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix	Adams et al., 2010	https://www.phenix-online.org/
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX	Goddard et al., 2018	https://www.cgl.ucsf.edu/chimerax/
Relion 3.0	Zivanov et al., 2018	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
UCSF MotionCor2	Zheng et al., 2017	https://emcore.ucsf.edu/ucsf-motioncor2
CTFFIND 4.1	Rohou and Grigorieff, 2015	https://grigoriefflab.umassmed.edu/ctffind4
SIMPLE	Elmlund and Elmlund, 2012	https://simplecryoem.com
ResMap	Swint-Kruse and Brown, 2005	http://resmap.sourceforge.net
EPU 2.3	ThermoFisher Scientific	N/A
Adobe Illustrator	Adobe	N/A
Xcalibur 4.2 Quan Browser	Thermo Scientific	N/A
Chromas Lite	Technelysium	
BLASTn	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Prism 8.2.1	GraphPad	N/A
Excel 2016	Microsoft	N/A
Powerpoint 2016	Microsoft	N/A
ImageJ 1.52p	NIH	https://imagej.nih.gov/ij/
GROMACS 2019	Van Der Spoel et al., 2005	http://www.gromacs.org/
VMD 1.9	Humphrey et al., 1996	https://www.ks.uiuc.edu/Research/vmd/
MDAnalysis	Michaud-Agrawal et al., 2011	https://www.mdanalysis.org/
Python		https://www.python.org/
Matplotlib	https://doi.org/10.5281/zenodo.2577644	https://matplotlib.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PyMOL	The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.	https://pymol.org/2/
Other		
Sephacryl S-300 HR	GE Healthcare	Cat#17059910
Micrococcal nuclease	New England Biolabs	Cat#M0247S
Rabbit reticulocyte lysate, nuclease treated	Promega	Cat#L4960
Albumin,bovine serum, protease free	ThermoFisher Scientific	Cat#11461655
Graphene oxide dispersion	Graphenea	Cat#947-768-1
Quantifoil R 1.2/1.3 200 Mesh, copper	Quantifoil	Cat#Q2100CR1.3
4–20% TGX Mini-Protean precast gels	BioRad	Cat#4561903
Spectra broad-range protein markers	Thermo Fisher Scientific	Cat#266234
Simply Blue	Thermo Fisher Scientific	Cat#10432072
Immobilon-P PVDF membrane	Thermo Fisher Scientific	Cat#15750319
Immobilon Western chemiluminescent substrate	Sigma Aldrich	Cat# WBKLS
McCoy's 5A (modified) Medium	ThermoFisher Scientific	Cat#11530646
Foetal Bovine Serum, heat inactivated, SA origin	ThermoFisher Scientific	Cat#11550356
Superscript IV	Life Technologies	Cat# 18090050
GoTaq G2 Green Mastermix	Promega	Cat# M7822
Dulbecco's Modified Eagle's Medium - high glucose	Sigma Aldrich	Cat#D6429
Fugene 6	Promega	Cat#E2691
Protease inhibitor cocktail	Sigma Aldrich	Cat#P8340

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rachel Simmonds rachel.simmonds@surrey.ac.uk

Materials Availability

All unique/stable reagents generated in this study (plasmids containing mutant human Sec61α and mycolactone-resistant cell lines) are available from the Lead Contact without restriction.

Data and Code Availability

The electron microscopy datasets generated during this study, in the form of maps, are available at the electron microscopy data bank (EMD-11064). The model for the Sec translocon with mycolactone bound, generated during this study, is available at the protein data bank PDB: (6Z3T). Other data associated with the manuscript is available at https://data.mendeley.com/datasets/cc92fyz9sv/1.

EXPERIMENTAL MODEL AND SUBJECT DETAIL

Cell lines

The male human colorectal cancer cell line HCT-116 (ATCC CCL-247) was maintained in McCoy's 5A (modified) medium supplemented with 10% fetal bovine serum (FBS) and 1x penicillin-streptomycin (both Thermo Fisher Scientific). TRex-293 cells, derived from the female human embryonic kidney line HEK293, were maintained in high glucose DMEM, supplemented with 10% FBS. Cells were routinely cultured at 37°C and 5% CO2. We do not routinely authenticate our cell lines.

CellPress
OPEN ACCESS

METHOD DETAILS

Mycolactone

We used synthetic mycolactone A/B, kindly provided by Prof Yoshito Kishi (Harvard University) (Song et al., 2002) for all of the experiments described in this manuscript. Mycolactone A/B is a 3:2 rapidly equilibrating mixture of *Z*-$\Delta^{4',5'}$ (A) and *E*-$\Delta^{4',5'}$ (B) geometric isomers at the second double bond in the southern fatty acid tail. Since the two isomers cannot be separated in light, as they rapidly re-equilibrate to the 3:2 ratio, we refer to them as mycolactone. Mycolactone was provided in ethyl-acetate, dried down under nitrogen and resuspended in DMSO at 0.5 mg/ml. Aliquots are protected from light and stored at −80°C.

For selection of stably transfected TRex-293 cells to derive overexpressing mycolactone resistance mutations, we used synthetic mycolactone A/B as a mixture of epimers at C12′, kindly provided by Dr Nicolas Blanchard (Saint-Auret et al., 2017a, 2017b) (French National Centre for Scientific Research), which is available in the greater amounts needed for cells under selection. However, tests performed with stable clones used mycolactone A/B (Song et al., 2002).

Purification of ribosome-translocon complexes

Ribosome-associated Sec61 complexes, known as ribosome-translocon complexes or RTCs, were prepared largely as described (Voorhees et al., 2014). Briefly, rough microsomal membranes (CPMM) were prepared from dog pancreas as described (Walter and Blobel, 1983) and stored at −80°C in 50 mM triethanolamine (pH 7.5), 250 mM sucrose, 1 mM DTT. Aliquots were incubated with 1 mM $CaCl_2$ and 150 U/ml micrococcal nuclease for 7 min at room temperature. The reaction was stopped by addition of 2 mM EGTA and microsomes were flash-frozen in 50 μl aliquots. Mycolactone (1 μg/ml) was prepared by dilution in 0.1% (w/v) BSA from a concentrated stock. Control samples contained 0.1% (v/v) DMSO in BSA. Microsomes were thawed on ice and mycolactone was added to 200 ng/ml (269.2 nM). After gentle mixing, samples were incubated on ice for 30 min. Microsomes were solubilised by incubation for 10 min on ice with an equal volume of 100 mM HEPES pH 7.5, 800 mM KOAc, 20 mM MgOAc, 3.5% (w/v) digitonin, 2 mM DTT. Samples were centrifuged at 20,000 x g for 15min at 4°C. An aliquot of supernatant (5μl) was removed for LC-MS analysis and the rest was applied to a 1 mL Sephacryl-300 column pre-equilibrated with ice cold column buffer (50mM HEPES pH7.5, 200 mM KOAc, 10mM MgOAc, 0.25% (w/v) digitonin, 1 mM DTT). Samples were eluted at 4°C in column buffer and collected in approximately 100 μl fractions. Absorbance at 260 nm was estimated using a Nanodrop (Thermo Fisher) and the initial peak eluting fractions were pooled. Samples were re-centrifuged at 20,000 x g for 15 min at 4°C and processed immediately for electron cryo-microscopy.

In vitro translation/translocation assays

In vitro translation/translocation assays were performed as previously described (Hall et al., 2014). Briefly, mycolactone was added to micronuclease-treated canine pancreatic microsomes and incubated on ice for 20 min in the dark. *In vitro* translation reactions were carried out in a 25 μl volume using nuclease-free rabbit reticulocyte lysates (Promega) according to manufacturer's instructions, with 0.5 μg prepro-α Factor mRNA and 35-S methionine. Microsomal membranes, when present, made up 10% of the final volume of the reaction mix. Samples were incubated for 30 min at 30°C and the reaction was stopped with the addition of an equal volume of 2x Laemmli sample buffer. Proteins were separated by SDS-PAGE on 4%–15% acrylamide gels, fixed, stained, and dried, then visualized with a Typhoon Phosphorimager (GE Healthcare).

Liquid chromatography mass spectrometry

Liquid chromatography mass spectrometry (LC-MS) was performed on a Thermo Scientific Ultimate3000 UHPLC system equipped with a binary solvent manager, column manager, and autosampler, coupled to a Thermo Orbitrap Q-Exactive Plus mass spectrometer operating with the standard electrospray ionization interface and the following conditions: spray voltage 4 kV, capillary temp. 375°C, sheath gas 60 psi, aux gas 20 psi, S lens RF level 100%, resolution 70,000, divert valve schedule (to divert sample buffer components away from MS): 0-2.3 min, waste; 2.3-3.0 min, MS; 3.0-5.0 min, waste. Liquid chromatography conditions were as follows: injection volume 10 μL; column: Phenomenex Luna Omega (C18 porous silica, 2.1 × 50 mm, 5 μm particle size, 100 Å pore size); column temp. 25°C, flow rate 0.5 mL min^{-1}; mobile phase Solvent A: aqueous ammonium acetate (10mM), pH 6.9; Solvent B: acetonitrile; gradient 0-1 min, 5% B; 1-2 min, 5%–90% B; 2-3.5 min, 90% B; 3.5-4.5 min,90%–5% B; 4.5-5 min, 5% B; autosampler temp. 10°C.

The identity and LC-MS retention time of pure mycolactone A/B bound to microsomes was confirmed by both accurate mass and MS/MS of the $[M+Na]^+$ ion which showed the same fragmentation pattern of sodium adducts reported previously (Hong et al., 2003) (core + northern chain *m/z* 429; southern fatty acid *m/z* 359).

For mycolactone quantification, a set of standards for an LC-MS calibration curve were set up using synthetic mycolactone A/B dissolved into acetonitrile/water (1:1). The standards were analyzed by LC-MS once before each repeat of experimental samples. Extracted ion chromatograms (EIC) for mass range 765.4721 – 765.5103 (mycolactone A/B $[M+Na]^+$) were produced using Thermo Scientific Xcalibur 4.2 Quan Browser. Peak areas were first generated automatically for retention time range 2.68 min with window 36 s (2.38 – 2.98 min), ICIS peak integration using smoothing points 7, baseline window 40, area noise factor 5, peak noise factor 10. This was followed by manual adjustment of the integration to include the entire EIC peak.

Experimental samples were: freshly prepared stocks of mycolactone (13.46 μM; 10 μg/ml) in 0.1% (w/v) BSA diluted 1:50 or 1:500 in acetonitrile/water (1:1), total extract from microsomes incubated with or without mycolactone diluted 1:10 in acetonitrile/water

(1:1), and pooled Sephacryl-300 peak fractions containing RTCs (undiluted). Each sample was run three times with blank runs in between. Mycolactone concentrations were determined relative to the calibration curve.

SDS-PAGE and immunoblotting

Aliquots of column eluted fractions were separated by SDS-PAGE on 4%–20% acrylamide gradient gels (Bio-Rad) alongside Spectra broad-range protein markers (Thermo Fisher Scientific). Proteins were stained with Simply Blue (Thermo Fisher Scientific) or blotted onto PVDF membranes. After transfer, blots were fixed with 5% acetic acid and blocked with 5% milk. Primary antibodies used were: mouse monoclonal anti-Sec61α; rabbit polyclonal anti-Sec61β (Kind gift from the Dobberstein lab) (Kelkar and Dobberstein, 2009); rabbit polyclonal anti-ribosomal protein S6 and mouse monoclonal anti-ribophorin 2 (Santa Cruz Biotechnology, sc-166421). Secondary antibodies were ECL Mouse IgG, HRP-linked (GE Healthcare, NXA931) and ECL Rabbit IgG, HRP-linked (GE Healthcare, NA934). Western blots were incubated with Immobilion Chemiluminescent HRP substrate and imaged using a Vilber Fusion FX imaging system. Images presented here were inverted using ImageJ v.1.52p.

Grid preparation and data acquisition

Carbon-coated holey grids (Quantifoil, R1.2/1.3, 200 mesh) were used for the preparation of graphene-oxide grids. The grids were plasma-cleaned using a Quorum device at 50 mA for 60 s. For each grid, 3 μL of 0.2 mg/ml graphene oxide dispersion (Graphenea) was applied on the carbon surface, manually blotted, and allowed to dry overnight.

A 4 μL sample with an absorbance of 10 at 260nm was applied on the surface of each grid and vitrified by plunge-freezing in liquid ethane using a Vitrobot (FEI) with an incubation time of 15 s, a blotting time of 5 s and a blotting force of -15 at 4°C.

Data were collected manually using the acquisition software EPU on a Titan Krios transmission electron microscope (FEI) at a defocus range between 1 and 2.5 μm. All data were recorded on a K2 detector under low dose conditions with a nominal pixel size of 0.822 Å/pixel on the object scale. A total of 5062 and 3833 micrographs were collected for RTCs with and without mycolactone, respectively. Each micrograph experienced a total exposure of $\sim$49 electrons per A^2 fractionated into 32 frames. Foil holes containing the thinnest regions of ice and darkest edges were manually selected as they mostly contained single layer graphene oxide and showed higher signal-to-noise ratio.

Image processing

Original movie frames were motion corrected and combined using MotionCor2 (Zheng et al., 2017). The contrast transfer function parameters were estimated via CTFFIND4.1 (Rohou and Grigorieff, 2015). All micrographs were screened manually for graphene oxide and ice quality. Particles were picked automatically with SIMPLE (Elmlund and Elmlund, 2012). All classifications and refinements were performed using Relion3 (Scheres, 2012; Zivanov et al., 2018). After reference-free 2D classification, particles (76,650 for RTCs with mycolactone, 35,366 for the control dataset) were subjected to an extensive 3D classification into 10 classes after an initial round of refinement.

The main differences between the three-dimensional classes was the absence or presence of eEF2, while most ribosomes contained tRNA in P/E-state or E-state. Two main classes were selected, which included ribosomes with eEF2 (32,456 particles in the presence of mycolactone, 8,607 in the control dataset) and without eEF2 (18,365 particles with mycolactone, 16,245 without). Comparison of these ribosome states with those seen by Voorhees et al. (2014), who used a similar protocol for purification of ribosome-translocon complexes, revealed similarities and differences. Voorhees observed 13% active ribosomes (with A/P and P/E-site tRNAs), 65% without tRNA but with eEF2 and 22% empty (Voorhees et al., 2014). The Voorhees et al. (2014) data share with our findings the presence of a majority of idle ribosomes and the surprising presence of a large fraction containing elongation factor eEF2. The major difference in ribosome conformation when compared with Voorhees is the presence of E or P/E site tRNAs in our RTCs, while Voorhees et al. (2014) observed predominantly tRNA-free ribosomes. We are not sure of the reasons for the differences here, but they might be due to different sources of microsomes (porcine versus canine) or subtle differences in the preparation procedure or time taken for purification.

For each state, 60S-Sec61 masked refinement, followed by particle subtraction of the ribosome core, were carried out. The subtracted particles were used for focused 3D classification with a mask applied around Sec61. The best classes were selected and subjected to 60S-Sec61 masked 3D refinement after particle polishing and CTF refinement. Particles from the two reconstructions obtained were combined, as no differences were observed in the conformation of the translocon, followed by a final 60S-Sec61 masked refinement. In the case of the control dataset, the same pipeline was used, without focused classification. All final reconstructions were subjected to post processing using a wide soft edge mask. This resulted in final resolutions of 2.63Å and 2.85Å, according to the FSC 0.143 criterion following the Relion gold-standard refinement (Scheres and Chen, 2012). Local resolution at the translocon was estimated using ResMap (Swint-Kruse and Brown, 2005). For better visualization and interpretation of the Sec density, and due to the heterogenous resolution across the translocon, the map used for model building was lowpass-filtered to 4Å.

Homology modeling and refinement of atomic models

The structure of Sec61 was built using the structures of the primed state (RCSB code 3J7Q) and that of the signal peptide-engaged state (RCSB code 3JC2) as reference models. The model was built in COOT (Emsley and Cowtan, 2004) and refined using Phenix (Adams et al., 2010) using Ramachandran and secondary structure restraints.

Molecular dynamics simulations

The final model derived from electron cryo-microscopy was used as the starting point for molecular dynamics simulations. The original conformation had mycolactone docked with the southern chain projecting toward the lipid phase while in the reversed conformation, the southern chain projected toward the translocon. The presumptive active form, the *E*-$\Delta^{4'5'}$ isomer, was selected for the simulations. The parameters for mycolactone were taken from (López et al., 2018), who developed them using the General Amber force field (GAFF) (Wang et al., 2004) in combination with the restrained electrostatic potential (RESP) approach (Bayley et al., 1993) employed for optimization of partial charges. The Amber 99SB-ILDN force field (Lindorff-Larsen et al., 2010) was adopted for Sec61 in conjunction with the TIP3P water model (Jorgensen et al., 1983). The POPC lipids were represented with the Slipids force field (Jämbeck and Lyubartsev, 2012a, 2012b, 2013).

The Sec61/mycolactone complex was first oriented with the translocon pore axis aligned coincident with the membrane normal of a pre-equilibrated POPC lipid bilayer containing 512 lipids, ensuring that the transmembrane helices of the Sec61α overlap with the hydrophobic region of the lipid membrane. Heavy atoms were position-restrained with a harmonic potential and a strong force constant of 10^5 kJ mol^{-1} nm^{-2} and the inflateGRO tool (Kandt et al., 2007) was executed to rescale the lipid atom coordinates by a factor of four. A series of 25 iterations of shrinking and energy minimizing was performed until the POPC area per lipid was 71 Å^2, a little above the equilibrium area per lipid of 65.8 Å^2 (Tieleman et al., 1998). The system was subsequently solvated and any water molecules inserted in the lipid phase were deleted. NaCl was added to neutralize the system and to provide a physiological ion concentration of 0.15 M.

Molecular dynamics simulations were performed using the GROMACS software package, version 2019 (Van Der Spoel et al., 2005). Initially, the system was energy minimized with the steepest descent algorithm until the maximum force did not exceed 100 kJ mol^{-1} nm^{-1}. A short equilibration run in the NVT ensemble was carried out for 5 ns and the temperature was stabilized at 310 K with the velocity-rescale thermostat (Bussi et al., 2007) and a coupling constant of 0.1 ps. An additional equilibration step was performed subsequently in the NPT ensemble for 10 ns. In this case, the temperature was maintained at 310 K using the Nosé–Hoover thermostat (Hoover, 1985; Nose, 1983a, 1988) and a time constant of 0.5 ps, while the target pressure of 1 bar was reached using the Parrinello-Rahman barostat (Nose and Klein, 1983; Parrinello and Rahman, 1981) and a semi-isotropic coupling with a time constant of 5 ps. For the MD production runs, the protein backbone atoms were position-restrained by applying a soft harmonic potential with a force constant of 500 kJ mol^{-1} nm^{-2}, in order to ensure that the Sec61 state obtained from electron microscopy would be maintained despite the absence of the ribosome during simulations. Periodic boundary conditions were applied in all three dimensions. The hydrogen-containing bonds were constrained using the LINCS algorithm (Hess, 2008). Neighbor searching, Lennard-Jones interactions and real-space Coulomb interactions were cut off at 10 Å. A dispersion correction was applied to energy and pressure to account for truncation of Lennard-Jones interactions. The particle-mesh Ewald summation was adopted for treatment of reciprocal-space electrostatic interactions with a cubic interpolation and a grid spacing of 1.2 Å. The time step was 2 fs and the leap-frog algorithm was selected for integration of the equations of motion. Coordinates were saved every 10 ps. Three repeats of 100 ns-long MD runs were subsequently performed for each system, amounting to a total simulation time of 600 ns.

To investigate the effect of mutations on mycolactone stability, three mutant systems were generated (S71F, G80W and S82F) using the mutagenesis tool of PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The starting configuration was again the model derived from electron cryo-microscopy, with mycolactone docked in the original conformation. For system setup, equilibration and production runs, the same parameters and procedure as described above were adopted, with the exception that for production runs, position restraints were applied only to the heavy atoms of the Sec61γ subunit and the transmembrane helices H6-H10 of Sec61α to allow the rest of Sec61α to adapt to conformational changes caused by the presence of the mutation. For each mutant system and the WT, three 100 ns-long production runs were performed, resulting in a total simulation time of 1.2 μs.

The VMD software (Humphrey et al., 1996) was used for visualization and the MDAnalysis package (Michaud-Agrawal et al., 2011) for trajectory analysis.

Forward genetic screen and viability assays

The forward genetic screen in DNA damage repair-resistant HCT-116 cells was performed as described (Junne et al., 2015; Ogbechi et al., 2018). The mutagen was ethyl methane sulphonate and selection was with 10 ng/ml (13.5 nM) mycolactone A/B. The cDNA for SEC61A1 was sequenced in resistant clones by RT-PCR. Total RNA (1 μg) was reverse transcribed using Superscript IV reverse transcriptase (Life Technologies) according to the manufacturer's instructions. The coding regions of SEC61A1 were amplified in four overlapping fragments using primers Sec61A1_frag1F/ Sec61A1_frag1R, Sec61A1_frag2F/ Sec61A1_frag2R, Sec61A1_frag3F/ Sec61A1_frag3R, Sec61A1_frag4F/ Sec61A1_frag4R with GoTaq G2 Green Mastermix (Promega). In some cases, confirmation of mutations used an additional fifth fragment using primers Sec61A1_frag5F/ Sec61A1_frag5R. The cycling conditions were 95C for 2 min, 35x cycles of 95C for 30 s, 58C for 30 s and 72C for 1 min, then 5 min at 72C. After clean-up, 70ng of each PCR product along with the corresponding forward primer were sent for Sanger sequencing using the Eurofins Mix2Seq service. Returned sequences, covering the entire coding region were analyzed using Chromas and BLASTn to identify the mutations. No homozygous mutations were identified. The parental cells and all resistant cell lines were heterozygous for a silent C/T transition in Thr445 (ACC/ACT). Almost all clones analyzed showed heterozygous mutations of single alleles. One exception was Hct7 which showed both Q127K and S82Y. Hct30 carried R66K and potentially P140T (as yet unconfirmed).

For viability assays, cells were seeded at 1000 cells/well and then treated with mycolactone, cotransin (Mackinnon et al., 2014) (CT8: kind gift of Jack Taunton), or ipomoeassin F (Cao et al., 2007) (kind gift of Wei Shi) for 5 days. Then resazurin was added to a final concentration of 0.01 mg/ml. After 4 h further incubation, viability was measured as emission at 620 nm at excitation 580 nm using a BMG Fluostar Optima Microplate Reader. Data are expressed as a viability index, representing the relative signal compared to the negative control (DMSO-treated cells).

Generation of mycolactone resistant cell lines

Synthetic C-terminal Flag-tagged Sec61A1 genes containing the mutations S71F, G80W and S85Y were generated (Thermo Fisher GeneArt) and cloned into the pNLF1-C vector (Promega), replacing the nanoluciferase insert. The plasmid was digested with *MluI* and 2 μg transfected into TRex-293 cells (Thermo Fisher) with Fugene (Promega) according to the manufacturer's instructions. Cells were selected for one week in the presence of 10 ng/ml (13.5 nM) mycolactone epimers (kind gift from Dr Nicolas Blanchard, CRNS France) (Saint-Auret et al., 2017a, 2017b) and then for 1 week with 600 μg/ml hygromycin (Thermo Fisher). Expression of the mutant gene was confirmed by immunoblotting, and/or immunofluorescence, for the Flag tag. Single clones were selected for analysis.

Immunofluorescence

Cells were transiently transfected with plasmids containing mutant Sec61 genes, incubated overnight, then seeded onto coverslips. After a further 24 hr incubation, cells were fixed with 4% paraformaldehyde in PBS and blocked with 1% BSA/PBS. Coverslips were incubated with anti-Flag antibody followed by Alex488 conjugated anti-rabbit antibody, then counterstained with DAPI.

Microsome preparation

Crude rough microsomes were prepared from cells by the method described in Chitwood et al. (2018) with minor variations. Briefly, wild-type TRex-293 cells and those clones overexpressing Sec61α mutants were grown to 70%–90% confluency on eight 500 cm^2 plates, collected in ice-cold PBS using cell lifters (Costar), centrifuged at 500 x g for 5 min at 4°C, and washed twice in ice-cold PBS. The cell pellet was resuspended in 3 pellet volumes of ice-cold buffer (10 mM HEPES, pH 7.4, 250 mM sucrose, 2 mM MgOAc). Cells were lysed in the cold (4°C) by 25 passes through a 25-gauge needle using a 10 mL syringe. The lysates were centrifuged at 3,800 x g for 30 min at 4°C. The supernatant was centrifuged at 75,000 x g for 1 h at 4°C in an SW55Ti rotor (Beckman Coulter). The membrane pellet was resuspended in microsome buffer (10 mM HEPES, pH 7.4, 250 mM sucrose, 1 mM DTT). Aliquots were taken at each stage for analysis by immunoblotting and assessment of absorbance at 260 nm. RTCs were isolated from cell derived microsomal preparations as described above for CPMM.

Figure preparation

Figures were prepared using the programs Chimera (Goddard et al., 2007), ChimeraX (Goddard et al., 2018) and GraphPad Prism versions 7.0 or 8.2.1 (GraphPad Software).

QUANTIFICATION AND STATISTICAL ANALYSIS

For mycolactone quantification by LC-MS, the calibration curve was established using the mean of the three repeat runs and the slope and intercept were calculated (Microsoft Excel 2016) for interpolation of mycolactone concentrations in samples. In order to compare TRex-293 cells to those overexpressing Sec61α mutants, the absorbance at 260 nm was used to estimate the concentration of RTCs in the recovered samples, using a μM extinction coefficient of 60.8 (Collins and Gilmore, 1991), for the 80S ribosome. The ratio of mycolactone:RTCs was normalized to wild-type cells. Alternatively, the pixel intensity of the Sec61α band from immunoblots of equal volumes of recovered RTCs (measured using ImageJ v.1.52p) was used to estimate the normalized relative ratio. Here, the individual data from the triplicate measurements from a single run is shown. Since the endogenous Sec61α was not depleted from the cells prior to the analysis, all preparations contain some wild-type Sec61α.

For quantification of cell viability, each biological repeat was performed in triplicate. The data were normalized to the mean of the absorbance at 620 nm for DMSO-treated cells (solvent control). When combining the data from independent biological repeats, the mean of each normalized technical triplicate was taken. The IC_{50} for wild-type cells was calculated using log(inhibitor) versus response – variable slope (four parameter) using Graphpad v. 8.2.1. No slopes were fitted to the resistant clones, as the fit was ambiguous, and no statistical comparison was performed. In the figure legends, n is the number of biological repeats (independent experiments).