

# **New approaches to accelerate structure-based small molecule discovery**

Alex James Flynn

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

School of Molecular and Cellular Biology

September 2023



## Intellectual property and publication statements

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 2 contains work from one jointly authored paper:

Flynn, A.J., Antonyuk, S. V., Eady, R.R., Muench, S.P. and Hasnain, S.S. 2023. A 2.2 Å cryoEM structure of a quinol-dependent NO Reductase shows close similarity to respiratory oxidases. *Nature Communications*. **14**(1), pp.1–12. The candidate set up the data collection and processed the cryo-EM images. S.V. Antonyuk performed the model building. All authors contributed to the analysing of the data and writing of the manuscript.

Chapter 3 contains work from one jointly authored paper:

Debski-Antoniak, O., Flynn, A., Klebl, D.P., Tiede, C., Wilson, I.A., Muench, S.P., Tomlinson, D. and Fontana, J. 2023. Exploiting the Affimer platform against influenza A virus. bioRxiv. 2023.08.22.554342. The candidate was involved with grid preparation and image processing. Debski-Antoniak performed the Affimer screen and subsequent testing. Klebl was involved with grid preparation, data collection and image processing. All authors contributed to the writing of the manuscript.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Alex James Flynn to be identified as Author of this work has been asserted by Alex James Flynn in accordance with the Copyright, Designs and Patents Act 1988.

## Acknowledgements

Thank you to my supervisors Stephen Muench, Juan Fontana and Katie Simmons for their help and guidance during my project. You gave me free rein to make this project my own and to develop my interests on the way. Thank you for helping me to start my journey into science.

There are also many people who have contributed work to this thesis who I must thank. Thank you to Oliver Debski-Antoniak for testing the OA compounds, Nattapon Pinthong for purifying cytochrome bc<sub>1</sub>, Svetlana Antonyuk for model building qNOR and cytochrome bc<sub>1</sub>, Amy Turner for cloning the BAP tag into NP, Alex Derry for performing the <sup>19</sup>F-NMR, and David Klebl and Iso Hirst for help spraying HA.

Thanks to all the Muench and Fontana group members both past and present who have trained me up and helped me get going, in particular, Anna, David, Iso, Ollie and Sammy. Also, a big thank you to everyone in the Level 10 EM office for answering my image processing queries, and for providing some great chat.

I am really grateful to the friends I made during my PhD, including Ben, Fran, Molly, Veronica, Ollie, Laura, Amy and Abid. You made my time in Leeds much more enjoyable. I will miss lunchtime chats, drinking Yorkshire beers and dancing with you all.

Last of all, I want to thank my family for all their support and encouragement. Philippa - thank you sis, I am very lucky to have you. Natalie - my other sis, I am glad our lives finally intersected in Leeds. And to my parents John and Juliet – I am incredibly fortunate to receive all the support you have given me, be that emotional or financial. When you inevitably read this cover to cover, you will finally understand what it is I have actually been doing for four whole years.

## Abstract

The drug discovery pipeline is lengthy and expensive but the early stages can be accelerated through the use of target structures in virtual screening and structure-based drug design. One method for protein structure determination is cryo electron microscopy, which has continued to show advancements in resolution and the speed of data collection, partly thanks to better electron detectors and energy filters.

To better understand how recent developments in hardware and software are pushing the limits of resolution, I have applied cryo-EM to two transmembrane proteins: quinol-dependent nitric oxide reductase (qNOR) from pathogenic bacteria *Alcaligenes xylosoxidans* and bovine cytochrome bc1. Both proteins were previously determined by the Muench group to ~4 Å resolution. However, with current cryo-EM setups, qNOR and cytochrome bc1 were re-imaged and determined to ~2 Å resolution. Structures of two small molecules bound to cytochrome bc1 were also determined to ~2 Å. The more detailed maps contained improved density for protein, waters, lipids and ligands. Furthermore, the cryo-EM data was better quality and collected faster, highlighting the increase in cryo-EM throughput that can be applied to small molecule discovery.

The thesis then focusses on a new route for protein structures to enable small molecule discovery. A novel method for developing small molecule inhibitors was generated whereby the loop structures from Affimers, which are small antibody-like proteins expressed in bacteria, act as starting templates for a ligand-based virtual screening workflow. For proof of concept, we have used two viral systems: Crimean-Congo Haemorrhagic Fever Virus and Influenza A Virus. In both cases, a hit was identified from the mimics using various biophysical and cell-based assays, highlighting how Affimers, commercial libraries and structural approaches can be combined to accelerate small molecule inhibitor discovery.

# Table of Contents

<b>Chapter 1 : Introduction .....</b>	<b>1</b>
1.1 <i>The drug discovery pipeline</i> .....	2
1.2 <i>Protein structure determination for pre-clinical drug discovery</i> .....	4
1.2.1 Applications for protein structures .....	4
1.2.1.1 Identification of ligand binding sites .....	4
1.2.1.2 Virtual screening .....	4
1.2.1.3 Structure-based drug design .....	5
1.2.1.4 Fragment-based drug design.....	6
1.2.2 Protein structure determination methods.....	6
1.2.2.1 X-ray crystallography .....	7
1.2.2.2 Cryo electron microscopy .....	8
1.3 <i>Drug modalities</i> .....	10
1.3.1.1 Small molecules.....	11
1.3.1.2 Biopharmaceuticals .....	11
1.3.1.3 Biopharmaceutical-guided small molecule development.....	13
1.4 <i>Objectives</i> .....	13
<b>Chapter 2 : Investigating the advances in cryo-EM structure determination .....</b>	<b>15</b>
2.1 <i>Introduction</i> .....	16
2.1.1 Improvements in cryo-EM structure determination .....	16
2.1.2 Quinol-dependent nitric oxide reductase .....	18
2.1.2.1 Ax qNOR as a drug target .....	18
2.1.2.2 Structural analysis of Ax qNOR .....	19
2.1.3 Cytochrome bc1.....	20
2.1.3.1 Cytochrome bc1 as a parasitic target .....	21
2.1.3.2 Cytochrome bc1 structural analysis.....	21
2.1.4 Objectives .....	22
2.2 <i>Materials and methods</i> .....	22
2.2.1 Overview of single particle cryo-EM analysis .....	22
2.2.1.1 Grid preparation.....	22
2.2.1.2 Cryo electron microscope setup .....	23
2.2.1.3 Data collection .....	24
2.2.1.4 Image processing .....	24
2.2.1.5 Model building.....	25
2.2.2 Structure determination of Ax qNOR by cryoEM .....	26
2.2.2.1 Protein purification and cryo-EM grid preparation .....	26

2.2.2.2 Data collection .....	26
2.2.2.3 Image processing .....	26
2.2.2.4 Model building.....	27
2.2.3 Bovine cytochrome bc1 cryo-EM structure determination .....	28
2.2.3.1 Protein purification .....	28
2.2.3.2 Grid preparation.....	28
2.2.3.3 Data collection .....	28
2.2.3.4 Image processing .....	28
2.2.3.5 Model building.....	29
2.2.4 Determining ligand-bound Bovine cytochrome bc1 structures by cryo-EM .....	29
2.2.4.1 Grid preparation.....	29
2.2.4.2 Data collection .....	29
2.2.4.3 SG114 image processing .....	30
2.2.4.4 CK267 image processing.....	30
2.3 Results.....	31
2.3.1 Ax qNOR was determined to 2.2 Å resolution using cryo-EM .....	31
2.3.1.1 Images collected on a Falcon 4 detector were processed to a 2.2 Å map .....	31
2.3.1.2 The positions of side chains and heme groups were less ambiguous in the higher resolution map .....	33
2.3.1.3 First structural evidence for the binding site of the ubiquinol electron donor by cryo-EM .....	33
2.3.1.4 Lipid molecules were present in the cryo-EM map .....	34
2.3.1.5 Only three hours of Ax qNOR cryo-EM data was required to generate a 3.0 Å 3D reconstruction .....	35
2.3.2 The second protein complex bovine cytochrome bc1 was determined to 2.1 Å resolution in an apo state by cryo-EM .....	37
2.3.2.1 Collecting on an upgraded detector improved the resolution by >2Å .....	37
2.3.2.2 New features in the cryo-EM density included protein, waters and lipids.....	39
2.3.2.3 Only 50 minutes of bovine bc1 cryo-EM data was required to generate a 3.0 Å reconstruction .....	40
2.3.3 Two small molecule-bound bovine cytochrome bc1 structures were determined to 2.0 Å resolution.....	42
2.3.3.1 A cryo-EM structure identified CK267 in both the Q <sub>i</sub> and Q <sub>o</sub> sites .....	42
2.3.3.2 A cryo-EM structure identified CK267 in the Q <sub>i</sub> site .....	44
2.4 Discussion .....	46
<b>Chapter 3 : Generating anti-viral small molecules using <i>in silico</i> approaches .....</b>	<b>49</b>
3.1 Introduction.....	50
3.1.1 Crimean Congo Hemorrhagic Fever Virus.....	50

3.1.1.1 CCHF Disease .....	50
3.1.1.2 Virus life cycle.....	51
3.1.1.3 Treatments.....	52
3.1.2 An Affimer inhibitor of CCHFV NP .....	53
3.1.2.1 Affimer development.....	53
3.1.2.2 Uses in drug discovery .....	54
3.1.3 Influenza A virus .....	54
3.1.3.1 Influenza disease.....	54
3.1.3.2 IAV lifecycle .....	54
3.1.3.3 Influenza treatments .....	55
3.1.4 Affimer inhibitors of hemagglutinin.....	55
3.1.4.1 Development.....	56
3.1.4.2 Use in drug discovery .....	56
3.1.5 Objectives .....	56
3.2 <i>Methods</i> .....	57
3.2.1 Overview of key methods .....	57
3.2.1.1 Protein preparation .....	57
3.2.1.2 Binding site prediction.....	58
3.2.1.3 Library preparation.....	58
3.2.1.4 Ligand-receptor docking .....	58
3.2.1.5 Ligand-based virtual screening.....	59
3.2.1.6 Visual inspection of compounds .....	59
3.2.2 NP-Affimer mimic development .....	60
3.2.2.1 Analysis of NP:NP-Affimer crystal structure .....	60
3.2.2.2 Ligand-based virtual screening.....	60
3.2.2.3 Investigating the binding modes of the AF1 series using docking programs.....	61
3.2.2.4 Prediction of small molecule binding sites .....	61
3.2.3 Structure-based virtual screening to identify modulators of the CCHFV NP RNA site .....	61
3.2.4 Determining Influenza A virus HA-Affimer structures by cryo-EM .....	61
3.2.4.1 Cryo grid preparation .....	62
3.2.4.2 Cryo-EM data collection.....	62
3.2.4.3 Image processing .....	62
3.2.4.4 HA-A31 model building .....	63
3.2.5 Developing Affimer A5 mimics.....	63
3.2.5.1 Affimer sequence alignment .....	63
3.2.5.2 IAV HA Affimer homology modelling.....	64
3.2.5.3 Ligand-based virtual screen using 3D shape similarity.....	64
3.3 <i>Results</i> .....	65

3.3.1 Small molecule mimics of NP-Affimer were developed using a NP:NP-Affimer crystal structure.....	65
3.3.1.1 NP-Affimer binds to CCHFV NP through loop 2 residues 106-110 .....	65
3.3.1.2 Small molecule mimics of the NP-Affimer were identified using shape and electrostatic similarity programs .....	67
3.3.1.3 <i>In silico</i> predictions of Affimer mimic binding identified the RNA-binding tunnel as a potentially druggable site on CCHFV NP.....	69
3.3.2 A traditional protein structure-guided virtual screening workflow identified potential small molecules binders of CCHFV NP .....	74
3.3.3 Mimics of the HA-Affimer were developed without prior structural information .....	75
3.3.3.1 Initial attempts identified issues with standard cryo-EM analysis of HA .....	76
3.3.3.2 Rapid cryo-EM grid preparation generated the first A31-bound HA structure at 3.4 Å resolution.....	78
3.3.3.3 Time-resolved cryo-EM generated the first A5-bound HA structure at 4.4 Å resolution .....	79
3.3.3.4 Software to predict shape and electrostatic similarity identified small molecule mimics of HA-Affimer .....	81
3.4 Discussion .....	84
<b>Chapter 4 : Characterising small molecule Affimer mimics in vitro .....</b>	<b>88</b>
4.1 Introduction .....	89
4.2 Materials and methods .....	89
4.2.1 Screening assays background.....	89
4.2.1.1 Cytotoxicity assays .....	89
4.2.1.2 TCID <sub>50</sub> assays.....	90
4.2.1.3 Mini-genome system assay .....	90
4.2.1.4 Surface Plasmon Resonance spectroscopy .....	91
4.2.1.5 Thermal shift assays by differential scanning fluorimetry .....	92
4.2.1.6 Thermal shift assays by Nano Differential Scanning Fluorimetry .....	92
4.2.1.7 <sup>19</sup> F-NMR .....	92
4.2.2 Materials .....	93
4.2.2.1 Plasmids .....	93
4.2.3 Cytotoxicity assay .....	93
4.2.3.1 Screening OA compounds.....	94
4.2.3.2 Screening AF compounds .....	94
4.2.4 CCHFV mini-genome system assay.....	94
4.2.5 IAV TCID <sub>50</sub> assay.....	95
4.2.6 CCHFV NP-BAP protein production .....	95
4.2.6.1 Cloning.....	96

4.2.6.2 NP-BAP protein expression .....	96
4.2.6.3 Protein purification buffers .....	96
4.2.6.4 Nickel affinity chromatography.....	97
4.2.6.5 Size exclusion chromatography.....	98
4.2.7 CCHFV NP protein production.....	98
4.2.7.1 Protein expression .....	98
4.2.7.2 Protein purification buffers .....	98
4.2.7.3 Nickel affinity chromatography.....	99
4.2.7.4 Size exclusion chromatography.....	99
4.2.8 NP-Affimer protein production .....	100
4.2.8.1 Protein expression .....	100
4.2.8.2 Protein purification buffers .....	100
4.2.8.3 First nickel affinity chromatography .....	101
4.2.8.4 Second nickel affinity chromatography .....	101
4.2.8.5 Third nickel affinity chromatography .....	101
4.2.9 Surface Plasmon Resonance spectroscopy .....	102
4.2.9.1 IAV HA screening .....	102
4.2.9.2 CCHFV NP screening.....	102
4.2.10 Thermal shift assays by differential scanning fluorimetry .....	102
4.2.11 Thermal shift assays by nanoDSF.....	103
4.2.12 X-ray crystallography .....	103
4.2.13 <sup>19</sup> F CPMG NMR .....	103
4.2.14 Cryo-EM analysis of HA and OA3 .....	104
4.2.14.1 Grid preparation.....	104
4.2.14.2 Data collection .....	104
4.2.14.3 Image processing .....	104
4.2.14.4 Model refinement.....	105
4.2.15 AF1-03 Analogue development .....	105
4.2.16 Figure making .....	105
4.3 Results.....	105
4.3.1 NP-Affimer mimics AF1-01 and AF1-03 bind to and inhibit CCHFV NP.....	106
4.3.1.1 AF1-01, AF1-03 and AF2-05 inhibit CCHFV replication in a mini-genome system assay .....	106
4.3.1.2 A new BAP-tagged NP construct was produced to facilitate immobilisation in SPR experiments.....	108
4.3.1.3 NP-Affimer was produced for use as a positive control in binding assays .....	110
4.3.1.4 Surface Plasmon Resonance failed to detect active compounds binding to CCHFV NP .....	111
4.3.1.5 Wild-type CCHFV NP was produced for screening assays .....	112

4.3.1.6 DSF-based thermal shift assays identified several AF2 compounds as binders to CCHFV NP .....	113
4.3.1.7 Dye-free nanoDSF-based thermal shift assays detected AF1-03 binding to CCHFV NP .....	114
4.3.1.8 X-ray crystallography did not identify the small-molecule binding sites.....	119
4.3.1.9 Summary of screening assays.....	121
4.3.2 Analogues of NP-Affimer mimic AF1-03 were inactive against CCHFV NP .....	122
4.3.2.1 Structural optimisation of AF1-03 .....	122
4.3.2.2 AF1-03 analogues were inactive against CCHFV NP .....	125
4.3.3 Affimer-derived small molecule OA3 inhibits influenza virus and binds to HA.....	128
4.3.3.1 OA3 inhibits live influenza virus in a TCID50 assay .....	128
4.3.3.2 Surface plasmon resonance spectroscopy did not identify OA3 binding to HA	129
4.3.3.3 DSF-based thermal shift assays did not identify compounds binding to HA ....	130
4.3.3.4 NanoDSF thermal shift assays also did not identify compounds binding to HA	132
4.3.3.5 <sup>19</sup> Fluorine NMR experiments identified OA3 binding to HA.....	133
4.3.3.6 Cryo-EM of HA with OA3 using a device for fast grid preparation generated an apo HA structure .....	134
<i>4.4 Discussion .....</i>	<i>135</i>
4.4.1 Recommended screening cascade .....	138
<b>Chapter 5 : Discussion.....</b>	<b>141</b>
<i>5.1 Discussion .....</i>	<i>142</i>
<i>5.2 Final thoughts .....</i>	<i>147</i>

## List of Figures

Figure 1.1 The drug discovery pipeline for small molecule therapeutics.....	2
Figure 1.2 Schematic for protein structure determination using X-ray crystallography and cryo electron microscopy. ....	7
Figure 1.3 The rise of cryo electron microscopy for protein structure determination. ....	10
Figure 1.4 Examples of small molecule and biopharmaceutical drug modalities.....	11
Figure 2.1 Ax qNOR structures determined by the Muench and Hasnain groups by cryo-EM. ....	19
Figure 2.2 Bovine cytochrome bc1 and its small molecule binding sites.....	21
Figure 2.3 A cryo-EM map of the transmembrane quinol-dependent nitric oxide reductase (qNOR) from the pathogenic bacterial species <i>Alcaligenes xylosoxidans</i> (Ax) was determined to 2.2 Å resolution. ....	32
Figure 2.4 A higher resolution cryo-EM map of Ax qNOR was generated using an upgraded cryo-EM set up.....	33
Figure 2.5 Electron donor ubiquinol was present in the cryoEM map of Ax qNOR. ....	34
Figure 2.6 Lipids and detergents within the Ax qNOR cryo-EM map. ....	35
Figure 2.7 A ResLog plot to assess the quality of Ax qNOR cryo-EM data collected on the Falcon 4 detector. ....	37
Figure 2.8 A cryo-EM reconstruction of apo bovine cytochrome bc1 to 2.1 Å resolution.	38
Figure 2.9 New features visible in the bovine cytochrome bc1 apo cryo-EM map. ....	40
Figure 2.10 A ResLog plot to measure the assess the quality of cryo-EM data collected of bovine cytochrome bc1 in the apo state. ....	42
Figure 2.11 Determining a structure of bovine bc1 bound to small molecule CK-267 by cryo-EM to 2.0 Å resolution. ....	43
Figure 2.12 The small molecule CK-267 bound to bovine cytochrome bc1 in a cryo-EM map. ....	44
Figure 2.13 Generating a structure of bovine cytochrome bc1 bound to the small molecule SG114. ....	45

Figure 2.14 SG114 bound to bovine cytochrome bc1 map in a cryo-EM map.....	46
Figure 3.1: Crimean-Congo haemorrhagic fever virus (CCHFV) is a tick-borne virus that causes human symptoms of fever, organ failure and haemorrhage.....	51
Figure 3.2 A schematic of a Crimean-Congo haemorrhagic fever virus virion structure..	52
Figure 3.3 Ribavirin and favipiravir are two small molecules used for the treatment of CCHFV. Both become incorporated into viral RNA causing lethal mutations.....	53
Figure 3.4 Example of influenza a virus structural proteins and small molecule drugs that target them. ....	55
Figure 3.5 The structure-based and ligand-based virtual screening pipelines and example programs used for this chapter.....	57
Figure 3.6 The X-ray crystal structure of Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) bound to an inhibitory Affimer (NP-Affimer). Residues 106-110 of the second Affimer loop is the core site of NP binding. ....	67
Figure 3.7 A ligand-based virtual screening workflow identified small molecules that mimic the nucleoprotein (NP) -binding region in an inhibitory Affimer (NP-Affimer). ....	69
Figure 3.8 The AF1 set of small molecules generated using a novel Affimer-guided approach.....	69
Figure 3.9 Docking experiments identified the RNA binding tunnel in Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) as a potentially drugable site.....	71
Figure 3.10 Docked poses compared to ROCS-aligned poses for example compounds from the AF1 series of small molecule NP-Affimer mimics.....	72
Figure 3.11 SiteMap (Schrödinger) analysis of Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) for potential ligand-binding sites. ....	73
Figure 3.12 A virtual screening workflow identified 15 potential binders of the Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) RNA-binding site.	75
Figure 3.13 Fifteen AF2 compounds selected using a traditional structure-based virtual screening workflow against the CCHFV NP RNA-binding site.....	75

Figure 3.14 Issues of hemagglutinin (HA) preferred orientation and on-grid Affimer A5-induced aggregation were solved by rapid mixing and rapid cryo grid preparation. ....	78
Figure 3.15 Cryo-EM map of inhibitory Affimer A31 bound to hemagglutinin (HA) generated by using rapid grid preparation. ....	79
Figure 3.16 Structure of inhibitory Affimer A5 bound to hemagglutinin (HA) generated by cryo-EM using rapid grid preparation.....	80
Figure 3.17 Development of small molecule mimics of Affimers which bind and inhibit influenza A virus hemagglutinin (HA). ....	82
Figure 3.18 Example small molecule mimics of Affimers A5 and A31 isolated against the hemagglutinin (HA) spike protein from influenza A virus. ....	83
Figure 3.19 Eight small molecules identified by ligand-based virtual screen that mimic an Affimer isolated against hemagglutinin (HA).....	84
Figure 3.20 Workflow for generating small molecule mimics of Affimer proteins depending on whether there is a target:Affimer structure to guide mimic development.....	87
Figure 4.1 A schematic for the CCHFV mini-genome system. ....	91
Figure 4.2 Workflow for small molecule cytotoxicity assay .....	94
Figure 4.3 Workflow for using the CCHFV mini-genome system assay to screen compounds .....	94
Figure 4.4 Workflow for generating BAP-tagged CCHFV NP .....	96
Figure 4.5 Workflow for producing wild-type CCHFV NP .....	98
Figure 4.6 Workflow for generating NP-Affimer.....	100
Figure 4.7 Screening the AF1 and AF2 sets of small molecules for inhibition of the CCHFV mini-genome system (MGS) identified three active compounds.....	108
Figure 4.8 Introduction of a BAP-tag into CCHFV NP construct to generate biotinylated NP.....	109
Figure 4.9 Purification of a BAP-tagged CCHFV NP from <i>E. coli</i> cells.....	110

Figure 4.10 Purification of the NP-Affimer which inhibits CCHFV NP for use as a control ligand in screening assays. ....	111
Figure 4.11 Surface plasmon resonance spectroscopy identified some AF2 compounds, which were discovered through traditional structure-based virtual screening, binding to CCHFV NP. ....	112
Figure 4.12 Purification of CCHFV NP for biophysical assays and structural studies....	113
Figure 4.13 DSF-based thermal shift assays identified binders of CCHFV NP from the AF1 and AF2 sets of small molecules. ....	114
Figure 4.14 Optimisation steps for NanoDSF thermal shift assays prior to screening the AF1 and AF2 compounds for binding to CCHFV NP. ....	116
Figure 4.15 Screening the AF1 and AF2 compound series at 200 $\mu$ M using thermal shift assays by nanoDSF. ....	117
Figure 4.16 Using nanoDSF to calculate KDs of AF1-03 and AF2-10. ....	119
Figure 4.17 X-ray crystallography of CCHFV NP crystallised in the presence of AF1-03 generated an apo structure of NP. ....	120
Figure 4.18 Analogues of Affimer-based compound AF1-03 were developed using virtual screens. ....	124
Figure 4.19 Analogues of the active small molecule AF1-03 were purchased and termed the AF3 compounds. ....	125
Figure 4.20 Screening analogues of AF1-03 in cell-based and biophysical assays showed they were inactive. ....	127
Figure 4.21 The Affimer-based OA compounds were screened for inhibition of influenza A virus in a TCID <sub>50</sub> assay. ....	129
Figure 4.22 Surface plasmon resonance (SPR) spectroscopy could not identify biologically active compound OA3 binding to HA due to its non-specific sticking to the chip. ....	130
Figure 4.23 Dye-based differential scanning fluorimetry (DSF) thermal shift assays did not identify any Affimer-based OA compounds binding to hemagglutinin (HA)....	132
Figure 4.24 NanoDSF thermal shift assays screened the remaining OA compounds but did not identify any binding to hemagglutinin (HA). ....	133

<b>Figure 4.25 19-Fluorine NMR experiments showing Affimer-based compound OA3 binding to hemagglutinin (HA). .....</b>	<b>134</b>
<b>Figure 4.26 Using a sprayer for rapid cryo-EM grid preparation to generate a map of HA with OA3 bound. The resulting map was ambiguous for the position of the ligand. ....</b>	<b>135</b>
<b>Figure 4.27 Template workflow for screening compounds developed using an Affimer-guided virtual screening workflow.....</b>	<b>140</b>

## List of Tables

Table 2.1 Differences in Ax qNOR cryo-EM data collections to evaluate effects from increased data quality and quantity.....	36
Table 2.2 Comparison of data collections generating apo bc1 data .....	41
Table 3.1 Visualisation programs.....	59
Table 3.2 Affimers used for structural alignment.....	63
Table 4.1 Plasmids used in this chapter.....	93
Table 4.2 Purified proteins used in this chapter.....	93
Table 4.3 Information on X-ray crystallography data collection and model refinement for CCHFV NP + AF1-03 .....	120
Table 4.4 Summary of AF1 and AF2 compound screening assays. ....	122
Table 5.1 Information on the cryo-EM data collection and model building for a structure of Ax qNOR.....	149
Table 5.2 Information on cryo-EM data collection for a structure of bovine cytochrome bc1.....	150
Table 5.3 Information on cryo-EM data collection for a structure of bovine cytochrome bc1 with small molecule SG-114 .....	152
Table 5.4 Information on cryo-EM data collection for a structure of bovine cytochrome bc1 with small molecule ligand CK-267.....	153

## Abbreviations

ADME	Absorption, distribution, metabolism and excretion
ASA	Accessible surface area
ATP	Adenosine Triphosphate
<i>Ax</i>	<i>Alcaligenes Xylooxidans</i>
BAP	Biotin acceptor peptide
BEI	Binding efficiency index
BSA	Buried surface area
BSL	Biosafety level
CCHF	Crimean Congo Hemorrhagic Fever
CCHFV	Crimean Congo Hemorrhagic Fever Virus
CPE	Cytopathic effects
Cryo-EM	Cryo electron microscopy
Cryo-ET	Cryo electron tomography
CF	Cystic Fibrosis
CTF	Contrast Transfer Function
Cyt bc1	Cytochrome bc1
DDM	n-dodecyl- $\beta$ -D-maltoside
DQE	Detective quantum efficiency
DSF	Differential scanning fluorimetry
DTM	n-Decyl- $\beta$ -D-Thiomaltoside
ELISA	Enzyme-linked immunosorbent assay
EM	Expectation-maximisation
EMDB	Electron microscopy databank
FBDD	Fragment based drug discovery
FDA	US Food and Drug Administration
FSC	Fourier Shell Correlation
GFP	Green fluorescent protein
GPCR	G-protein coupled receptor

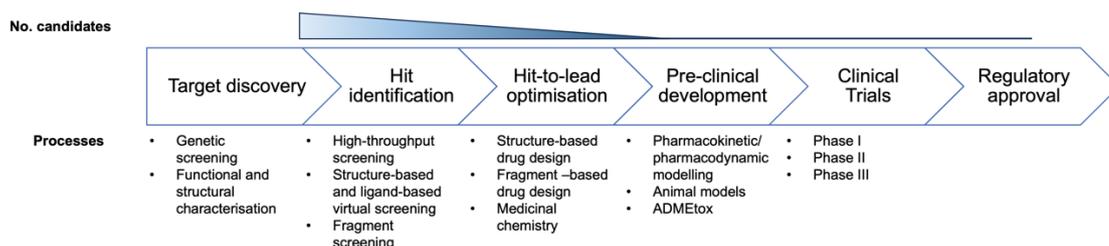
GUI	Graphical user interface
HA	Hemagglutinin
HIV-1	Human immunodeficiency virus 1
HTS	High throughput screening
HTVS	High throughput virtual screening
IAV	Influenza A virus
LBVS	Ligand-based virtual screening
LOP	Lauryl Oleyl Phosphatidyl-ethanolamine
mAb	Monoclonal antibody
MGS	Mini-genome system
NMR	Nuclear magnetic resonance
NP	Nucleoprotein
OG	Octyl glucoside
ORF	Open reading frame
PAINS	Pan-assay interference compounds
<i>Pf</i>	<i>Plasmodium falciparum</i>
PPI	Protein-protein interactions
PROTACs	Proteolysis targeting chimera
Q	Ubiquinone
QH2	Ubiquinol
qNOR	Quinol-dependent nitric oxide reductase
RBD	Receptor binding domain
RMSD	Root mean squared deviation
RNP	Ribonucleoprotein
RUs	Response units
SA	Streptavidin
SAR	Structure activity relationship
SBDD	Structure-based drug design
SBVS	Structure-based virtual screening
SEC	Size exclusion chromatography

SNR	Signal-to-noise ratio
SP	Standard precision
SPR	Surface plasmon resonance
STA	Sub-tomogram averaging
TCID50	The 50% tissue culture infectious dose
T <sub>m</sub>	Melting temperature
TSA	Thermal shift assays
VDAM	Variable-metric Gradient Descent with Adaptive Moments
VS	Virtual screening
WHO	World health organization
XP	Extra precision
XRC	X-ray crystallography

# Chapter 1 : Introduction

## 1.1 The drug discovery pipeline

The discovery of small molecule drugs for treating human disease is a cornerstone of modern medicine, with small molecules being used to treat a range of diseases including cancers, neurological diseases and infectious diseases. The constant, rapid development of new drugs is key to fighting emerging diseases and pathogens including during the recent Covid-19 pandemic <sup>1</sup>. However, the drug discovery process is by no means easy; it typically takes 10-15 years from patent filing to approval by the U.S. Food and Drug administration (FDA) and it requires a vast range of scientific disciplines <sup>2</sup>. Because of this, the financial cost per approved drug is estimated to be between \$161 million and \$4.54 billion, depending on how the cost is calculated, and has been steadily increasing over time <sup>3</sup>. The cost of drug development must outweigh the predicted profits for it to be deemed commercially viable. Despite the rising costs, the number of new chemical entities has been increasing since 2000 <sup>4</sup>. The process required to translate a drug from the bench into patients is termed the drug discovery pipeline and covers the stages from target identification through to lead compound optimisation, clinical trials and regulatory approval (Figure 1.1).



**Figure 1.1 The drug discovery pipeline for small molecule therapeutics.**

In target discovery, when a human disease is identified, a biological macromolecule is selected whose modulation should lead to a modulation of disease status <sup>5</sup>. Potential targets can be identified through large-scale genetic screening of patients (e.g. for cancers) and/or by functional studies (e.g. for pathogens). The range that can be modulated is being expanded; for example, targets have traditionally been proteins but now, there are therapies designed to target DNA and RNA <sup>6</sup>, and historically undruggable targets such as non-enzymes and proteins without deep binding sites are now targeted via new therapeutic modalities <sup>7</sup>. This broadening of potential targets is highlighted by the fact that 54% of novel FDA approvals in 2022 were first-in-class and another 54% were for rare diseases <sup>8</sup>.

In hit identification, small molecule ‘hits’ which bind and modulate a given target are detected. Traditionally, this involved physically screening a library of thousands or millions of compounds in high-throughput screening (HTS) assays <sup>9</sup>. Screening assays

can measure target binding or inhibition, either in an isolated or cellular context, or more general phenotypic changes. Nowadays, advances in laboratory automation and robotic liquid handling allow for >100,000 compounds to be screened per day, provided the assay workflow is well-established<sup>10</sup>. Examples of HTS-derived compounds include maraviroc<sup>11</sup> with potent anti-human immunodeficiency virus type 1 (HIV-1) activity, and eltrombopag<sup>12,13</sup> for the treatment of immune thrombocytopenia. Furthermore, advances in computer modelling have allowed for chemical libraries to be screened for target binding *in silico* using structure-based virtual screening (SBVS) programs prior to the *in vitro* assaying of a subset of predicted binders. SBVS was used in the development of many compounds including captopril for the treatment of hypertension; saquinavir, ritonavir and indinavir for the treatment of human immunodeficiency virus; and boceprevir for the treatment of hepatitis C virus<sup>14</sup>.

Hits are then converted into drug-like lead molecules during hit-to-lead optimisation. Iterative cycles of drug design and testing are employed to identify more active, and selective lead molecules. This iterative approach can improve the potency of hits by orders of magnitude, as was the case in the development of apixaban, which had a 480,000-fold improved inhibitory constant and 30,000-fold selectivity over other human coagulation proteases compared to the initial hit molecule<sup>15</sup>. The testing is more varied in this optimisation stage, including early ADME (absorption, distribution, metabolism and excretion) profiling and using more complex *in vivo* systems like organoid or animal models. Furthermore, the compounds are optimised to improve their pharmacokinetic properties and safety profiles prior to them being tested in humans.

Clinical trials involve the testing of therapeutics in humans. This stage is typically the most expensive within the pipeline and is where most drugs will fail due to toxic off-target effects, earning it the nickname 'The Valley of Death'<sup>16</sup>. Clinical trials are divided into phases I-III<sup>17</sup>. The objective of phase I trials is to determine the safety and tolerated dose of the drug and it involves 20-50 healthy volunteers and can last 2-6 months. Phase II gathers additional efficacy and safety data on the drug, involving more patients with the disease and lasting between 6 months and two years. In phase III trials, the drug is tested in hundreds or thousands of patients for 1-4 years. The objective is to gather safety and efficacy data over the long term and to compare the new treatment to existing treatments or a placebo. Phase III trials have the highest average cost of the three phases due to their complexities and long time frames.

Accelerating the drug discovery pipeline is important so that we can modulate critical diseases and respond quickly to emerging diseases. One method of acceleration is to expand the breadth of the discovery funnel by increasing the rate of hit identification and development into lead compounds<sup>18</sup>. There are several ways that information from high-resolution 3D protein structures can feed into these pre-clinical stages to accelerate their progress, which will be discussed in this introduction.

## 1.2 Protein structure determination for pre-clinical drug discovery

### 1.2.1 Applications for protein structures

#### 1.2.1.1 Identification of ligand binding sites

Some experimental ligand screens, like antibody screens, are unbiased towards a particular binding site and can identify novel regions to target. Determining a ligand-bound target structure can be used to identify the site at which the ligand binds, allowing further targeting of this site with small molecules. This structural information can feed into various processes such as virtual screening, fragment-based drug design or structure-based drug design. For example, when 17 inhibitory single-domain antibodies were isolated against the influenza virus RNA polymerase, their binding sites were mapped for downstream targeting with small molecules<sup>19,20</sup>. Although, these sites were likely limited to only the accessible positions on the surface, which may not be the most targetable with small molecules. Ligand-bound structures can also identify cryptic pockets, binding sites which were previously concealed in the apo structure but opened up upon ligand-binding<sup>21</sup>. These are attractive sites to target with drugs as they could allow for high subtype specificity and reduced off-target effects<sup>22</sup>.

Beyond simply identifying their location and structure, comparing the same ligand-binding site in related proteins allows chemists to identify non-conserved regions which may allow the development of selective inhibitors for the target and reduce off-target effects. In the case of rapidly evolving viruses, comparing the conservation of residues in the binding site of different strains allows the engineering of broad resistance or the targeting of specific drug resistant strains.

There is still useful information to be gained from target structures without ligands bound. Computational programs can predict binding sites in protein structures based on geometric properties<sup>23,24</sup>, energetics<sup>25</sup>, sequence conservation<sup>26</sup> and more recently, machine learning<sup>27</sup>. A program also exists to predict cryptic pockets from single protein structures<sup>28</sup>. Furthermore, even if a ligand binding site is already known, these predictive tools can identify new sites, expanding the protein search space and widening the drug discovery funnel.

#### 1.2.1.2 Virtual screening

Another major use for protein structures is in virtual screening (VS), which can be split into two categories: structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS). SBVS is the *in silico* screening of a compound library for predicted binding to a protein structure. It relies on molecular docking algorithms which predict the optimum target-ligand binding pose and the corresponding binding affinity<sup>29</sup>. The

programs use scoring functions based on experimentally determined rules of thermodynamics and protein-drug interactions. Usually multiple SBVS rounds occur, with initial rounds using faster docking at the expense of potentially reduced accuracy to allow for screening of a large number of compounds, and further rounds that involve re-docking the top scoring compounds with more rigorous sampling and scoring applied. Consensus scoring, whereby a library is docked using multiple algorithms that employ different scoring functions, is one way of reducing the error of binding prediction<sup>30</sup>. Docking is only a computational prediction, so compounds must also be tested experimentally; but by removing incompatible compounds and enriching the library for potential binders, the number of molecules requiring testing is greatly reduced.

LBVS is a complementary method where a library is screened for structural similarity to an existing hit using programs for the comparison of 2D fingerprints<sup>31</sup> or 3D shapes<sup>32</sup>, which makes the top scorers more likely to be active themselves. LBVS is regularly used to find analogues of active compounds such as during the development of Epalrestat<sup>33</sup>, but it can also be used to generate inhibitors of protein-protein interactions by identifying compounds similar to the interacting residues as identified in a high-resolution protein complex structure<sup>34</sup>.

The trend in VS is to screen larger libraries faster, because sampling a greater chemical space increases the chance of identifying more potent and higher affinity hits<sup>35</sup>. Readily available compound libraries such as the Enamine screening collection (4.0 million) and Chembridge library (1.3 million) allow wide sampling by removing the physical requirement of hosting a library locally. Databases such as the eMolecules database (20.3 million) combine libraries from multiple suppliers to increase the size further. Searching through these enormous libraries requires faster computers. The speed of VS programs scales up with the number of computing cores, so large computing clusters are key for modern VS, such as ARC3 and ARC4 hosted at the University of Leeds, with a combined total of 12,000 computing cores.

More recently, the libraries and programs have become even less tangible. Libraries do not have to exist physically thanks to virtual make-on-demand libraries like the Enamine REAL library<sup>36</sup> (6 billion), Enamine REAL space library (36 million) and the ZINC22 database<sup>37</sup> (37 billion). The computing resources to search these enormous libraries are not required to exist locally either. Instead, VS can employ cloud computing such as servers hosted by the likes of Amazon, Google or Microsoft. The Cherkasov and Gorgulla groups have now managed to dock >1 billion compounds using these systems<sup>38,39</sup>.

### **1.2.1.3 Structure-based drug design**

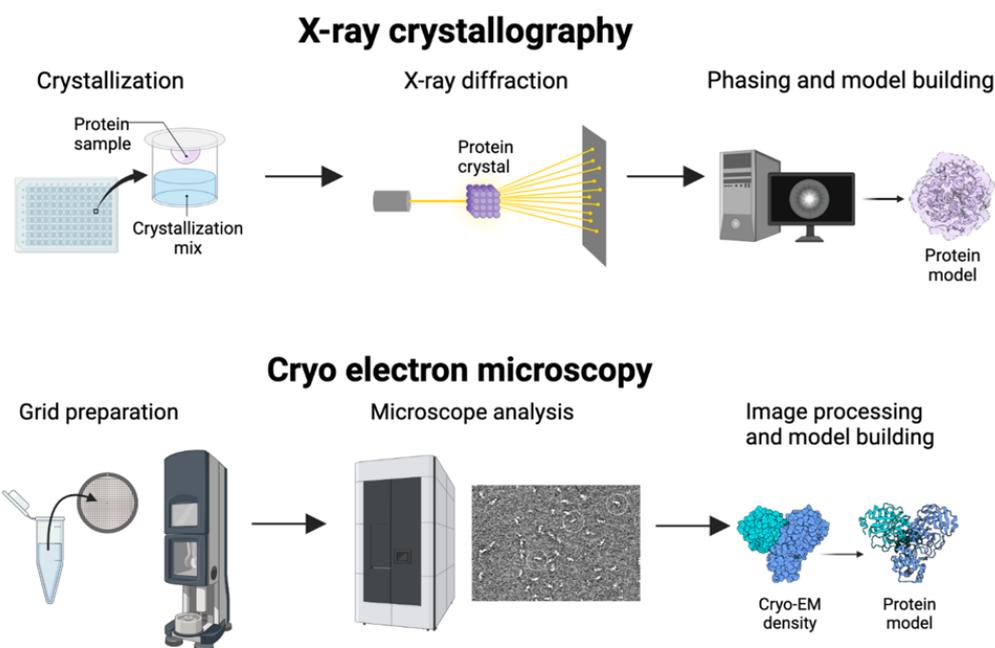
Although not essential, a small-molecule bound target structure can greatly accelerate hit-to-lead optimisation through structure-based drug design (SBDD). The structure can be used to identify regions of the binding site into which the molecule could be expanded to improve potency, functional groups on the molecule that could be switched for more complementary ones and non-conserved protein regions that would allow the introduction of selectivity by modifying the compound in a complementary way. Through organic synthesis, new analogues of hit compounds which satisfy these criteria can be synthesised, assayed and their target-bound conformation elucidated, in order to develop a structure-activity relationship (SAR). Typically, this cycle is iterated multiple times during the lead optimisation stage. The first major example of this rational design was in the development of the HIV protease inhibitors in the 1990s<sup>40</sup>.

#### **1.2.1.4 Fragment-based drug design**

Fragment-based drug discovery (FBDD) is a similar method to SBDD. First, small (150-300 Da) compound fragments are screened against a target in assays or through determining fragment-bound target structures<sup>41</sup>. The initial fragment hits are then grown or linked together using medicinal chemistry, typically guided by protein structures so that they better fit into the target site and increase in potency. A disadvantage of FBDD is that the small size of initial fragment hits means they tend to have lower target affinity, making them difficult to detect by cell-based assays. However, weak fragment binders can be detected by some biophysical assays including structure-determination methods like X-ray crystallography and cryo electron microscopy<sup>42,43</sup>. Several compounds have been developed using FBDD, including the anti-cancer drugs erdafitinib and vemurafenib, with Astex Pharmaceuticals pioneering this approach in industry<sup>44</sup>.

#### **1.2.2 Protein structure determination methods**

Two key methods for visualising proteins at the detail required for structure-based drug discovery are X-ray crystallography (XRC) and cryo electron microscopy (cryo-EM) (Figure 1.2). New programs now exist for the prediction of protein structures based on their primary amino acid sequence such as AlphaFold<sup>45</sup> and Rosetta Fold<sup>46</sup>. These predicted structures show real promise for aiding early stage drug discovery<sup>47</sup>, but they still require some form of experimental validation. Therefore, the focus in this thesis are the experimental methods for protein structure determination.



**Figure 1.2 Schematic for protein structure determination using X-ray crystallography and cryo electron microscopy.**

#### 1.2.2.1 X-ray crystallography

For many years, XRC was the primary method for generating high-resolution protein structures. First, purified protein is crystallised into an ordered solid lattice through the inclusion of certain precipitants, salts and buffer components<sup>48,49</sup>. Usually a large range of conditions are screened because the crystallisation conditions differ for each target<sup>49</sup>. Once crystals are obtained, these are shot with X-rays to generate an X-ray diffraction pattern. The diffraction pattern records the amplitude of the diffracted X-rays but not their phases, so the phases are calculated experimentally or predicted based on phases from closely related proteins<sup>50</sup>. From this information, the diffraction pattern is converted via an inverse Fourier transform into an interpretable electron density in which an atomic model is built<sup>50</sup>.

XRC has been revolutionary for pre-clinical drug discovery. It is high throughput, highly automated and can regularly generate  $<2 \text{ \AA}$  resolution structures for some systems, allowing the unambiguous placing of atoms and the visualising of interactions<sup>51</sup>. Therefore, it lends itself to the high throughput required for drug discovery. For example, XChem is a facility for FBDD at the UK Diamond Light Source where compound fragments are soaked into crystals and these are shot with X-rays, generating ligand-bound protein structures for any hits<sup>52</sup>. Hundreds of fragments can be screened in 2 days using XChem, with 80% of generated structures being higher resolution than the  $2.8 \text{ \AA}$  threshold required to accurately identify and place fragments within the density<sup>52</sup>.

However, crystallisation is the main bottleneck to solving structures by XRC. Proteins that do not easily crystallise include those that are large, flexible, dynamic, and/or have transmembrane portions. Therefore, despite the emergence of crystallisation methods like lipidic cubic phase, membrane proteins, which represent the majority of drug targets, are not well suited to XRC<sup>53</sup>. In order to obtain crystals, targets are often truncated, mutated or switched for more thermostable homologues, limiting the structural interpretability. Furthermore, once crystallised, the protein is locked by crystal contacts into a single conformation which is unlikely to be representative of its entire conformational landscape.

### **1.2.2.2 Cryo electron microscopy**

Cryo electron microscopy (cryo-EM) requires a thin, frozen film of sample through which a beam of electrons is transmitted. To achieve this, purified protein is applied to a mesh grid, and the majority is wicked away using blotting paper<sup>54</sup>. This leaves a thin layer of sample which is rapidly frozen by plunging the grid into liquid ethane, generating vitreous ice which preserves protein integrity and reduces electron-induced damage. Within the cryo electron microscope, a focussed beam of electrons is fired through the frozen sample onto a detector. Due to the radiation damage caused by electrons, images are first collected using a low dose of electrons which leads to images with poor protein signal. Later on, images of protein particles are averaged together to increase the protein signal compared to the background noise.

For cryo-EM single particle analysis (SPA), images of different copies of the protein are aligned and averaged together, first in 2D then in 3D to generate a cryo-EM density map into which a protein model can be built. For the complementary technique cryo electron tomography (cryo-ET), images of the same particle are collected at different tilt angles, then aligned and averaged together. Cryo-ET has limited resolution compared to SPA, but it is starting to show promise, with it recently generating an antibiotic-bound ribosome structure in cells at 3.5 Å resolution<sup>55</sup>. However, cryo-ET is still in its infancy regarding drug discovery, so the term cryo-EM will refer to SPA from here on.

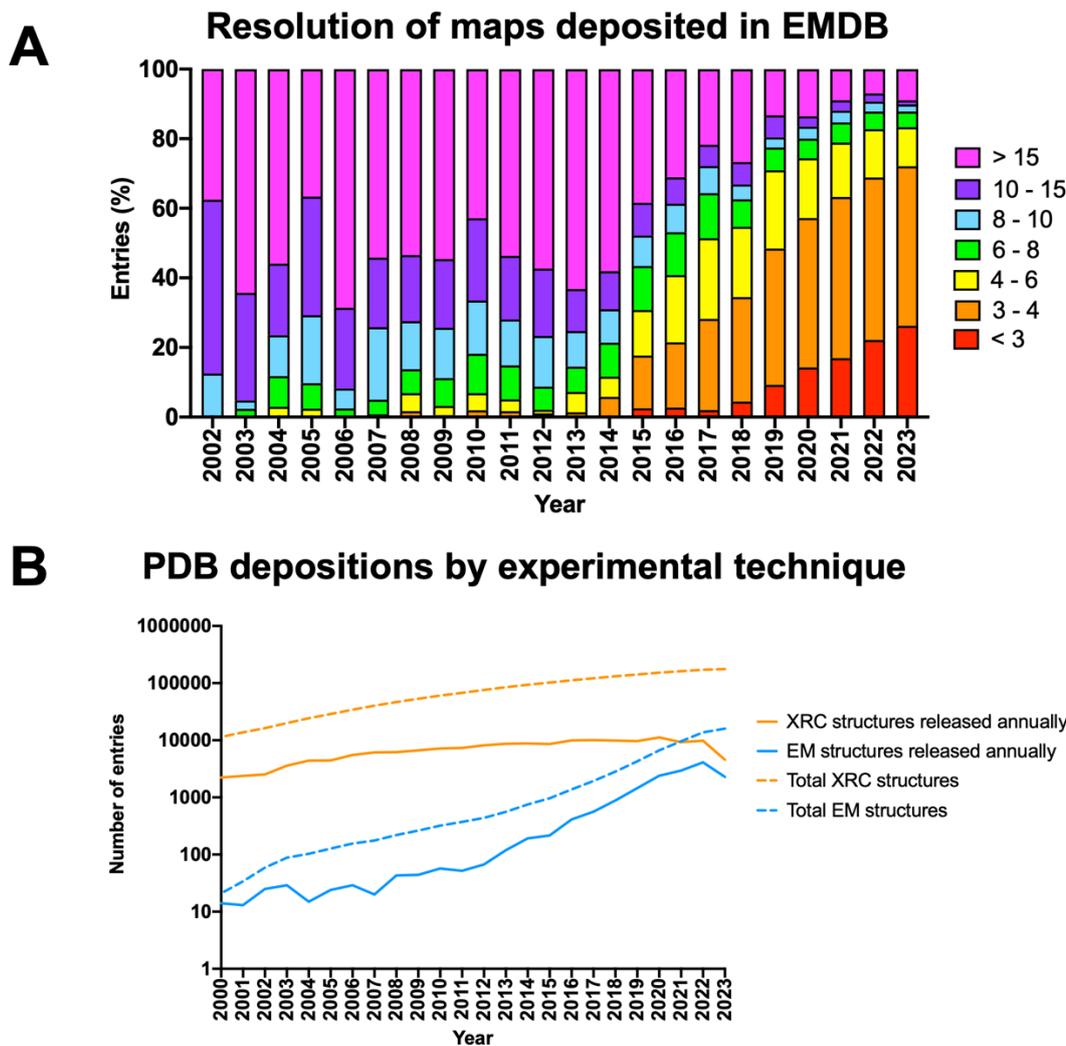
Around 2014, improvements in microscope stability, detector hardware and image processing software meant that cryo-EM regularly delivered sub-4.0 Å maps, causing the so-called 'resolution revolution'<sup>56</sup>. This shift is visible in the resolution of maps deposited in the electron microscopy database (EMDB) each year (Figure 1.3A). The resolution of deposited maps has only continued to improve since then, with <3 Å resolution maps now more frequent.

By circumventing the crystallisation bottleneck, cryo-EM has opened up structural determination and structure-based drug discovery to a wide range of targets. It has been particularly impactful for membrane proteins like G protein-coupled receptors (GPCRs), transporters and ion channels, which make up the majority of small molecule

targets but are challenging to crystallise<sup>57,58</sup>. In 2018, cryo-EM overtook XRC as the method which developed the majority of membrane protein structures. However, even in 2023, more deposited structures in the PDB are solved by XRC than cryo-EM, but the trends predict this changing in the coming years (Figure 1.3B).

The limits of cryo-EM are still being pushed beyond just resolution. Proteins smaller than 100 kDa were once not amenable to cryo-EM due to weak protein signal and minimal features for alignment. Through more sensitive detectors, grids with graphene support which leads to thinner, more uniform ice, and methods to bulk up targets, several proteins with masses of <50 kDa have now been determined to <4.0 Å resolution by cryo-EM<sup>59-61</sup>. The size limit has also been pushed in the opposite direction. Although cryo-EM has always been suited to large proteins and protein complexes, even 2.8 MDa and 14 MDa structures have now been determined by cryo-EM<sup>62,63</sup>. Unlike crystallography which often locks proteins in a single conformation, cryo-EM allows protein dynamics to be teased out by classifying multiple conformation states, even from a single dataset<sup>64</sup>. Furthermore, changes in the conformational landscape over time can be probed by time-resolved cryo-EM<sup>65-67</sup>.

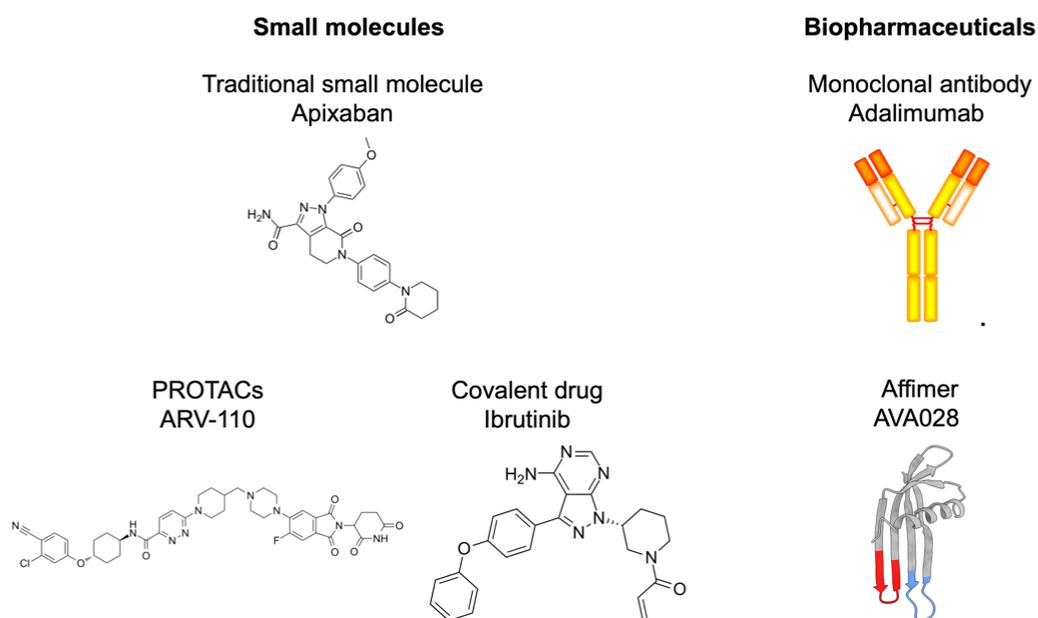
Even after the 2014 resolution revolution, the maximum resolution of structures determined by cryo-EM continues to improve year on year. For example, the highest resolution structure of the cryo-EM test sample apoferritin sits at 1.25 Å resolution and even a challenging protein like the  $\beta$ 3 GABA<sub>A</sub> receptor was resolved to 1.7 Å resolution<sup>68,69</sup>. Furthermore, the throughput of cryo-EM is increasing thanks to the increasing speed of data collection on new detectors. Although the throughput of cryo-EM will likely never rival that of XRC, multiple ligand-bound structures can now be determined from a single 24 hour data collection session<sup>70</sup>. All this progress means that many pharmaceutical companies across the world use cryo-EM for protein structure determination<sup>42,71</sup>.



**Figure 1.3 The rise of cryo electron microscopy for protein structure determination.** Data for 2023 correct as of July 2023. A) The resolution ( $\text{\AA}$ ) of maps deposited in the electron microscopy database (EMDB) per year since 2002. B) The total and annual number of structures deposited in the protein databank (PDB) per year.

### 1.3 Drug modalities

Although the focus so far has been on the discovery of small molecules, they are not the only class of therapeutics. Therapeutics can be binarized into two categories: small molecules, which are typically synthetic chemicals with masses  $< 500$  Da, and biologics, which are produced in cells and made of natural building blocks like amino acids (Figure 1.4). There have been developments in both modalities as well as the ability to switch between them.



**Figure 1.4** Examples of small molecule and biopharmaceutical drug modalities.

### 1.3.1.1 Small molecules

Since the advent of protein structure determination, small molecules have been designed to directly inhibit the function of a protein target through non-covalent binding to either the active site or an allosteric site. Most approved small molecules fit into this category, such as apixaban which is used to treat blood clots and binds non-covalently to the active site of coagulation factor Xa <sup>72</sup>. However, new modalities are challenging this paradigm. Covalently binding drugs can confer high selectivity through binding to non-conserved residues and they can inhibit ‘undruggable’ proteins without deep pockets <sup>73</sup>. For example, ibrutinib is used for the treatment of B-cell cancers and binds covalently to the active site cysteine of Bruton’s tyrosine kinase (BTK) <sup>74</sup>. In addition, proteolysis targeting chimeras (PROTACs) contain one portion of the molecule which binds the target and another portion which binds an E3 ubiquitin ligase, therein catalysing the degradation of the target. PROTACs allow for the blocking of non-enzyme proteins that cannot be inhibited traditionally, as well as proteins without deep binding pockets <sup>7</sup>. So far, only a handful of PROTACs have shown efficacy in patients, such as ARV110 for the treatment of prostate cancer, but none have completed clinical trials <sup>75</sup>. These new types of small molecule are fulfilling needs within drug discovery, but SBDD and VS programs are still more advanced for traditional non-covalent small molecules. Therefore, the latter will be the focus of this work.

### 1.3.1.2 Biopharmaceuticals

Although more recent additions to the drug discovery toolbox, biopharmaceuticals have cemented themselves as major therapeutics. The number of newly approved biologics

overtake small molecules for the first time in 2022 and they come with shorter development times on average compared to small molecules<sup>2,76</sup>. Biopharmaceuticals is a broad category of therapeutics including vaccines, recombinantly produced hormones and monoclonal antibodies (mAbs). Like small molecules, mAbs bind a target and block its function. However, their large size means they often perform this by blocking the target interacting with its binding partners. For example, adalimumab, the top selling mAb of 2022, is approved for the treatment of rheumatoid arthritis<sup>77</sup>. It is a human IgG molecule which binds the cytokine tumour necrosis factor  $\alpha$  (TNF $\alpha$ ) and prevents it binding its cell surface receptors<sup>78</sup>.

Despite the >100 mAbs approved for therapy<sup>79</sup>, they have limitations. Their large sizes (~150 kDa) mean mAbs cannot penetrate deep or densely packed tissues. mAbs are usually produced in a complex, animal-dependent process whereby the antigen-specific plasma cells are fused with myeloma cells to form antigen-producing hybridomas<sup>80</sup>. Hybridomas produce antibodies at relatively poor yield, so must be greatly scaled-up, resulting in high financial costs of mAb treatment<sup>81</sup>. Although, there are now animal-free in-vitro techniques for antibody production<sup>80,82</sup>. Furthermore, mAbs can also show issues with stability and aggregation.

These issues with mAbs have led to the development of other biopharmaceutical modalities such as artificial binding proteins like Affimers<sup>83</sup>. Affimers contain two hypervariable loops displayed on a synthetic, highly thermostable scaffold<sup>84</sup>. The two loops are responsible for target binding and each contain nine randomised residues of all naturally occurring amino acids except cysteine. The Affimers which contain the loops with the highest target affinity are isolated through phage display from a library of  $1 \times 10^{10}$  clones, then expressed in *E. coli* bacteria to very high yields.

Affimers have been isolated against a broad range of therapeutic targets involved in cancer<sup>85</sup>, viral infection<sup>86,87</sup> and pain<sup>83</sup>. Their small size (~12 kDa) means Affimers circumvent many of the issues associated with mAbs; they bind with a small surface area, they can penetrate more deeply into tissues and they can also be multimerised to broaden their uses. As well as being tools for labelling in microscopy<sup>88,89</sup> and diagnostics<sup>86,90</sup>, they can also be employed as therapeutics. For example, AVA028 is an Affimer entering the pre-clinical stage for the treatment of cancer through the inhibition of the PD-L1 cell surface protein<sup>91</sup>.

Despite the promises of biopharmaceuticals, both established mAbs and more recently developed Affimers suffer from some of the same issues. Firstly, neither can cross the blood brain barrier, meaning many diseases of the central nervous system are yet to be tackled by biopharmaceuticals. Secondly, although Affimers and mAbs can target secreted and cell surface proteins, neither can penetrate the cell membrane and target important cytosolic proteins. Finally, being proteinaceous reagents, neither have oral bioavailability due to their rapid breakdown by gastrointestinal proteases, so they

require administration by injection which is more invasive. Whereas, small molecules typically fulfil all of these criteria, and it is often easier to optimise the pharmacokinetics of a small molecule, so they are more appropriate than biopharmaceuticals for certain cases. This means that it is sometimes useful to switch modality from a biopharmaceutical to a small molecule.

### **1.3.1.3 Biopharmaceutical-guided small molecule development**

Different methods of antibody-guided small molecule discovery have been proposed. Many researchers have postulated that antibodies which stabilise inactive target conformations could be mimicked by therapeutic small molecules<sup>92</sup>. Similarly, since antibodies can stabilise a conformation which opens up novel, previously occluded ligand binding sites, they can act as clamps to open up this site for the screening of small molecule binders by HTS or SBVS<sup>93</sup>.

In a more direct switching of modality, in 1991, a mAb binding region was used to guide the development of a cyclic organic small molecule via a peptide mimetic intermediate molecule without the guidance from an antibody-bound target structure<sup>94</sup>. The small molecule mimetic acted through the same pathways as the original template mAb. More recently, a broadly neutralising antibody against the hemagglutinin spike protein in influenza A virus was used to generate a small molecule using HTS<sup>95</sup>. The resulting small molecule was similarly broadly acting, bound to the same location on HA as the template antibody but had the benefit of being orally bioavailable. With protein structures easier than ever to determine to high resolution, more LBVS programs and bigger commercially accessible small molecule libraries, antibody-guided small molecule development could become much higher throughput and with more structure-guided rational design than these two examples.

The large binding interface of antibodies mean other biopharmaceuticals may be better suited as templates for small molecule development. Affimers can access smaller pockets on a protein surface than antibodies due to their small binding sites. The loop residues in an Affimer responsible for binding and inhibiting KRAS showed high structural similarity to a small molecule inhibitor that bound at the same site, leading to speculation that the Affimer could be used to develop small molecule mimics in the future<sup>85</sup>. This could take a LBVS approach using the key Affimer residues responsible for binding as the template structure.

## **1.4 Objectives**

In this thesis, I will show how the structure-based drug discovery pipeline can be strengthened by new tools in cryo-EM and VS. These tools will be applied to targets involved in pathogenic infection including viruses, bacteria and parasites.

In chapter 3, advancements in cryo-EM detectors will be applied to two transmembrane drug targets. Data collected on current detectors was reconstructed to maps with significantly higher resolution compared to data collected on previous generations of detectors. The new maps allowed visualization of waters, lipids, natural substrates and small molecules and allowed for better protein modelling to guide small molecule discovery.

In chapter 4, a new *in silico* LBVS method for using Affimers as templates for small molecule discovery will be demonstrated against two viral targets.

In chapter 5, the Affimer-guided small molecules will be screened for their inhibition of each viral target, leading to the identification of one hit from each compound series.

Together this work shows that the types of questions that cryo-EM can answer in the drug discovery field are rapidly changing, and highlights the various routes that protein structures can feed into drug discovery to generate small molecule inhibitors.

# **Chapter 2 : Investigating the advances in cryo-EM structure determination**

## 2.1 Introduction

3D protein structures can aid drug discovery by characterizing ligand binding sites, being templates for virtual screening and for guiding hit-to-lead optimization. Compared to X-ray crystallography, cryo-EM is a less well established method for generating high-resolution structures. By circumventing the requirement for protein crystals, cryo-EM has been particularly useful for visualising proteins that are less amenable to crystallography such as membrane proteins and protein complexes. In 2015, there was a 'resolution revolution' in cryo-EM which was brought about by improved microscope stability, more sensitive electron detectors and more advanced software for image processing. However, even since this revolution, the resolution of cryo-EM maps has improved year on year. This has been primarily due to further advances in the electron detectors and new image processing algorithms.

### 2.1.1 Improvements in cryo-EM structure determination

The path of electrons in a cryo electron microscope begins with their emission from an electron gun, followed by their focussing through condenser lenses to form a beam which passes through the frozen sample within a grid cooled to liquid nitrogen temperatures (-196 °C)<sup>96</sup>. Then, the scattered electrons are focused onto a detector. Since the invention of stable microscopes that allow data collection over several days, the bulk of hardware advances have come in the systems following image formation. In particular, there have been great advancements in the electron detectors which have become more sensitive, most importantly at higher spatial frequencies.

The performance and sensitivity of a detector is quantified by its detective quantum efficiency (DQE). The DQE describes how much of the signal is captured by a detector at a certain spatial frequency on a scale of 0 to 1. The DQE for a detector varies with the spatial frequency of the electron events and it is described at frequencies relative to Nyquist, the maximum spatial frequency at which a detector can distinguish individual electron events. The Nyquist limit depends on the distance visible in each pixel, which varies based on the magnification applied to the scattered electrons. The DQE is described in McMullan *et al.* as Equation 2:1<sup>97</sup>.

**Equation 2:1 The detective quantum efficiency of cryo-EM detectors.**

$$\text{DQE} = \text{SNR}(\text{output})^2 / \text{SNR}(\text{input})^2$$

The DQEs of detectors have increased steadily over time as new generations have become available. For example, the DQEs of the K2 summit (Gatan) and K3 summit (Gatan) improved at 1x Nyquist, going from 0.23 to 0.40, and at 0.5x Nyquist, going from 0.56 to 0.71<sup>97,98</sup>. In addition, progressing from the Falcon II to Falcon III EC to Falcon 4 (All produced by Thermo Fisher Scientific), the DQE at 1x Nyquist improved from 0.27 to 0.34 to 0.50, and at 0.5x Nyquist, it improved from 0.54 to 0.78, then

decreased slightly to 0.72<sup>99,100</sup>. This means that data collected on the newer detectors have higher SNRs, especially at the high spatial frequencies close to Nyquist, meaning reconstructions may contain more high resolution information.

Energy filters come between the stage and the detector and remove inelastically scattered electrons that contribute to noise. This further boosts the SNR, particularly at high spatial frequencies. Energy filters have been included in cryo electron microscopes in some capacity since their invention, but the recently released Selectris energy filter (Thermo Fisher Scientific) has shown great promise due to its more stable and more narrow slit width, meaning fewer of the inelastically scattered electrons pass through, compared to its predecessors<sup>101</sup>. Recently, a cryo-EM set up with a Falcon 4 and Selectris energy filter was used to determine the highest resolution structure of the cryo-EM test specimen apoferritin to 1.2 Å as well as the transmembrane  $\beta 3$  GABA<sub>A</sub> receptor at 1.7 Å<sup>69</sup>. Since thicker ice leads to greater inelastic electron scattering, energy filtering has a greater effect in these areas. Energy filtering using the Selectris can allow for protein imaging in ice that was previously too thick for sample visualisation. This can be particularly useful when images are collected on a tilt, because the tilting increases the distance the beam must travel through.

Furthermore, each generation of camera has improved data collection speeds on its predecessor. For example, the K3 collects data 3.75x faster than the K2<sup>98</sup>, and the frame rates for the Falcon cameras has improved from 40 frames per second (fps) for the Falcon 3EC to 250 fps and 320 fps for the Falcon 4 and Falcon 4i, respectively<sup>99,100,102</sup>. Therefore, data is not only better quality, but it is also collected faster, meaning more structures can be determined in a shorter period of time and multiple ligand-bound structures can be determined within 24 hours<sup>70</sup>.

Relion and cryoSPARC are two software suites that have made image processing easier and faster, and have contributed to cryo-EM reconstructions becoming higher resolution over time<sup>103,104</sup>. In general, the general workflows for cryo-EM image processing have not changed drastically, but new processing algorithms have been released recently which have improved processing speeds and accuracy. With the rate of data collection rising, faster algorithms are essential. In previous versions of Relion, the traditional Expectation-Maximization (EM) algorithm employed for particle classification would iterate through the full dataset 25 times when used with the default settings. In the most recent Relion release, the new algorithm for 2D classification and 3D initial model generation termed Variable-metric Gradient Descent with Adaptive Moments estimation (VDAM) performs initial classification steps using small subsets of particles, with each iteration using more particles until the entire dataset is classified only five times under the default settings. This means the VDAM algorithm iterates through the entire data set fewer times than the EM algorithm, and classifying particles using the VDAM algorithm is on average five times faster<sup>105</sup>. In addition, the non-

uniform refinement algorithm exclusive to cryoSPARC treats regions of spatial variability differently, as is common for membrane proteins and protein complexes<sup>106</sup>. This algorithm improved the resolution of three membrane proteins compared to a traditional uniform refinement during testing<sup>106</sup>.

Two proteins that are useful test cases to benchmark the advancements in cryo-EM are quinol-dependent nitric oxide reductase (qNOR) from the pathogenic bacterial species *Alcaligenes xylosoxidans* (*Ax*) and cytochrome bc1 from the malaria parasite *Plasmodium falciparum* (*Pf*). They were both determined by the Muench group to ~4 Å resolution using former generations of detector and older image processing algorithms<sup>107,108</sup>. Higher resolution maps that can likely be generated using the newer cryo-EM setups could improve the information on protein function, structure and inhibitor binding, and as potential drug targets, this information can feed into important small molecule development. In addition, as membrane proteins, they are more challenging targets to work on than the traditional test proteins like apoferritin. For example, both proteins are solubilised in detergents which can reduce the contrast around the transmembrane region.

### 2.1.2 Quinol-dependent nitric oxide reductase

qNOR is an integral membrane metalloenzyme that sits in the cytoplasmic membrane of some bacterial species and belongs to a broader family of nitric oxide reductases (NORs). NORs are involved in a process called denitrification which functions as a respiratory pathway in oxygen-limited environments. In this process, subsequent oxidoreductases convert nitrate (NO<sub>3</sub><sup>-</sup>) to NO<sub>2</sub>, NO, N<sub>2</sub>O and finally nitrogen gas (N<sub>2</sub>). NORs catalyse the intermediary conversion of nitric oxide (NO) to nitrous oxide (N<sub>2</sub>O) in the reaction:  $2\text{NO} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{N}_2\text{O} + \text{H}_2\text{O}$ . qNORs are used to degrade NO produced by the host immune system in certain species of pathogenic bacteria including *Neisseria meningitidis*, which causes meningitis and septicaemia, *Neisseria gonorrhoeae*, which causes gonorrhoea and *Alcaligenes xylosoxidans* (*Ax*), an opportunistic bacteria<sup>107,109,110</sup>. NO is directly toxic to bacteria and it functions as a pro-inflammatory signalling molecule for the innate immune system to activate anti-bacterial immune cells<sup>111</sup>. Therefore, inhibition of qNOR is a potential route for anti-bacterial small molecules by increasing the susceptibility of the bacteria to the toxic effects of nitric oxide.

#### 2.1.2.1 *Ax* qNOR as a drug target

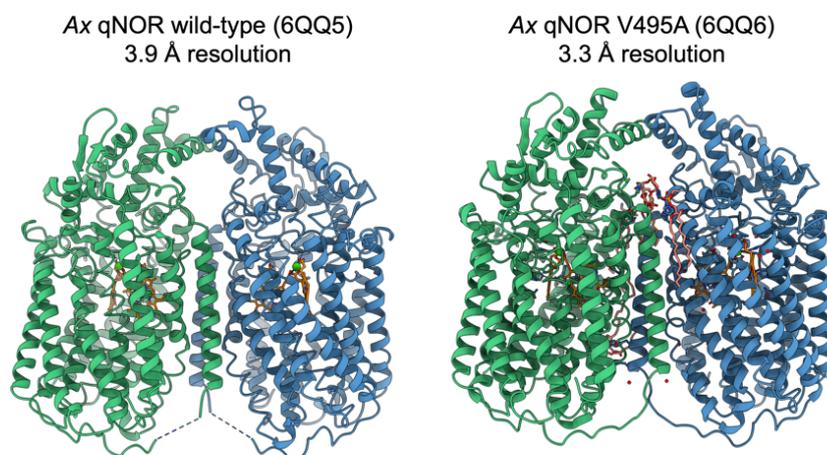
*Alcaligenes xylosoxidans* is a qNOR-containing pathogenic bacteria that mainly infects immunocompromised patients. Often, *Ax* infection is healthcare associated and leads to bacteraemia, the bacterial infection of the bloodstream<sup>112,113</sup>. *Ax* also infects the lungs of patients with cystic fibrosis (CF) and with 60,000 patients worldwide, CF is the most common life-shortening disease caused by a recessive autosomal gene when

measured within the Caucasian population <sup>114</sup>. CF is caused by a defect in an ion channel that causes a build-up of mucus in the lungs which becomes colonised by various bacterial species including *Ax*, leading to chronic pulmonary inflammation <sup>115</sup>. In one trial, *Ax* was found in mucus from 8.7% of CF patients, but it is unclear whether *Ax* contributes to CF disease or simply colonises the respiratory tract <sup>115,116</sup>.

Currently, *Ax* is treated with antibiotics but antibiotic resistance is one of the biggest threats to global health. New antibiotics are needed to continue treating bacterial infection and this involves finding new protein targets. Therefore, as well as being a test case for advancements in cryo-EM, *Ax* qNOR is also a potential antibiotic target to maintain the flow of effective therapeutics and a high resolution structure could aid this.

### 2.1.2.2 Structural analysis of *Ax* qNOR

The cryo-EM structures of wild-type and an activity-enhanced mutant of *Ax* qNOR were determined previously by the Muench and Hasnain groups to 3.9 Å and 3.3 Å resolution, respectively <sup>107</sup>. Cryo-EM identified qNOR to be a dimer, when it had been repeatedly solved as a monomer by X-ray crystallography, likely due to the binding of zinc from the crystallization conditions (Figure 2.1) <sup>109,117–119</sup>. Each monomer within *Ax* qNOR contains 18 helices that arrange around two heme groups termed heme b and heme b<sub>3</sub> in the protein core, with heme b<sub>3</sub> forming the binuclear centre with a non-heme iron atom.



**Figure 2.1** *Ax* qNOR structures determined by the Muench and Hasnain groups by cryo-EM. Subunit A and B shown as green and blue subunits, respectively. Heme groups, LOP lipid and detergents shown as orange, light salmon and dark salmon sticks, respectively. Water molecules shown as red spheres.

The *Ax* qNOR maps demonstrated the advantages of cryo-EM for studying membrane proteins in more native conformations but drug discovery based on the structures was limited by the low resolution. Although almost every protein residue could be placed in the cryo-EM map, precise side chain conformations were often missing and information

on metalloenzyme chemistry in the redox centre only becomes visible higher than  $\sim 2$  Å resolution. Also, it was suggested that water-mediated proton transport was important for Ax qNOR function. But at 3.9 Å resolution, no waters were visible in the wild-type Ax qNOR map and only 18 were visible in the mutant map, which is much fewer compared to higher resolution qNOR crystal structures from other species<sup>118</sup>. The location and conformation of the electron donor binding site was not visible in the Ax qNOR map, but it had been identified for qNOR from other species<sup>118</sup>. Finally, there were likely many more lipids and detergents to be visualised than the few that were visible in the mutant Ax qNOR.

### 2.1.3 Cytochrome bc1

Cytochrome bc1 (Cyt bc1) is another metalloenzyme that was previously visualised by the Muench and Hasnain groups using cryo-EM<sup>108</sup>. Cyt bc1 is found in prokaryotic energy-transducing and eukaryotic mitochondrial membranes. In the latter, it forms complex III of the mitochondrial electron transport chain and couples the oxidation of cytochrome c with the transmembrane pumping of protons. This generates a proton gradient that is harnessed by ATP synthase. The reaction that Cyt bc1 catalyses is:

$$\text{Ubiquinol (QH}_2\text{)} + 2 \text{ Cytochrome c (Fe}^{3+}\text{)} + 2 \text{ H}^+ \rightarrow \text{Ubiquinone (Q)} + 2 \text{ Cytochrome c (Fe}^{2+}\text{)} + 4 \text{ H}^+$$

Cyt bc1 forms a dimer, with each monomer containing eleven protein subunits that arrange around two heme groups heme  $b_H$  and heme  $b_L$ . There are two main ligand binding sites on Cyt bc1, the  $Q_o$  site where ubiquinol binds and the  $Q_i$  site where ubiquinone binds. These are the main inhibitor binding sites including for the small molecules antimycin A, which binds in the  $Q_i$  site, and myxothiazol and stigmatellin, which bind in the  $Q_o$  site<sup>120,121</sup>.

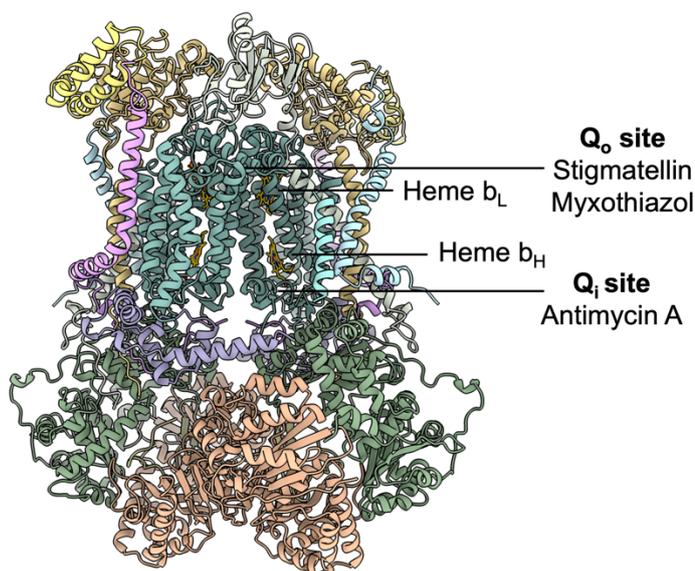


Figure 2.2 Bovine cytochrome bc1 and its small molecule binding sites.

### 2.1.3.1 Cytochrome bc1 as a parasitic target

Cyt bc1 in the human malaria parasite *Plasmodium falciparum* (*Pf*) is an antimalarial target. For example, atovaquone, an approved antimalarial small molecule, binds at the Q<sub>o</sub> site of Cyt bc1<sup>122</sup>. Drug selectivity is engineered in by exploiting sequence differences between the human and *Pf* versions of Cyt bc1.

Malaria is a disease caused by single celled microorganisms of the *Plasmodium* group which are spread by *Anopheles* mosquitoes<sup>123</sup>. There are five *Plasmodium* species that cause malaria, with *Plasmodium falciparum* causing the most severe form of malaria, resulting in the most deaths. Symptoms of malaria include chills and fever as well as headache, fatigue and muscle pain, and they arise from the infection of the parasite of the red blood cells<sup>123</sup>.

The number of malaria cases worldwide is high, with 247 million in 2021 and 619,000 estimated deaths from malaria that year<sup>123</sup>. Malaria infections primarily occur in South America, Asia and Africa, with the majority of deaths occurring in Africa. Since 2021, a malaria vaccine has been available<sup>123</sup>. Several drugs are available for malaria treatment, including the traditional cure quinine hydrochloride. However, there are issues with quinine hydrochloride including severe side effects and drug resistance, so the development of new, more selective drugs is needed.

### 2.1.3.2 Cytochrome bc1 structural analysis

Due to its essential role, Cyt bc1 has been studied extensively. The first structures of Cyt bc1 were of the bovine homolog generated by X-ray crystallography published in 1997 and 1998<sup>124,125</sup>. Since then, structures of Cyt bc1 from other species have been determined, including that of the chicken homolog by X-ray crystallography<sup>121</sup>, and the homologs from humans<sup>126</sup> and the purple bacterium *Rhodobacter Sphaeroides*<sup>127</sup> by cryo-EM. However, the *Pf* structure has not yet been determined, meaning it cannot feed into structure-based drug discovery of antimalarial inhibitors. Instead, homolog structures, particularly the bovine homolog, have been determined in complex with inhibitors to understand their general binding modes and to rationally engineer in selectivity for *Pf* using homology modelling<sup>128</sup>.

The main reason for the lack of a *Pf* cyt bc1 structure is the challenge in purifying sufficient quantities for X-ray crystallography. With eleven subunits in higher animals, recombinant protein expression of cytochrome bc1 into the native, functional conformation is incredibly challenging. To address this, the Muench and Hasnain groups demonstrated that a structure of bovine cyt bc1 purified from native tissue could be determined by cryo-EM, which has much lower sample requirements compared to crystallography<sup>108</sup>. The structures were also determined with the ligands GSK932121

and SCR0911. Since then, the Hasnain group has been developing a workflow for purifying native *Pf* Cyt bc1 from infected tissue and I have been involved in assessing the suitability of these samples for cryo-EM. However, this has been a challenging project and to date, *Pf* bc1 has not been visualised by EM, despite promising activity assays and protein analysis gels, and therefore it is not reported within the thesis. Despite these challenges, further improvements to these samples may allow for the generation of the first *Pf* bc1 structure for developing antimalarial compounds.

There is still information to be gained from structures of homologs bound to other *Pf* bc1 inhibitors. The previous ligand-bound cryo-EM structures were only determined to 4.1 Å resolution, meaning the exact conformations of ligands were unclear<sup>108</sup>. So, new ligand-bound structures would also demonstrate the information available in the higher resolution maps generated today thanks to the advances in cryo-EM outlined in section 2.1.1.

#### **2.1.4 Objectives**

The objectives of this chapter were to investigate the effects that advances in cryo-EM hardware and software can have on two therapeutic membrane protein targets: *Ax* qNOR and bovine cytochrome bc1. Specifically, what is the new information visible in high resolution maps and how can it feed into structure-based drug discovery? To investigate this, an improved cryo-EM workflow was used to improve the resolution of a map of *Ax* qNOR, with cryo-EM density visible for electron donor, waters and lipids. Attention was then switched to improving the resolution of a bovine Cyt bc1 cryo-EM structure in the apo state and bound to two small molecules.

## **2.2 Materials and methods**

Section 2.2.1 will focus on the general cryo-EM methodology from grid preparation through to atomic model building and sections 2.2.2-2.2.4 will focus on the exact methods used to generate the structures presented in this chapter.

### **2.2.1 Overview of single particle cryo-EM analysis**

#### **2.2.1.1 Grid preparation**

Single particle cryo-EM requires a protein to be suspended in a thin film within a grid under cryo-conditions. EM grids are typically made of copper or gold and contain a carbon or gold mesh, usually with regularly spaced holes within the grid bar support<sup>129</sup>. Before sample application, the grids require glow discharging, which increases the hydrophilicity of the mesh, increasing the adsorption of the biological material onto the grid.

The standard approach for cryo-EM grid preparation is the side blotting method, which can be performed on commercial machines like the Vitrobot (Thermo Fisher Scientific). In order to generate a thin layer for imaging, once sample is applied and it fills the holes, the majority is wicked away with blotting paper. The remaining thin layer is rapidly plunged into liquid ethane, which has a high heat capacity, meaning it transfers heat from the sample so quickly that the protein solution freezes as vitreous ice. Vitreous ice is non-crystalline and does not diffract electrons like typical crystalline ice, reducing the background noise. Once vitrified, the grids are clipped by encasing them in a metal ring to increase their stability and aid automatic grid handling in the microscope. Clipped grids are stored in liquid nitrogen until use, and must be kept below -150 °C for imaging.

### **2.2.1.2 Cryo electron microscope setup**

Grids are then imaged in a cryo electron microscope in which electrons travel in the path as follows. Firstly, electrons are generated from a source, typically a field emission gun (FEG), then focussed into a beam using the condenser lens system. The beam hits the frozen sample and when the electrons encounter electrostatic potential in the protein, they are scattered and their phases are shifted. This scattering can be elastic, which is useful for structure determination, or inelastic, which does not contribute to phase contrast and adds noise to the images. An energy filter placed after the stage can be used to remove most of these inelastically scattered electrons. Finally, the elastically scattered electrons are focussed using an objective lens system onto a detector which records the phases and amplitudes of the electrons that interact with it.

When electrons interact with proteins, they transfer energy and cause motion in the sample. In a single exposed image, uncorrected motion would blur the image and limit the resolution of any reconstruction. To circumvent this, cryo-EM detectors record multi-frame movies in which particles can be tracked and beam-induced motion can be corrected for <sup>130</sup>.

The detector contains an array of pixels with a pixel size based on the magnification applied to the image. The physical resolution limit termed Nyquist resolution is always double the pixel size because only electron events >2 pixels apart can be detected. Therefore, a choice must be made about the magnification at which the data is collected. A lower magnification with a higher pixel size leads to a wider field of view, increasing the number of particles within the images. The higher pixel size limits the resolution, but not all biological questions require high-resolution information to answer them. A high magnification with a lower pixel size increases the amount of high-resolution information within the movies and lowers the Nyquist resolution limit. However, this also narrows the field of view and may require the collection of more images.

### 2.2.1.3 Data collection

An ideal grid for data collection would have various features. Firstly, the ice should be entirely vitreous. Vitreous ice can devitrify to crystalline ice if it is warmed above -137 °C, which can occur in error during grid preparation or transport<sup>131</sup>. Protein in crystalline ice has lower contrast and is likely to be damaged by the ice crystals. Secondly, the ice should be thin enough to visualise protein particles with good contrast but not too thin that the sample is dehydrated or denatured. Thirdly, the sample should be in its full orientation landscape, to aid alignment and to provide information on all regions of the protein, and to prevent map anisotropy. Protein preferred orientation occurs with various samples including hemagglutinin and bacterial large ribosomal subunit assemblies due to interactions of the protein with the air-water interface<sup>132</sup>.

One or many of these features may be compromised, making a grid unsuitable for cryo-EM data collection, so multiple grids of the same condition are produced and screened in the microscope to identify the most suitable ones. Once a satisfactory grid has been identified, movies are collected at different locations around the grid. This process is aided by automation now, with Titan Krios microscopes being controlled by the EPU software (Thermo Fisher Scientific) which can automatically select squares and holes with suitable ice thickness and move around the grid to collect data at those locations<sup>133</sup>.

As the image forms, it is distorted by the contrast transfer function (CTF), which causes the loss of signal at certain spatial frequencies, depending on the defocus applied to the image. To obtain movies with a range of CTF signals, movies are collected at a range of defocus values, allowing a final reconstruction with information at all frequencies.

Data is collected at low dose to reduce electron-induced sample damage such as the breaking of disulphide-bonds, decarboxylation of carboxylic acid side chains and macroscopic bubble formation, resulting in data with a low signal to noise ratio<sup>134</sup>. The signal is boosted by collecting many (>1000) movies and averaging the particles from these movies together. With increases in detector frame rates and other tools, now up to 1000 movies can be collected per hour. And computational features like multigrid EPU (Thermo Fisher Scientific) now allow for the automated collection on multiple grids, meaning that multiple ligand-bound structures can be determined overnight without the need for user intervention<sup>135</sup>.

### 2.2.1.4 Image processing

Once movies are collected, the 2D projections of the sample are processed into a 3D reconstruction, usually either the Relion or cryoSPARC software suites<sup>103,104</sup>. Relion,

which is based on a maximum likelihood algorithm for the alignment of particles, is open source and was the most used software for maps deposited in the Electron Microscopy Data Bank (EMDB) until 2023, when cryoSPARC overtook it <sup>103,136</sup>. Compared to Relion, cryoSPARC follows a similar processing workflow, but it has a more intuitive GUI and its algorithms run faster. The main disadvantage of cryoSPARC is that it requires a license which starts at ~£100,000 for industrial users, but is free for academic users <sup>104</sup>.

In pre-processing, the movie frames are averaged together into single micrograph in which the beam-induced motion is corrected for <sup>137</sup>. Then, the CTF is estimated for each micrograph, so that it can be corrected for in later processing <sup>138</sup>. Reconstructing requires images of individual particles, so the areas of the micrograph containing protein particles are selected. Nowadays, particles are picked automatically, often using machine learning based programs like crYOLO <sup>139</sup>, which have been trained to identify protein particles based on their contrast, size and appearance. The picked particles are extracted with some surrounding ice into a box with dimensions according to particle size. In order to speed up initial particle curation, the particle stacks can be binned by averaging neighbouring pixels together. This limits the maximum resolution but high resolution features are not important at this early processing stage.

During particle classification, particles are first classified in 2D, by aligning particles against other similar particles to generate discrete classes showing the common orientations. The particles in each class are averaged together, boosting the SNR in the classes. 'Junk' particles which contain ice contaminants or deformed particles will classify into blurry, feature-less classes and these can be removed and excluded from downstream averages. Whereas, identical, homogeneous, particles are sorted into high-resolution classes. The 2D classes demonstrate the variety of particle orientations within the micrographs. A broad range of particle orientations are required to generate a high resolution reconstruction, and particle preferential orientation leads to map anisotropy which limits the resolution. Particles then undergo 3D classification to further clean up the particle stack.

During 3D refinement, small changes in the sampling of optimum angles and translation are made, leading to an increase in the high resolution features within the map. Furthermore, for each individual particle, the beam-induced motion is corrected for and the CTF is estimated, leading to a further increase in resolution. Post-processing is used after 3D refinement to mask out the solvent region and sharpen the map by restoring high resolution features during refinement.

#### **2.2.1.5 Model building**

Once a suitable map has been determined, an atomic protein model can be built into it. The protein coordinates are refined based on the cryo-EM density in the map, as well as defined protein geometry such as bond lengths and angles.

## **2.2.2 Structure determination of Ax qNOR by cryoEM**

### **2.2.2.1 Protein purification and cryo-EM grid preparation**

The purification of Ax qNOR and its vitrification on a cryo-EM grid was performed previously by Gopalasingam *et al.*<sup>119</sup>. Protein was expressed recombinantly as His-tagged Ax qNOR-BRIL in C41 (DE3) *E. coli* cells grown in 2xYT media. Harvested cells were lysed by sonication and protein was extracted from purified membrane fractions in 50 mM tris (pH 7.0), 150 mM NaCl, and 1% (v/v) n-Dodecyl- $\beta$ -D-maltoside (DDM) (Anatrace) for 2 hours at 4°C. Solubilised material was applied to a pre-equilibrated 5 mL HisTrap column (GE Healthcare), washed in 3 column volumes (CV) of 50 mM tris (pH 7.0), 150 mM NaCl, 20 mM imidazole, 0.05% DDM, then 2 CV of the same buffer with 35 mM imidazole. Fractions containing Ax qNOR were concentrated using Amicon Ultra 100K concentrators and then applied to a Superdex 200 10/300 Increase (GE Healthcare) column equilibrated in 50 mM tris (pH 7.0), 150 mM NaCl, and 0.05% (v/v) decyl-thio-maltoside (DTM) (Anatrace). Fractions containing Ax qNOR were  $A_{410}/A_{280} > 0.7$  were concentrated to 20 mg/mL and flash-frozen in liquid nitrogen prior to storage at  $-80^{\circ}\text{C}$ .

Cryo-EM grids were prepared of the protein by applying sample at 3 mg/mL to glow-discharged Quantifoil Au R1.2/1.3 grids and plunge freezing them using a Vitrobot Mark IV (FEI).

### **2.2.2.2 Data collection**

All cryoEM data were collected on a Titan Krios cryo microscope (Thermo Fisher Scientific) at the Astbury Biostructure Laboratory at the University of Leeds. The microscope was operated at 300 kV and was equipped with a Falcon 4 camera and a Selectris energy filter set to 10 e<sup>-</sup> width (Appendix Table 5.1). In total, 5466 movies were collected at a nominal magnification of 130,000x, a pixel size of 0.91 Å/pixel and over a defocus range of -0.9 to -2.7  $\mu\text{m}$ . A total dose of 34.90 e<sup>-</sup>/Å<sup>2</sup> was applied to the movies over 6.11 seconds exposure time, corresponding to a dose of 5.71per Å<sup>2</sup>/second. The movies were split into 34 frames, giving a dose per frame of 0.8 e<sup>-</sup>/Å<sup>2</sup>.

### **2.2.2.3 Image processing**

All image processing was performed in RELION 3.1<sup>140</sup> unless otherwise stated. For pre-processing, movie frames were averaged to a single micrograph and beam-induced motion correction was performed using RELION's own implementation of MotionCor2<sup>137</sup>. Then, the CTF of each micrograph was estimated using CTFFIND-4.1

<sup>138,141</sup>. The particles were picked on the micrographs by crYOLO 1.6.1 (Sphire) <sup>139</sup> using the weights from its general model and a picking threshold of 0.1. In total, ~3 million particles were picked and extracted from the micrographs and underwent two rounds of 2D classification. After removing particles that belonged to poorly resolved or “bad” classes, 1.8 million particles remained. These particles then underwent 3D classification using a previously solved Ax qNOR structure (PDB: 6QQ5) <sup>119</sup> as a 3D reference that was low-pass filtered to 60Å resolution to avoid introducing model bias. The resulting 730k particles were refined using RELION's auto-refine program, then underwent a further round of 3D classification leaving 404,950 particles. The particles were extracted without re-scaling into a box with 240 pixel width and refined in C2 symmetry to give a resolution of 3.8 Å. The refined map was post-processed in RELION, whereby it was sharpened according to the global B-factor with a mask applied around the protein region, which improved the resolution to 3.3 Å. After two rounds of both Bayesian Polishing and per-particle CTF refinement, the resolution had improved further to 2.7 Å. In order to better model the de-localised CTF signals at the protein exterior, and to follow the guidance that smaller proteins require larger box sizes <sup>142</sup>, the particles were re-extracted into a larger box with a width of 400 pixels. The particles underwent three further rounds of both per-particle motion correction and CTF refinement. After a final 3D refinement, followed by masking out the solvent and detergent micelle, the final resolution was 2.2 Å, as calculated by the gold standard half map criteria with a Fourier Shell Correlation (FSC) cut-off of 0.143.

#### **2.2.2.4 Model building**

Model building was performed by Dr Svetlana Antonyuk. Using a starting model of Ax qNOR(PDB ID: 6QQ5), the protein coordinates were refined into the cryo-EM map using the molecular replacement program MOLREP <sup>143</sup> of the CCPEM suite <sup>144</sup>. This was followed by jelly-body refinement in REFMAC5 <sup>145</sup> in CCPEM against the cryo-EM half maps with final FSC average of 0.85 and root-mean-square deviation (RMSD) of the map against the model of 0.012 Å.

The Ax qNOR model (PDB ID: 6QQ5) was then built into the post-processed map that was sharpened according to the global B-factor. Density modification was performed on the map using the cryo-EM half maps and a refined model of Ax qNOR <sup>146</sup>. Density modification uses a maximum-likelihood function to boost high resolution features identified in the protein model that are present in the experimental cryo-EM data. This gave a map with improved density, so further model building was performed on this map and model refinement was performed using the non-sharpened, non-modified map followed by manual rebuilding in Coot <sup>147</sup>. Model building was assisted using cryoEM density difference maps, which highlight regions of unmodelled or incorrectly

modelled cryo-EM density, produced by Refmac-Servalcat in the CCP-EM suite <sup>148</sup>. Maps were visualised and figures were produced using ChimeraX <sup>149</sup>.

## **2.2.3 Bovine cytochrome bc<sub>1</sub> cryo-EM structure determination**

### **2.2.3.1 Protein purification**

Bovine cytochrome bc<sub>1</sub> was produced previously by Ampornnanai *et al.* according to the workflow described previously <sup>108</sup>. To summarise, bovine mitochondria were purified from fresh bovine hearts. Mitochondrial protein was extracted in buffer containing DDM, then Cytochrome bc<sub>1</sub> was purified using anion exchange chromatography, size exclusion chromatography and PEG precipitation.

### **2.2.3.2 Grid preparation**

Grids were prepared by Ampornnanai *et al.* following the method described previously <sup>108</sup>. To summarise, purified bovine cytochrome bc<sub>1</sub> was buffer-exchanged into 25 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA and 0.015% DDM and the protein concentration adjusted to 5 mg/mL. 3  $\mu$ L of the sample was applied to Quantifoil Cu R1.2/1.3 holey carbon grids, which had been glow-discharged for 30 s using a Pelco glow-discharge unit. The grids were blotted using a Vitrobot (FEI) for 6 seconds, at a blot force of 6, at 100% humidity and at 4°C before being plunge frozen in liquid ethane.

### **2.2.3.3 Data collection**

All cryoEM data were collected on a FEI Titan Krios TEM at the Astbury Biostructure Laboratory at the University of Leeds. The 300 kV microscope was equipped with a Falcon 4i camera (Thermo Fisher Scientific) and a selectris energy filter set to 10 e<sup>-</sup> width (Appendix Table 5.2). In total, 8,000 images were taken at a nominal magnification of 165,000x, a pixel size of 0.74 Å/pixel and a defocus range was used of -0.9 to -2.7  $\mu$ m. A total dose of 40.57 e<sup>-</sup>/Å<sup>2</sup> was applied to the movies over an exposure time of 3.22 seconds, corresponding to a dose per Å<sup>2</sup>/second of 12.60. The movies were split into 51 frames, giving a dose per frame of 0.8 e<sup>-</sup>/Å<sup>2</sup>.

### **2.2.3.4 Image processing**

All image processing was performed in RELION 4.0 <sup>105</sup> unless otherwise stated. For pre-processing, beam-induced motion correction was performed on the movies using RELION's own implementation MotionCor2 <sup>137</sup>, then the contrast transfer function (CTF) of each micrograph was estimated and corrected for using CTFFIND-4.1 <sup>138,141</sup>. The particles were picked on the motion corrected micrographs by crYOLO 1.8.0 (Sphire) <sup>139</sup> using the weights from its general model and a picking threshold of 0.1. In total ~1.5 million particles were picked and extracted from the micrographs and

underwent one rounds of 2D classification using the VDAM classification algorithm. After removing particles that belonged to poorly resolved or “bad” classes, 430k particles remained. These particles then underwent 3D classification using a previously published bovine cytochrome bc1 apo map (EMDB: EMD-4288)<sup>119</sup> as a 3D reference that was low-pass filtered to 60 Å resolution to avoid model bias.

The resulting 290k particles were re-extracted in a 500 pixel box without re-scaling, then refined using RELION’s auto-refine program in C1 symmetry to generate a final map with 3.08 Å resolution. The particles went through three rounds of both Bayesian Polishing and per-particle CTF refinement, with 3D refinement and post processing in between. After the first round, 3D refinement was performed with C2 symmetry. The final map had a global resolution of 2.06 Å as calculated by the gold standard half map criteria with a 0.143 cut-off.

### **2.2.3.5 Model building**

Model building was performed by myself and Dr Svetlana Antonyuk. A previously published apo cytochrome bc1 structure (PDB: 6FO2) was used as a starting model. Extra protein regions, lipids and water molecules were built using Coot<sup>147</sup> and refined against the cryo-EM half maps using Refmac-Servalcat<sup>148</sup>.

## **2.2.4 Determining ligand-bound Bovine cytochrome bc1 structures by cryo-EM**

### **2.2.4.1 Grid preparation**

Purified bovine cytochrome bc1 was produced by Nattapon Pinthong according to the method described in Ampornchai et al.<sup>108</sup>. The small molecules CK267 and SG114 were supplied in 100% DMSO at 100 mM and 10 mM, respectively. The protein was mixed with the compound so that the final bc1 concentration was 5 mg/mL and the DMSO concentration was 1% DMSO. Samples were incubated at 4 °C for 30 mins. There was aggregation in the sample with CK267, likely due to compound insolubility at 1 mM and 1% DMSO, so aggregates were removed by centrifugation at 16,000 g for 5 minutes.

Samples were applied to glow discharged Quantifoil Cu R1.2/1.3 carbon cryo-EM grids and UltraFoil Au R1.2/1.3 gold cryo-EM grids. Using a Vitrobot IV (Thermo Fisher Scientific) at 100% humidity and 6 °C, the sample was blotted off with a blot force of 1 for 6 seconds, then plunge frozen in liquid ethane before storing in liquid nitrogen.

### **2.2.4.2 Data collection**

All cryoEM data were collected on a FEI Titan Krios TEM at the Astbury Biostructure Laboratory at the University of Leeds. The 300 keV microscope was equipped with a

Falcon 4i camera and a selectris energy filter set to 10 e<sup>-</sup> width (Appendix Table 5.3 and Table 5.4). Initially, 1,936 and 2,500 images were taken of bc1 with CK267 and SG114, respectively. This was at a nominal magnification of 165,000x, a pixel size of 0.74 Å/pixel and a defocus range of -0.9 to -2.7 μm. A total dose of 35.51 e<sup>-</sup>/Å<sup>2</sup> was applied to the movies over 2.49 seconds, corresponding to a dose of 14.26 per Å<sup>2</sup>/second. The movies were split into 44 frames, giving a dose per frame of 0.8 e<sup>-</sup>/Å<sup>2</sup>.

A second data collection was performed for bc1 with CK267 due to a low number of particles from the first collection. This was done using the same parameters as above except using a total dose of 35.48 e<sup>-</sup>/Å<sup>2</sup> over an exposure time of 2.78 seconds, corresponding to a dose of 12.76 per Å<sup>2</sup>/second. The movies were split into 44 frames, giving a dose per frame of 0.8 e<sup>-</sup>/Å<sup>2</sup>.

#### **2.2.4.3 SG114 image processing**

All image processing was performed in RELION 4.0<sup>105</sup> unless otherwise stated. Pre-processing followed the same method described in section 2.2.2.3.. Particle picking using crYOLO 1.8.0 (Sphire)<sup>139</sup> used a picking threshold of 0.05. In total, ~220k particles were picked and extracted from the CTF corrected micrographs and put through one round of 2D classification using the VDAM algorithm. After removing poorly aligned particles and those that belong to poorly resolved or “bad” classes, 160,000 particles remained. These particles then underwent 3D classification using a map generated *ab initio* from the same dataset as a reference. The 122,000 particles from two classes containing bovine bc1 were re-extracted in a 500 pixel box without re-scaling, then refined using RELION’s auto-refine program in C1 symmetry to generate a final map with 3.0 Å resolution.

The particles went through three rounds of both Bayesian Polishing and per-particle CTF refinement, with 3D refinement and post processing in between. After the first round, 3D refinement was performed with C2 symmetry. The final map had a global resolution of 2.0 Å as calculated by the gold standard half map criteria with a 0.143 cut-off.

#### **2.2.4.4 CK267 image processing**

Both bc1 + CK267 datasets were processed using RELION 4.0<sup>105</sup> and the same workflow unless otherwise stated. Pre-processing followed the same method described in section 2.2.2.3.. Particle picking using crYOLO 1.8.0 (Sphire)<sup>139</sup> used a picking threshold of 0.05. From the two datasets, ~100k and ~450k particles were picked and extracted from the CTF corrected micrographs and put through one round of 2D classification using the VDAM algorithm. After removing poorly aligned particles and those that belong to poorly resolved or “bad” classes, 60k and 270k particles remained. These particles then underwent 3D classification using a map generated *ab initio* from

the same dataset as a reference. The 122k particles from two classes containing bovine bc1 were re-extracted in a 500 pixel box without re-scaling, then refined using RELION's auto-refine program in C1 symmetry to generate final maps with 3.11 Å and 3.03 Å resolution.

The particles went through three rounds of both Bayesian Polishing and per-particle CTF refinement, with 3D refinement and post processing in between. After the first round, 3D refinement was performed with C2 symmetry. After the third round, the particles were combined and refined together, followed by a final round of polishing and CTF refinement. The final map had a global resolution of 2.03 Å as calculated by the gold standard half map criteria with a 0.143 cut-off.

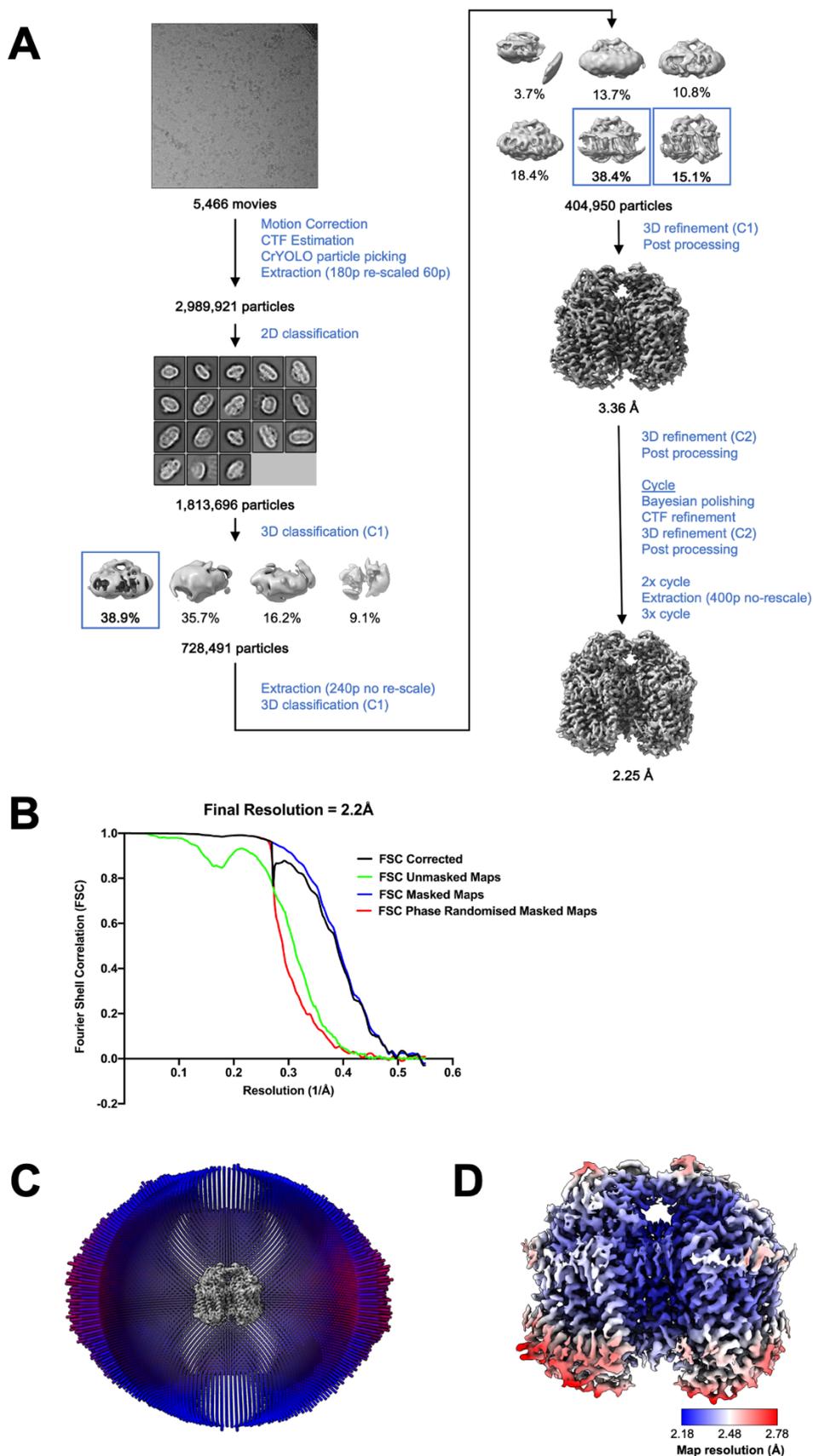
## **2.3 Results**

The objectives of this chapter were to use advances in cryo-EM to improve the resolution of maps for two pathogenic targets and to obtain new structural information. These targets were qNOR from the bacterial species *Ax* and cytochrome bc1 from malaria parasite *Pf* or its bovine homolog.

### **2.3.1 *Ax* qNOR was determined to 2.2 Å resolution using cryo-EM**

#### **2.3.1.1 Images collected on a Falcon 4 detector were processed to a 2.2 Å map**

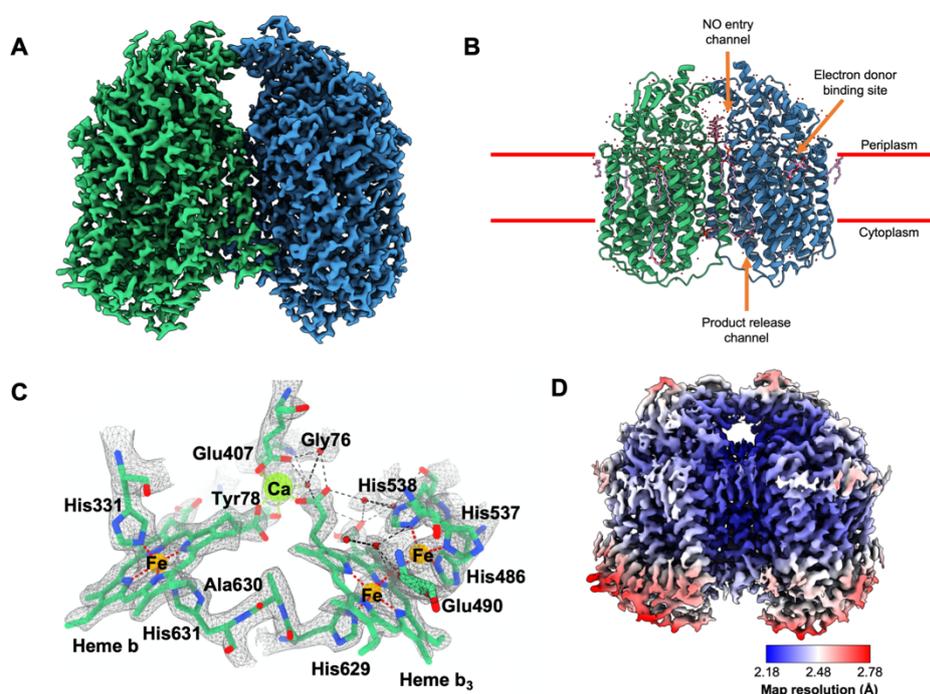
To test how much the resolution of the *Ax* qNOR map could be improved by collecting on a new cryo-EM setup, grids of *Ax* qNOR produced by Dr Rachel Johnson at the same time as those that led to the 3.9 Å structure in 2019<sup>107</sup> were imaged in an upgraded Titan Krios cryo electron microscope fitted with a Falcon 4 detector and a Selectris energy filter (Appendix Table 5.1)<sup>150</sup>. The movies were processed in Relion following a standard workflow through pre-processing, particle curation and 3D refinement (Figure 2.3A). C2 symmetry was only applied for later refinements, once C2 symmetry was clearly visible in map refined in C1. This generated a final map with a global resolution of 2.2 Å (Figure 2.3B). The orientations in the final reconstruction were well distributed, with a slight preference for side views (Figure 2.3C). The global resolution of the new map was 1.7 Å better than that of the previous *Ax* qNOR map and placed it in the top 1% highest resolution cryo-EM structures according to the EMDB as of July 2023<sup>151,152</sup>.



**Figure 2.3** A cryo-EM map of the transmembrane quinol-dependent nitric oxide reductase (qNOR) from the pathogenic bacterial species *Alcaligenes xylosoxidans* (Ax) was determined to 2.2 Å resolution. A) Image processing workflow performed in Relion 3.1. B) The Fourier Shell Correlation (FSC) for the final cryo-EM half maps. The global resolution was 2.2 Å according to the gold standard cut off at 0.143. C) The distribution of particle orientations in the final reconstruction with a minor preference for side views.

### 2.3.1.2 The positions of side chains and heme groups were less ambiguous in the higher resolution map

Ax qNOR Model building was performed by Dr Svetlana Antonyuk and the improved resolution allowed for the placement of all sidechains within the cryo-EM map (Figure 2.4A-B). The overall fold of 18  $\alpha$ -helices packed together into each monomer was the same as the previous Ax qNOR structure. However, there was also cryo-EM density corresponding to several hundred water molecules, a number of lipids, detergent molecules and an electron donor, none of which was visible in the previous map (Figure 2.4B). The Ax qNOR active site, formed of heme *b* and the binuclear Fe<sub>b</sub>-heme *b*<sub>3</sub> centres, is located within the core of each monomer where the map has the highest resolution (Figure 2.4C and Figure 2.4D).



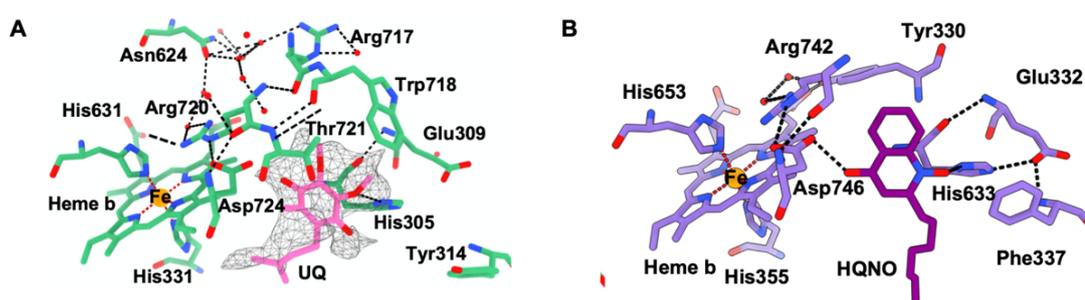
**Figure 2.4** A higher resolution cryo-EM map of Ax qNOR was generated using an upgraded cryo-EM set up. A) The cryo-EM map with 2.2 Å resolution. The two monomers are highlighted in green and blue. B) Protein model built into the cryo-EM map. Key regions are highlighted on one monomer. C) The catalytic core formed of heme *b* and heme *b*<sub>3</sub> linked by a calcium ion. Cryo-EM density set at a threshold of 0.034 except for that corresponding to water molecules and Glu 490 which is at 0.020. D) Local resolution of the cryo-EM map.

### 2.3.1.3 First structural evidence for the binding site of the ubiquinol electron donor by cryo-EM

A weak cryo-EM density corresponding to a non-protein molecule was visible away from the transmembrane domain close to Asp724 and His305 (Figure 2.5A). In the structure of qNOR from a related bacterial species in complex with quinol analogue 2-heptyl hydroxyquinoline *N*-oxide (HQNO), a weak density corresponding to HQNO was observed in a similar position<sup>118</sup> (Figure 2.5B). The density in this new map of Ax qNOR likely corresponds to an endogenous electron donor that was carried through

from the *E. coli* in which the protein was expressed, since no such molecule or analogue was added during sample preparation. Therefore, the native bacterial electron donor ubiquinol was modelled in that position.

The structure was then assessed for the feasibility of electron donation from this site to heme *b*. The ubiquinol is hydrogen bonded to His305 and Asp724, which are well poised to become protonated upon ubiquinol oxidation and to facilitate proton coupled electron transfer. Arg720 is hydrogen bonded to the Asp724 and makes a stacking interaction with heme *b*, positioning the heme group favourably for electron transfer and allowing for a potential route of electron transfer. In addition, Arg717, Arg720 and Asn624 create a water filled cavity close to heme *b*, making water-mediated electron transfer feasible<sup>153,154</sup>.



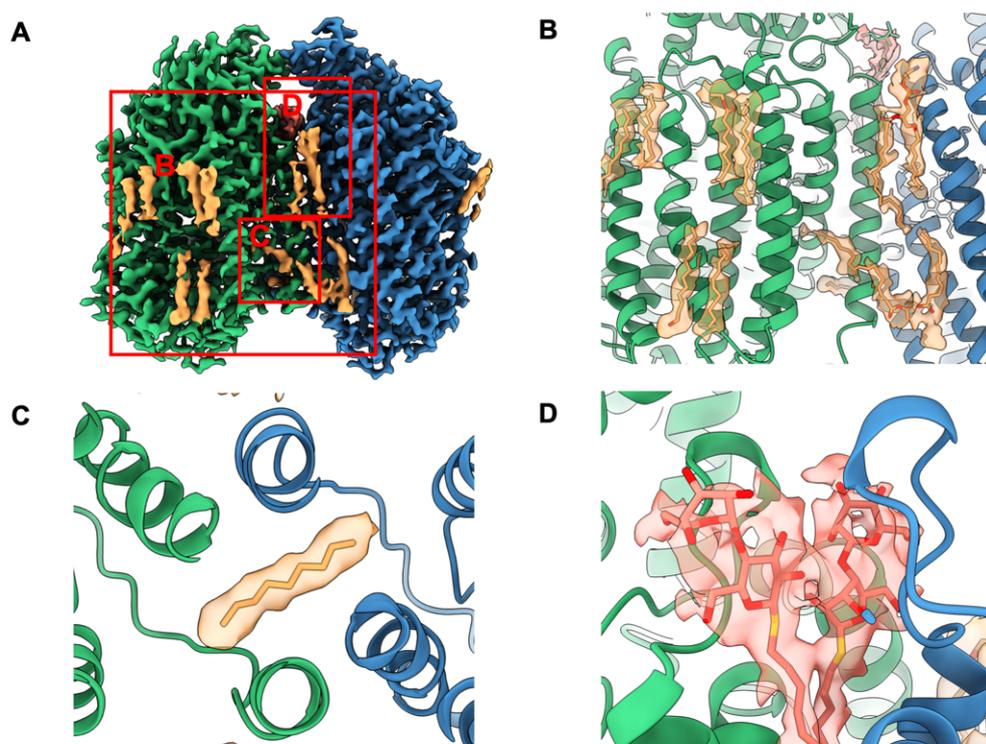
**Figure 2.5 Electron donor ubiquinol was present in the cryoEM map of Ax qNOR.** A) Density corresponding to ubiquinol visible in a pocket adjacent to heme *b*. Protein and ubiquinol are shown as green and pink sticks, respectively. Map threshold set to 0.01. B) The quinol analogue 2-heptyl hydroxyquinoline *N*-oxide (HQNO) (dark purple) found at the same site in qNOR from the related species *Neisseria meningitidis* (*Nm*) (light purple).

### 2.3.1.4 Lipid molecules were present in the cryo-EM map

Several areas of additional long, tubular density were observed in the cryo-EM map, which are consistent with the shape of bound lipids (Figure 2.6A-C). This extra density is located within the qNOR transmembrane region with the lipid tails pointing towards the middle. No additional lipids were re-introduced during protein purification, so they were carried through the extraction and purification process. They were modelled in as the native *E. coli* lipid Lauryl Oleyl Phosphatidyl-ethanolamine (LOP), but their exact identity remains unknown. As is common with detergents and lipids, the density is poorly defined in the tail regions, likely indicative of lipid flexibility. Where the lipid density is weak, lipids have not been modelled in. The lipids make interactions with the exterior of the transmembrane helices but do not make any specific interactions with the substrate tunnels or the bound cofactors. The presence of the lipids within the dimeric structure further reinforces that the integrity of the functional dimer is well preserved in this cryo-EM structure.

Further cryo-EM density at the dimeric interface is not consistent with the two-tailed shape expected for a lipid. Instead, the strong head density and single tail is more

typical of a bound DDM detergent molecule that was used during Ax qNOR purification (Figure 2.6D). Its location at the dimeric interface suggests that the detergent DTM replaced a bound lipid molecule during protein extraction and that a hydrophobic agent at the dimeric interface, be that lipid or detergent, is key for dimeric stability.



**Figure 2.6 Lipids and detergents within the Ax qNOR cryo-EM map.** A) Overview of additional non-protein density in the AxqNOR map. Cryo-EM density for chain A, chain B, LOP lipid and DTM density are shown as green, blue, orange and salmon surfaces respectively. Density for protein is contoured to a threshold of 0.030 and density for detergent and lipids is contoured to a threshold of 0.008. This scheme is used throughout the figure. B) Modelled LOP lipids found at the protein exterior and dimer interface. C) Modelled fragment of LOP crossing the dimer interface and protruding into product release channel. D) DTM detergent at the dimer interface.

### 2.3.1.5 Only three hours of Ax qNOR cryo-EM data was required to generate a 3.0 Å 3D reconstruction

Collecting Ax qNOR data on the upgraded Falcon 4 detector led to a higher resolution map with new features. Following this, I evaluated whether this improvement was caused by the increased data quality or quantity compared to the previous reconstruction from the K2 Summit. For example, 1.7 times more micrographs were collected on the Falcon 4 and 4.3 times more particles were picked on the newer micrographs (Table 2.1). This means that of the micrographs collected, 2.5 times more particles were picked per micrograph. Since both grids were produced at the same time, it could be assumed that the concentration of particles was similar across both grids. So, why were more particles picked on the newer micrographs?

A possible explanation is that the new picking algorithms might be better at identifying particles within the ice, particularly those with weaker SNRs. For example, whereas the K2 dataset was picked using the Relion template picker which requires pre-defined views of the protein, the new dataset was picked using the machine learning-based picker crYOLO, which is trained on a range of cryo-EM datasets and uses more general information on how particles look within the ice. It is also possible that having a more sensitive detector and a higher SNR through due to the Selectris energy filter means that data can be collected on thicker ice where there tends to be a higher concentration of particles. Particles in thicker areas of the ice with lower signal might have become more visible in the micrographs due to the increased contrast from the Selectris. In order to evaluate whether more particles were picked due to a higher concentration of particles on the grid or whether the newer picking algorithm identified more particles, it would have been helpful to re-process the previous dataset collected on the K2 using more recent programs for processing like crYOLO. However, the data was not available to do this.

That being said, even though 4.3x more particles were picked in total, there were 7.2x more particles in the final reconstruction, meaning picked particles were 1.7x more likely to end up in the final reconstruction. The algorithms for 3D refinement did not change significantly between Relion versions 3.0 and 3.1 used for the two datasets, so this improvement was likely to have been inherent in the data, due to it being better quality.

**Table 2.1 Differences in Ax qNOR cryo-EM data collections to evaluate effects from increased data quality and quantity**

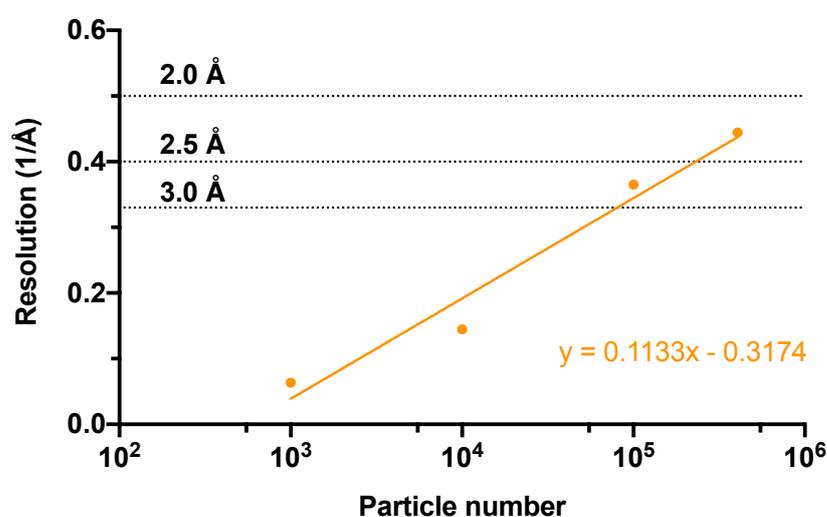
	<b>Previous dataset</b>	<b>New dataset</b>	<b>Increase</b>	<b>Increase adjusted from previous step</b>
<b>Detector</b>	K2 Summit	Falcon 4 + Selectris	-	-
<b>Movies collected</b>	3,213	5,466	1.7x	-
<b>Particles picked</b>	700,000	2,999,921	4.3x	2.5x
<b>Final particles in reconstruction</b>	56,134	404,950	7.2x	1.7x
<b>Resolution (Å)</b>	3.9	2.2	-	-

The quality of the new data was also assessed using a ResLog plot, where the resolution of reconstructions from different numbers of random particles are plotted <sup>155</sup> (Figure 2.7). The points were fitted with a linear regression and from this it was

calculated that 56,134 random particles (the number of particles that generated the previous reconstruction) from the new dataset would have generated a 3.26 Å reconstruction and a 3.9 Å resolution reconstruction (the final resolution of the previous map) would have been generated from 26,516 particles. Therefore, the new Ax qNOR cryo-EM data was higher quality because fewer particles were required to generate higher resolution maps compared to the previous data.

Not only was the data quality better, but the new detectors allow for data to be collected faster. The average collection rate for the Falcon 4 with AFIS was estimated by the Leeds EM facility to be ~400 movies per hour, whereas for the K2 it was ~300 movies per hour. This is partly due lower exposure times, with the K2 exposing for ~12 seconds/movie and the Falcon 4 exposing for ~6 seconds/movie.

The ResLog plot also identified that a 3.0 Å map would have been generated from 85k particles. With an average of 74.1 particles/micrograph for the new dataset, a 3.0 Å map could have theoretically been generated from 1,147 micrographs in only 2 hours and 56 minutes. This indicates that eight 3.0 Å ligand-bound Ax QNOR structures could potentially be produced from 24 hours of data collection, altering how cryo-EM can impact drug discovery projects.



**Figure 2.7 A ResLog plot to assess the quality of Ax qNOR cryo-EM data collected on the Falcon 4 detector.** The final masked resolution was plotted from the refinement of random particle subsets (N=3, except for the highest particle number which was N=1).

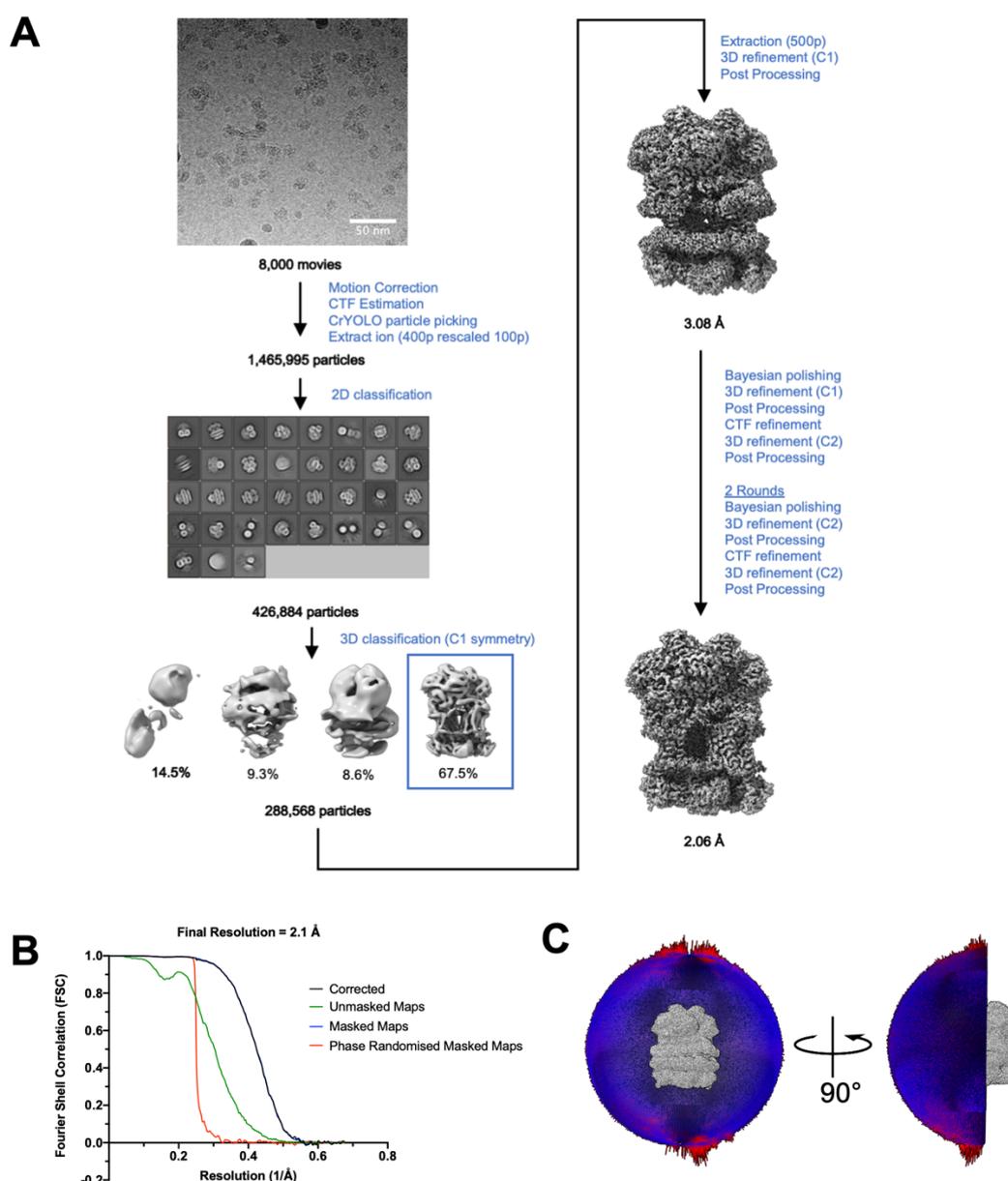
### 2.3.2 The second protein complex bovine cytochrome bc1 was determined to 2.1 Å resolution in an apo state by cryo-EM

#### 2.3.2.1 Collecting on an upgraded detector improved the resolution by >2Å

Due to the issues with determining a *Pf* bc1 structure, we focused on improving the resolution of the bovine structure so that small molecule binding could be visualised in

more detail. We applied the same method that was successful for Ax qNOR, whereby we imaged grids that were produced at the same time as those that generated the previous bovine bc1 structure, but using an upgraded Falcon 4 (now Falcon 4i) detector with a Selectris energy filter (Appendix Table 5.2).

Movies of bc1 were processed in Relion (Figure 2.8A). Picked particles were first classified in 2D and 3D and after 3 rounds of Bayesian polishing and CTF refinement, the final map had a global resolution of 2.1 Å, which was even higher than the resolution for the Ax qNOR map (Figure 2.8B). The angular distribution was consistent across the reconstruction with a slight over-representation of top/bottom views.



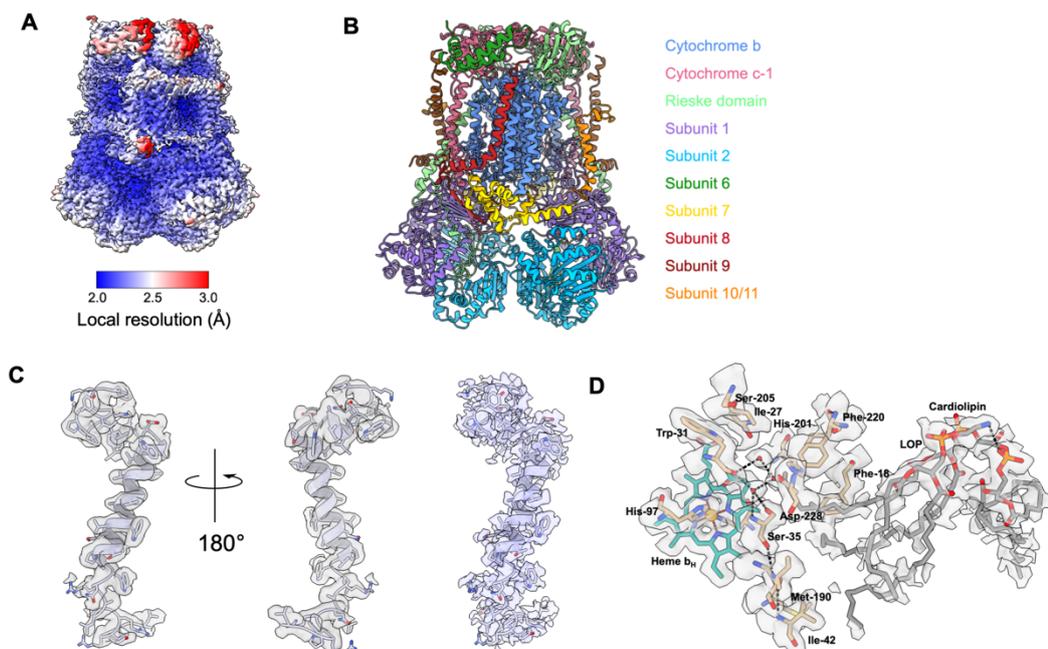
**Figure 2.8 A cryo-EM reconstruction of apo bovine cytochrome bc1 to 2.1 Å resolution.** Image processing workflow performed in Relion 4.0. B) FSC curves for the final map. C) Angular distributions of particles within the final reconstruction.

### 2.3.2.2 New features in the cryo-EM density included protein, waters and lipids

The resolution across the cryo-EM map was consistently high, meaning all eleven subunits of cytochrome bc<sub>1</sub> could be modelled in by Dr Svetlana Antonyuk (Figure 2.9A and Figure 2.9B). This included subunit 11, which was not resolved in either the cryo-EM or X-ray crystallography structures of bc<sub>1</sub> published by the Muench and Hasnain groups<sup>108</sup> (Figure 2.9C). The previous lack of subunit 11 was presumed to be caused by dissociation during protein isolation. However, its location within this cryo-EM map, which was reconstructed from grids produced at the same time as those that led to the previous cryo-EM structure, indicates that subunit 11 was present before but the signal was not strong enough to visualize it. The weak signal at this subunit in the higher resolution map was over sharpened compared to the rest of the map (Figure 2.9C). This indicates subunit 11 is more flexible and the high spatial frequencies required to visualise it were simply not present in the previous map. Subunit 11 was visible in the first complete bovine bc<sub>1</sub> crystal structure<sup>125</sup>, suggesting it was locked into position by the crystal contacts.

The Q<sub>i</sub> site, where ubiquinone is reduced to ubiquinol, is one of the key inhibitor binding sites. The resolution is particularly high in this region, with strong density for the protein as well as for new features like water molecules and lipids (Figure 2.9D). Three water molecules form a cluster in the Q<sub>i</sub> site, forming hydrogen bonds with Trp 31, Asp 228 and Ser 35. It is possible they are involved in water-mediated electron transport from heme b<sub>H</sub> to ubiquinone. Although their corresponding map is weaker than that of the surrounding protein, the lipids are clearly visible at the entrance to the pocket. One molecule has four fatty acid tails, distinctive of cardiolipin, and the other has two tails and has been modelled in as the standard lipid LOP. Since lipids were not supplemented during purification, the exact identify of the latter endogenous lipid remains unknown. The density within the Q<sub>i</sub> site corresponds to one of the cardiolipin tails, rather than an electron donor. It suggests that cardiolipin plays a regulatory role and is displaced by a UQ molecule during catalysis.

In summary, we greatly improved the resolution of the bovine cytochrome bc<sub>1</sub> apo cryo-EM map by collecting on an upgraded F4i detector. This new map contains not only the entire bc<sub>1</sub> protein, but also newly identified water molecules and lipid molecules, providing new information on how these ligands affect bc<sub>1</sub> function. The improved resolution allows for better placing of protein side chains, making docking models more accurate. Furthermore, the high resolution at the Q<sub>i</sub> site meant that density for small molecules located at the same site in future ligand-bound structures would likely also be high resolution and unambiguous in their positioning.



**Figure 2.9 New features visible in the bovine cytochrome bc1 apo cryo-EM map.** Local map resolution. B) Model built from the map, coloured by chain. C) The 6.4 kDa subunit 11, which was not visible in the previous cryo-EM reconstruction. Non-sharpened map shown in grey and globally sharpened map shown in pale purple. D) The Qi site in the apo cryo-EM map. Cytochrome bc1 protein shown as sand coloured sticks, heme  $b_H$  highlighted as blue sticks. The lipids lauryl oleyl phosphatidyl-ethanolamine (LOP) and cardiolipin are shown as grey sticks. Waters are shown as red spheres.

### 2.3.2.3 Only 50 minutes of bovine bc1 cryo-EM data was required to generate a 3.0 Å reconstruction

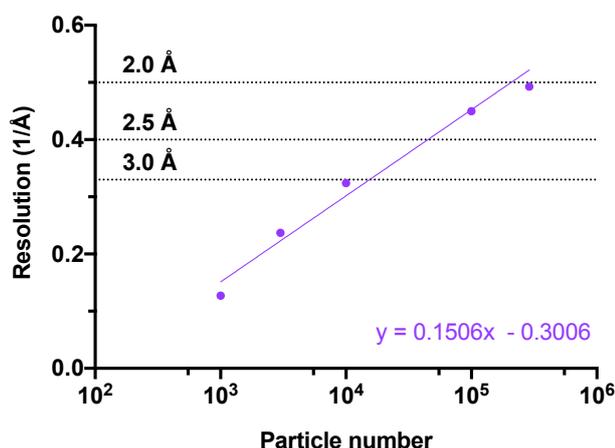
As with Ax qNOR, the new cryo-EM reconstruction was used to compare the quality of the bc1 data collected previously on the K2 Summit and recently on the Falcon 4i. Considering that 2.5 times more movies were collected for the newer dataset, 5.6 times more particles were picked on the micrographs, which means particles were 2.2 times more likely to be picked on each micrograph. Again, since the original bc1 data could not be re-processed, it cannot be verified whether the particle concentration was higher on the recently imaged grid, whether the newer picking algorithms were better at picking or whether the increased SNR meant more particles were visible in the micrographs. Considering the 5.6 times higher number of picked particles, there were only 5.0 times more particles in the final reconstruction, indicating some of the extra picked particles were likely junk particles. However, considering only 2.5 times more movies were collected, a 5.0 times higher number of particles in the final reconstruction suggests bc1 particles were more likely to end up in the final reconstruction when collected on the Falcon 4i compared to the K2 Summit, which was also seen for the new Ax qNOR reconstruction.

**Table 2.2 Comparison of data collections generating apo bc1 data**

<b>Metric</b>	<b>Previous dataset</b>	<b>New dataset</b>	<b>Increase</b>	<b>Increase adjusted for previous step</b>
Detector	K2 Summit (Gatan)	Falcon 4i (Thermo fisher scientific)	-	-
Movies collected	3,256	8,000	2.5x	-
Particles picked	260,201	1,465,995	5.6x	2.2x
Particles in final reconstruction	57,571	288,568	5.0x	0.9x
Resolution (Å)	4.4	2.1	-	-

A ResLog plot was used as a measure data quality and to compare the amount of data required to obtain reconstructions with specific resolutions (Figure 2.10). For example, 57,571 particles (the number of particles averaged in the final reconstruction of K2 data) collected on the F4i would have led to a 2.4 Å resolution reconstruction, a large improvement on the K2 Summit. Whereas, a 4.4 Å reconstruction (the same resolution as the final reconstruction from the K2 data) would have been generated from only 3,200 particles, meaning 18 times fewer particles were required to obtain the same resolution as on the K2.

A 3.0 Å reconstruction would have been generated from only 16,196 particles. For the F4i data, there were 36.1 good particles per micrograph, so only 449 micrographs are needed to generate the 3.0 Å reconstruction. A data collection rate of ~550 movies/hour was estimated by the Leeds EM facility for the F4i using AFIS on grids with 1.2/1.3 spacing. This means that the 3.0 Å reconstruction could have been generated from only 49 minutes of data collection, so 29 ligand-bound structures at 3.0 Å resolution could be determined from a 24 hour data collection.



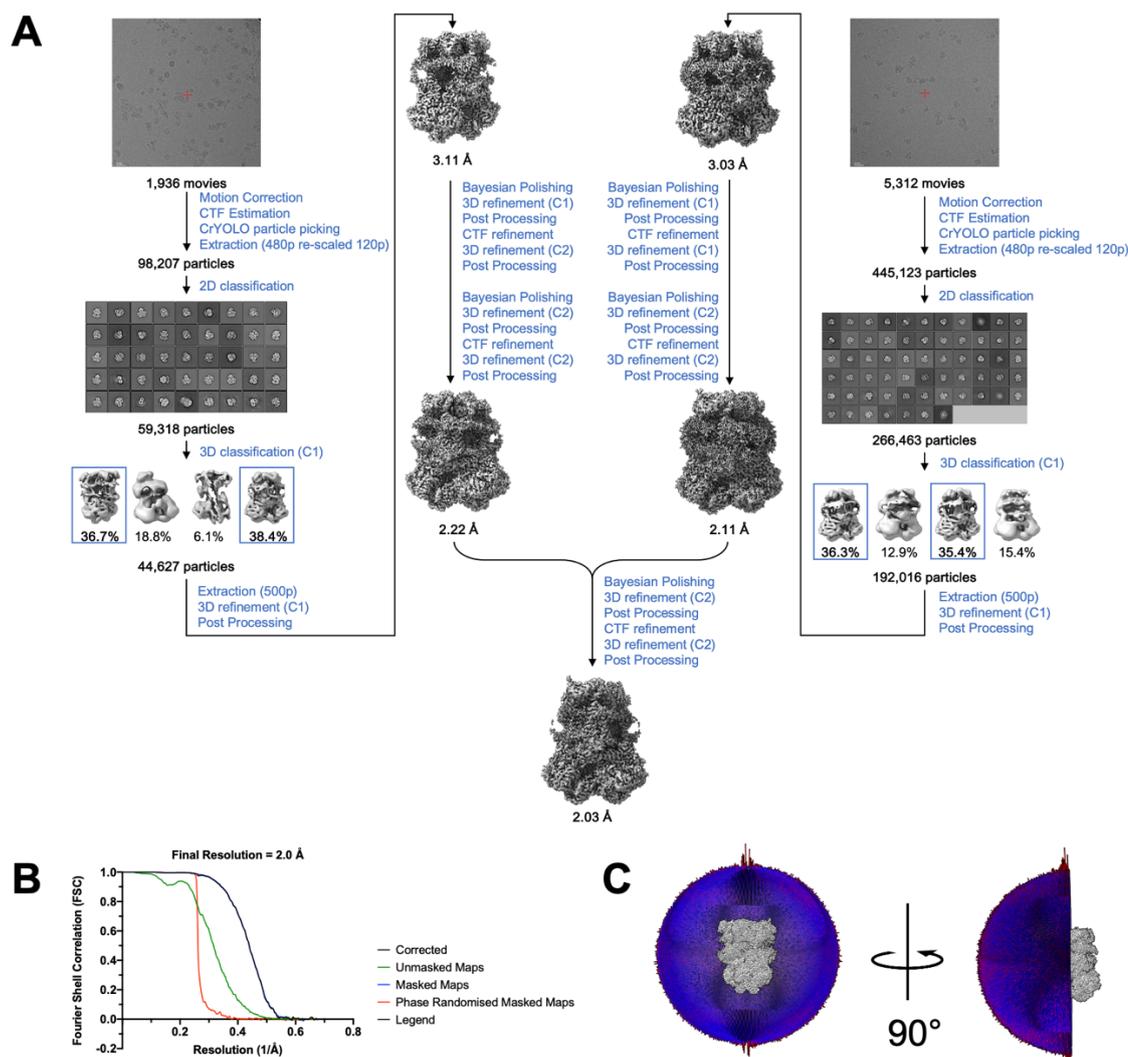
**Figure 2.10 A ResLog plot to measure the assess the quality of cryo-EM data collected of bovine cytochrome bc1 in the apo state.** The final masked resolution was plotted from the refinement of random particle subsets. (N=3, except for the highest particle number which was N=1).

### **2.3.3 Two small molecule-bound bovine cytochrome bc1 structures were determined to 2.0 Å resolution**

The vast improvement in the resolution of the bovine Cyt bc1 apo map compared with the previous reconstruction meant there would be much more detail about inhibitor binding in ligand-bound maps. Rather than simply identifying the location of the small molecules and docking into an “envelope” with some ambiguity, as was done for the previous ligand-bound structures at 4.1 Å<sup>108</sup>, at 2.0 Å resolution, it would be possible to observe the ligand interactions with bc1 residues directly. To obtain ligand-bound structures, fresh bovine bc1 was purified by Dr Nattapon Pinthong, together we mixed the protein with two small molecules CK267 and SG114 and froze the mixtures on cryo grids. Grid screening, data collection and image processing was done by myself.

#### **2.3.3.1 A cryo-EM structure identified CK267 in both the Q<sub>I</sub> and Q<sub>O</sub> sites**

For bc1 with CK267, data was merged from two data collections to generate a final reconstruction with 2.0 Å resolution (Figure 2.11A and Figure 2.11B, Appendix Table 5.3). Similar to the apo structure, the final reconstruction had a minor overrepresentation of top/bottom views (Figure 2.11C).



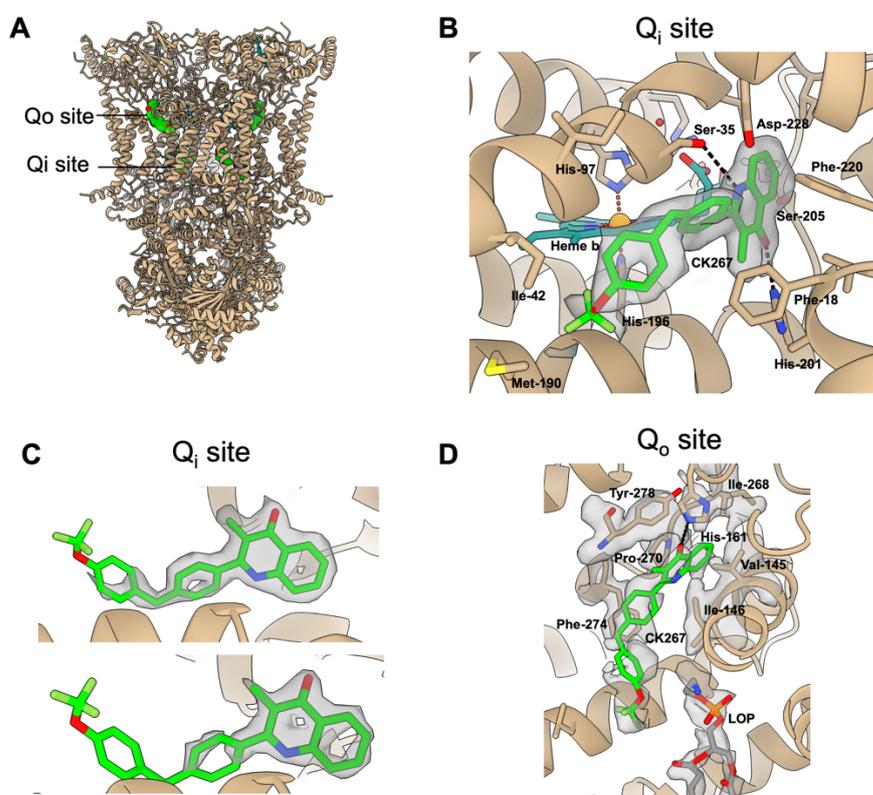
**Figure 2.11 Determining a structure of bovine bc1 bound to small molecule CK-267 by cryo-EM to 2.0 Å resolution.** A) Image processing workflow performed in Relion 4.0. B) The FSC curve for the final cryo-EM half maps. C) The distribution of particle orientations in the final reconstruction.

Within the cryo-EM map generated for dimeric bc1, CK267 is present in both the Qi site, where the density is strong, and in the Qo site, where the density is weaker (Figure 2.12A). The Qi site is the same location in which the small molecules SRC0911 and GSK932121 were found in the previous cryo-EM and X-ray crystallography studies<sup>108</sup> (Figure 2.12B). This makes sense as both SRC0911 and CK267 both contain quinolone scaffolds.

The small molecule density is strikingly better resolved. Holes within cryo-EM density for aromatic rings is a sign of high resolution and there is hole within all three of CK267's rings (Figure 2.12C). The high resolution means the precise orientation of CK267 within the map is unambiguous, whereas this was not the case for the previous structures. The quinolone double ring sits towards Phe-220 and Ser-205, with Phe-220 making a pi-pi stacking interaction with the carbon ring. In the adjacent nitrogen-containing ring, the nitrogen makes a hydrogen bond with Ser-35 and His-201 makes a hydrogen bond with the ring's carbonyl group. Moving along the small molecule, the

central ring stacks with Phe-18. The remainder of the molecule does not appear to make any strong interactions like hydrogen bonds or stacking. The density is slightly weaker around the trifluoro moiety, likely because there is some rotation around the ether bond and general flexibility. This structure could be used to guide the development of more active compounds. For example, there is room for CK267 to grow into the Qi site and make an interaction with Ser-205.

The slightly weaker density at the Qo site also fits CK267 well (Figure 2.12D). CK267 sits in the Qo site next to a LOP lipid molecule, with the nitrogen in the quinolone ring making a H bond with His-161 and Phe-274 stacking with the central benzene ring. It is possible that the trifluoro moiety interacts with the head group of the adjacent lipid molecule, but the density is particularly weak in this area.

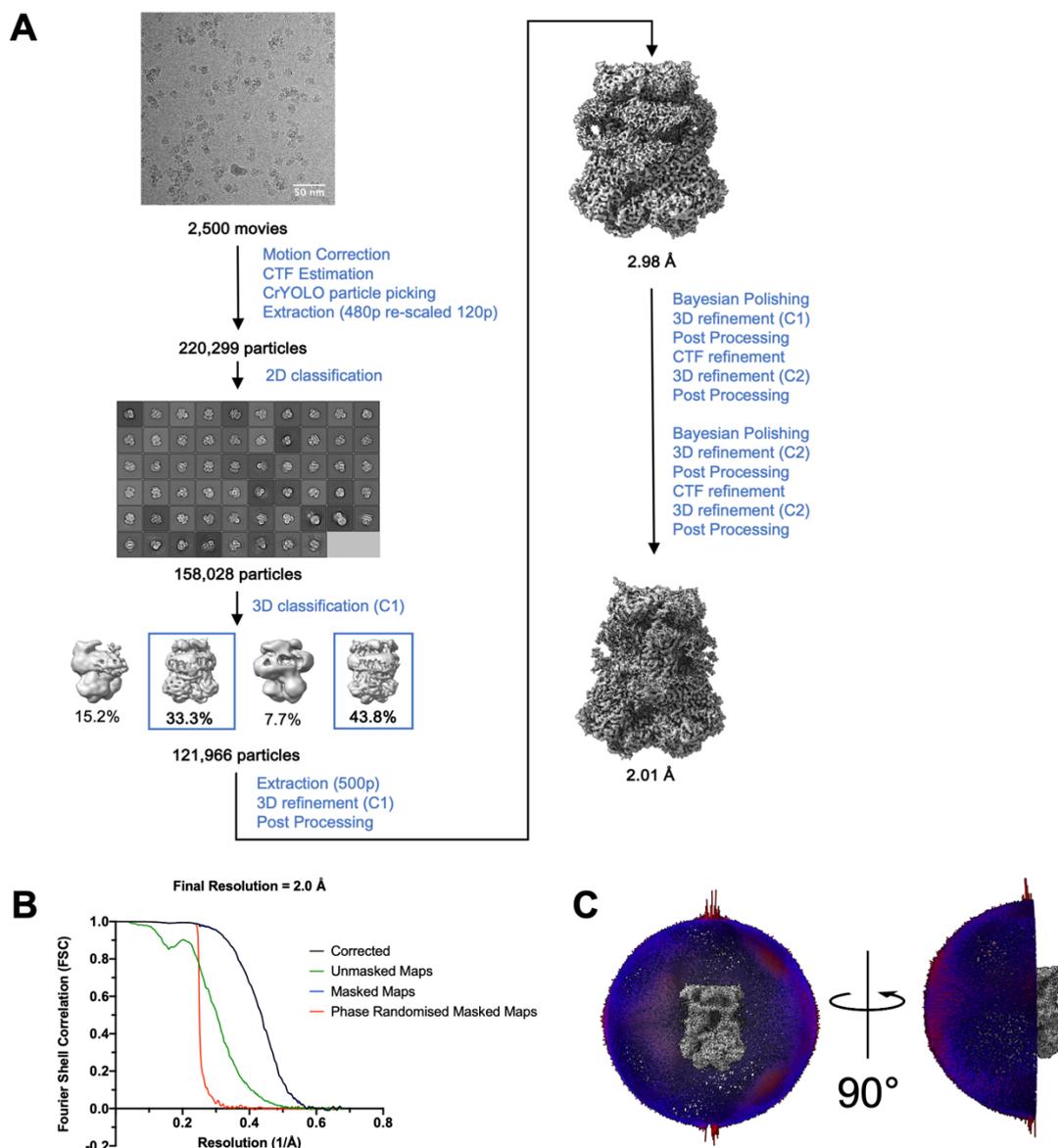


**Figure 2.12 The small molecule CK-267 bound to bovine cytochrome bc1 in a cryo-EM map.** A) Atomic model showing the CK267 binding sites. Cytochrome bc1 shown as beige ribbons and CK267 shown as green spheres. B) CK267 in the Q<sub>i</sub> site. Cytochrome bc1 shown as beige ribbons and CK267 shown as green sticks. Key bc1 residues are highlighted as beige sticks. The cryo-EM density for CK267 is shown as a transparent grey surface at a threshold of 0.0081. This scheme is used throughout the figure. C) The cryo-EM density for CK267 at thresholds of 0.018 (top) and 0.0216 (bottom), showing holes within the aromatic rings. D) CK267 in the Q<sub>o</sub> site. The cryo-EM density for CK267 is set at a threshold of 0.0065. LOP lipid is shown as grey sticks.

### 2.3.3.2 A cryo-EM structure identified CK267 in the Q<sub>i</sub> site

For bc1 with SG114, there was little difference between particle concentration and dispersion on gold- or carbon-coated grids, so images were collected on carbon-coated

grids. Movies were collected and processed in Relion using the same workflow and parameters that were successful for the apo and CK267-bound structure (Figure 2.13A, Appendix Table 5.4). The final reconstruction had a global resolution of 2.0 Å, which was ~2.1 Å better than the SCR0911- and GSK932121-bound cryo-EM maps published previously<sup>108</sup> (Figure 2.13B). Similar to the apo structure, the angular distribution was consistent throughout the reconstruction but with a slight preference for top/bottom views (Figure 2.13C).



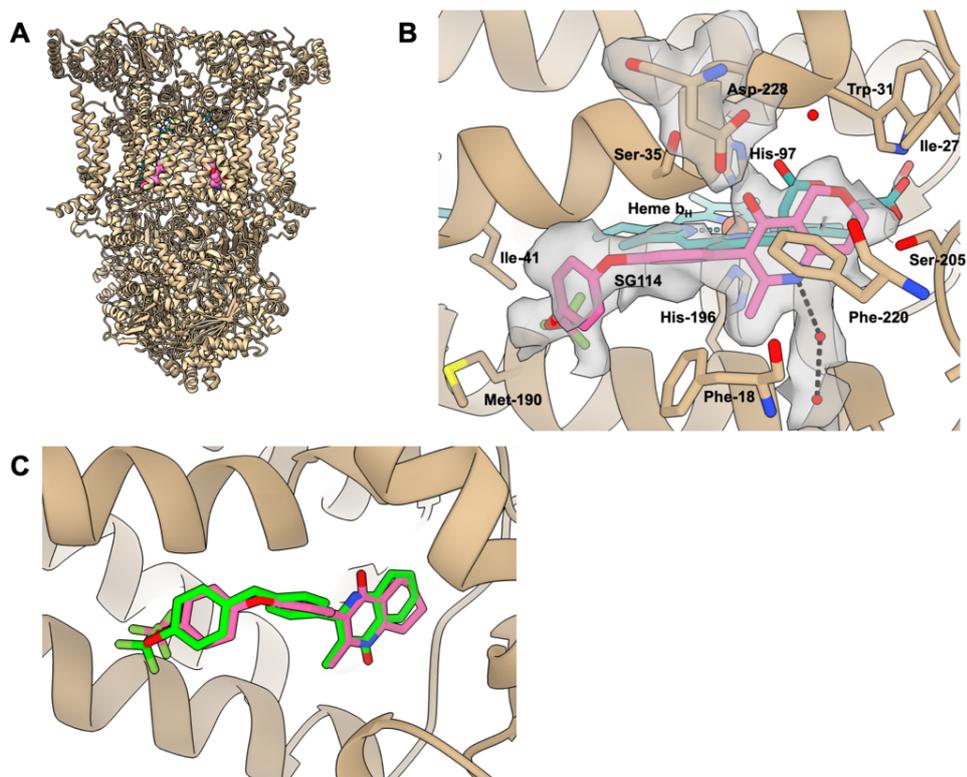
**Figure 2.13** Generating a structure of bovine cytochrome bc1 bound to the small molecule SG114.

A) Image processing workflow performed in Relion 4.0. B) The FSC curve for the final cryo-EM half maps.

C) The distribution of particle orientations in the final reconstruction.

There is extra density in the bc1 dimer map for two SG114 molecules (Figure 2.14A). The molecules are found in the Qi site in the same location as CK267 but making different interactions. The nitrogen of the quinolone scaffold makes a hydrogen bond

with a water molecule rather than Ser-35 and the carbonyl oxygen makes a hydrogen bond with Asp-228 (Figure 2.14B). Overlaying the positions of SG114 and CK267 shows they sit in exactly the same site but the quinolone rings are rotated, with the carbonyl pointing down towards His-201 in CK267 and up towards Asp-228 in SG114. Such small differences in orientation are only visible in high resolution maps like those represented here and indicates how significant the advances in cryo-EM hardware are.



**Figure 2.14 SG114 bound to bovine cytochrome bc1 map in a cryo-EM map.** A) Model of cytochrome bc1 (beige ribbons) with SG114 (pink spheres). B) The bc1 Q<sub>i</sub> site with unambiguous density for SG114. Cytochrome bc1 shown as beige ribbons with key residues highlighted as sticks. SG114 shown as pink sticks. Water molecules shown as red spheres with hydrogen bonds shown as dashed black lines. Cryo-EM density at a threshold of 0.005. C) Overlaid positions of SG114 (pink) and CK267 (green) in the Q<sub>i</sub> site.

## 2.4 Discussion

The objective of this chapter was to investigate the impact of advances in cryo-EM, such as new detectors and energy filters, on generating structures to use in drug discovery. Two metalloenzyme drug targets were used as test subjects for this by collecting cryo-EM data on upgraded microscope set ups and processing them into 3D reconstructions. The two proteins were Ax qNOR and Cyt bc1 from either the malaria parasite *Pf* or the bovine homolog. Both Ax qNOR and bovine cytochrome bc1 were solved to ~4 Å.

For Ax qNOR, cryoEM data collected on a Falcon 4 with a Selectris energy filter was processed to a map at 2.2 Å resolution. This was a major improvement on the previous resolution of 3.9 Å and allowed the visualisation of many new features within the map. Side chains could be better placed within the map, increasing the confidence in the

protein model generated from this map, which is important for downstream SBVS. There were also hundreds of waters visible in the new map, highlighting their role in water-mediated proton and electron transport.

We also visualised the *E. coli* electron donor ubiquinol bound to Ax qNOR, meaning the two were co-purified. The presence of ubiquinol highlights that the conditions used for Ax qNOR purification, like the DDM/DTM detergents included in buffers, were gentle enough to maintain qNOR-ubiquinol binding and supports the idea that Ax qNOR is in a native conformation in this map. Visualising the electron donor allowed us to identify the route of electrons from its binding site to the catalytic core. Since Ax qNOR is a potential antibiotic target, the electron donor binding site is a promising site to target with small molecule inhibitors. These could be generated using a LBVS workflow to identify compounds similar to ubiquinol, or using a SBVS workflow to identify more unique and specific inhibitors. Thanks to the high resolution at the site, this structure can provide residue-level information about electron donor binding to guide VS.

Overall, the qNOR structure is a powerful demonstration of the combination of the Falcon 4 and Selectris energy filter. At ~180 kDa, qNOR is relatively small for cryo-EM structure determination, giving it a low SNR in micrographs, and the presence of a detergent micelle which surrounds much of the protein further reduces the SNR. Despite both of these challenges, the new cryo-EM data was processed into a 2.2 Å reconstruction using a standard image processing workflow without using complex processing techniques to boost the SNR.

In a similar process, we also used a previously prepared grid of bovine Cyt bc1 to generate a new map to 2.1 Å resolution. This again indicated that current cryo-EM set ups can improve the resolution of protein reconstructions by ~2 Å compared to previous setups. Not only were there waters and lipids present in the map, but the 6.4 kDa subunit 11 was visible for the first time by cryo-EM. It confirms that it had not dissociated during protein purification, as was suspected previously.

The presence of lipids in the Ax qNOR and bovine Cyt bc1 maps also indicates that the proteins are in native states that resemble the environment in the lipid bilayer. The DDM and DTM detergents used for purification were gentle enough to maintain the lipids throughout purification. That the cryo-EM density for the lipids is strong suggests that they are important for protein stability or function.

For both qNOR and Cyt bc1, the new cryo-EM data was better quality and was collected faster than the previous data, meaning that higher resolution structures can be generated more quickly thanks to the new detectors. It highlights how cryo-EM data collections are changing, moving to shorter sessions per structure, with the potential for multiple structures to be generated per day, which is particularly important for small molecule discovery. Although multigrid EPU for the Titan Krios microscopes automate

the collecting of multiple cryo-EM datasets, we need more downstream tools to organise and automate image processing pipelines.

Two 2.0 Å cryo-EM maps of bovine Cyt bc1 were determined bound to two small molecule inhibitors CK-267 and SG-114. The maps identified that while both compounds bound in the  $Q_i$  site as expected, only CK-267 also binds the  $Q_o$  site. This dual binding mode might make CK-267 a more promising inhibitor, as two routes of inhibition might be more challenging to evolve resistance to.

A cardiolipin molecule was present in the  $Q_i$  site in the apo Cyt bc1 structure, but in both the CK267- and SG114-bound structures, the compound had displaced the lipid. This displacing of lipids by small molecules has been seen previously for the TRPC5 ion channel, where the inhibitor Pico-145 displaces a lipid in its binding site<sup>156</sup>. For TRPC5, this displacement was thought to contribute to the inhibitory effect of Pico-145, but this cannot be stated for Cyt bc1. There are not many examples of small molecules having this effect, possibly due to the ability of cryo-EM to determine maps with the resolution to visualise lipids only occurring recently.

In summary, cryo-EM has demonstrated promise at determining therapeutically-relevant membrane protein structures to the resolution at which waters, side chains, lipids and small molecule ligands can be visualised. In only five years, much more structural information is visible in cryo-EM maps thanks to advances in the field. Following this, it is worth investigating how high resolution structures like those of Ax qNOR and Cyt bc1 can feed into the discovery of new small molecule inhibitors, beyond simply being searched in traditional structure-based virtual screening approaches.

# **Chapter 3 : Generating anti-viral small molecules using *in silico* approaches**

## 3.1 Introduction

Affimers are protein reagents that form a new class of biopharmaceuticals which circumvent some of the issues associated with monoclonal antibodies (mAbs). Affimers are ~10-times smaller than mAbs, providing better tissue penetration, they are more thermostable and they can be produced to high yields in *E. coli*<sup>83</sup>. However, neither mAbs nor Affimers can permeate the blood-brain barrier or the cell surface membrane, making them unsuitable for treating diseases of the central nervous system or for inhibiting intracellular targets, and they are not orally bioavailable, so they require administration by injection. Using the predicted or validated binding mode of the Affimer to develop a small molecule could circumvent these issues by providing a more suitable scaffold, whilst maintaining binding at the same site. A novel method of Affimer-guided small molecule discovery was tested using two sets of Affimers that inhibited viral proteins: an Affimer that inhibited the nucleoprotein (NP) from Crimean-Congo Haemorrhagic Fever Virus (CCHFV) as described in Alvarez-Rodriguez *et al.*<sup>86</sup>, and a set of Affimers that inhibited the hemagglutinin (HA) spike protein from influenza A virus (IAV) as described in Debski-Antoniak *et al.*<sup>157</sup>.

### 3.1.1 Crimean Congo Hemorrhagic Fever Virus

#### 3.1.1.1 CCHF Disease

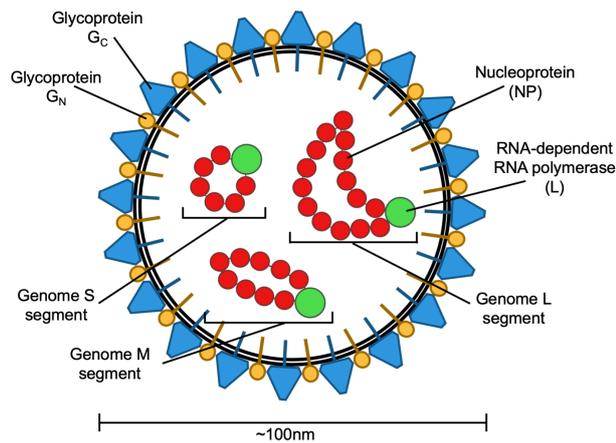
CCHFV is a member of the *Bunyavirales* order of enveloped viruses and the causative agent of Crimean-Congo Hemorrhagic Fever (CCHF) disease<sup>158,159</sup>. Symptoms of CCHF include mild fever, nausea, organ failure and haemorrhage and CCHF has a fatality rate estimated at 5-30% (Figure 3.1A)<sup>160,161</sup>. CCHFV is spread by arthropods, specifically *Hyalomma* ticks which bite humans and release the virus into the bloodstream<sup>162-164</sup>. The vast tick distribution across Africa, Asia, the Middle East and Southern and Eastern Europe leads to regular CCHFV outbreaks in these regions, making CCHFV the most wide-spread tick-borne virus and the only biosafety level 4 (BSL4) virus that is endemic in Europe (Figure 3.1B)<sup>165</sup>. It was predicted that higher temperatures caused by global warming would allow the *Hyalomma* ticks to survive and transmit CCHFV in new locations like the UK<sup>166</sup>. Although adult *Hyalomma* ticks were identified on birds in the UK and cases of CCHFV have been imported into the UK, as of yet, no ticks carrying CCHFV have been identified<sup>167-170</sup>. However, its severity and wide distribution caused the world health organisation (WHO) to rank CCHF in the top priority diseases for research and development.



**Figure 3.1: Crimean-Congo haemorrhagic fever virus (CCHFV) is a tick-borne virus that causes human symptoms of fever, organ failure and haemorrhage.** A) A patient suffering from a haemorrhage in the arm from CCHFV infection. Images reproduced from Bente *et al.*, which is distributed under the creative commons attribution license <sup>165</sup>. B) The global distribution of the *Hyalomma* species (spp.) of ticks, which are the primary vector for CCHFV. Image reprinted from Bonnet *et al.* with permission from Elsevier <sup>171</sup>.

### 3.1.1.2 Virus life cycle

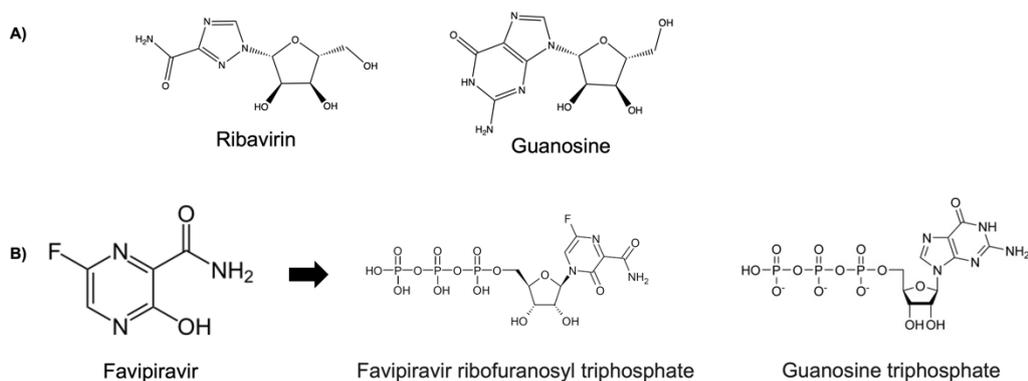
Like all viruses, CCHFV has a complex intracellular lifecycle <sup>172</sup>. Two glycoproteins  $G_c$  and  $G_n$  sit in the virus envelope, forming an external layer, and bind to the cell-surface receptors for CCHFV which include DC-SIGN found on macrophages and dendritic cells and Nucleolin found on macrophages and endothelial cells (Figure 3.2) <sup>173,174</sup>. Once bound, CCHFV enters the cell via receptor-mediated endocytosis, being enclosed by endosomes <sup>175</sup>. During endosomal maturation, changes in pH and ion concentration trigger the virus glycoproteins to mediate fusion of the endosomal and viral membranes <sup>176,177</sup>. This releases the viral ribonucleoprotein (RNP) complexes into the cytoplasm where viral replication occurs. The RNPs are formed of nucleoprotein (NP) wrapped around the tri-segmented negative-sense RNA genome, packaged with an RNA-dependent L RNA polymerase (RdRp or L). L uses the viral RNA as a template to make positive sense RNA which is translated by ribosomes into viral proteins. L is also involved in viral replication, which involves producing copies of the viral RNA segments and allows the production of more negative-sense genome segments that are packaged into RNPs and bundled into enveloped particles. The viruses are then released from the cell via the secretory pathway. Overall, all of CCHFV's lifecycle occurs in the cytoplasm and NP is only exposed from inside the envelope in the cytoplasm, making it more suited to targeting by small molecules than biopharmaceuticals.



**Figure 3.2 A schematic of a Crimean-Congo haemorrhagic fever virus virion structure.**

### 3.1.1.3 Treatments

Treatment for CCHFV infection is currently limited to Ribavirin, a broadly-acting guanosine analogue which has several proposed modes of action including direct inhibition of L polymerase, inhibition of enzymes involved in guanosine metabolism and incorporation into viral RNA products by L polymerase which introduces lethal mutations that result in defective virus particles (Figure 3.3A) <sup>178</sup>. It is licensed for treating chronic hepatitis C infection and recommended for the treatment of some RNA viruses like CCHFV <sup>179</sup>. Although in-vitro assays and clinical observational studies indicate ribavirin treatment inhibits CCHFV infection, a Cochrane review of all relevant studies relating to Ribavirin found insufficient evidence to demonstrate the efficacy of Ribavirin in treating CCHF <sup>180,181</sup>. Favipiravir is a guanosine analogue which has similar modes of action to ribavirin and causes lethal mutations in CCHFV, but it has not yet been approved for human use (Figure 3.3B) <sup>182</sup>. mAb therapy has also been trialled for treating CCHFV. mAbs against G<sub>C</sub> and G<sub>N</sub> protected mice from a lethal CCHFV challenge and three of these mAbs were broadly neutralising against glycoproteins from a range of CCHFV strains, preventing CCHFV from entering the required cells, but these mAbs are yet to be tested in humans <sup>183,184</sup>. Furthermore, there are no new CCHFV treatments in early stage clinical trials either. The potential for a CCHFV epidemic combined with the lack of effective therapies caused the World Health Organisation to rank CCHF in the top ten highest priority diseases for research and development <sup>185</sup>.



**Figure 3.3 Ribavirin and favipiravir are two small molecules used for the treatment of CCHFV. Both become incorporated into viral RNA causing lethal mutations.** A) Ribavirin is an analogue of guanosine recommended for treating CCHFV. B) Favipiravir is ribosylated and triphosphorylated to become an analogue of guanosine triphosphate. Its efficacy has not yet been measured in humans.

### 3.1.2 An Affimer inhibitor of CCHFV NP

The need for new reagents to treat and diagnose CCHFV was the basis for the development of an Affimer named NP-Affimer, which inhibited CCHFV NP<sup>86</sup>. NP-Affimer showed promise as a diagnostic but its inability to enter cells hinders its use as a therapeutic.

#### 3.1.2.1 Affimer development

NP-Affimer was identified using the established Affimer isolation method based on phage display, whereby a library of  $10^{10}$  bacteriophages, each presenting an Affimer scaffold with randomised hypervariable loop sequence on their outside, is screened for binding to an immobilised protein target<sup>186</sup>. Since the bacteriophages also encode the DNA sequence of their corresponding Affimer, the bacteriophage DNA is sequenced to identify the loop sequence which confers binding. Multiple panning rounds can be used to remove non-specific binders or to select for isoform-specific binders.

During NP-Affimer isolation, recombinantly expressed CCHFV NP was immobilised to capture Affimer binders<sup>86</sup>. Seven unique Affimers were identified following phage display, but only one termed NP-Affimer showed specific binding to CCHFV NP via a pull-down assay. Then, NP-Affimer was measured for binding to CCHFV NP, first using surface plasmon resonance (SPR) spectroscopy, in which it bound with a dissociation constant ( $K_D$ ) of 5.69 nM, close to the median  $K_D$  of 3 nM for mAbs with available target-bound structures<sup>187</sup>. Then, in circular dichroism-based thermal shift assays, NP-Affimer bound CCHFV NP and generated a 11.6 °C increase in its melting temperature ( $T_m$ ). It also inhibited the binding of CCHFV NP to RNA as well as the expression of CCHFV genes in a CCHFV mini-genome assay. In order to determine how NP-Affimer blocked NP, an NP:NP-Affimer crystal structure was solved to 2.84 Å, revealing that NP-Affimer likely blocks NP oligomerisation required for RNA packaging. Finally, it was

shown that NP-Affimer could be used in a lateral flow diagnostic device for detecting NP within a human serum sample <sup>86</sup>.

### **3.1.2.2 Uses in drug discovery**

Although the main purpose for NP-Affimer development was to increase the tools for CCHFV diagnostics, its inhibitory effect on CCHFV highlighted a potential use as a therapeutic Affimer. Inhibition of CCHFV via binding at the proposed NP oligomerisation site validated this area for further therapeutic targeting. However, as mentioned previously, a small molecule would be a more appropriate modality for targeting the NP. Combined with the fact that we are lacking effective small molecule inhibitors of CCHFV and there is a NP:NP-Affimer crystal structure to guide mimic development, NP-Affimer acts as a useful test case for a new Affimer-guided small molecule development method.

### **3.1.3 Influenza A virus**

Influenza A virus (IAV) is also an enveloped negative-sense RNA virus similar to CCHFV but from the *Orthomyxoviridae* family. Affimers were previously developed against the HA spike glycoprotein previously by Dr Oliver Debski-Antoniak and these represent a second contrasting test case for Affimer-guided small molecule development.

#### **3.1.3.1 Influenza disease**

IAV causes 'the flu', a respiratory illness with symptoms including a sudden high temperature, coughing, headache and fatigue <sup>188</sup>. The severity of flu infection can depend on how the strain emerged. The error-prone nature of the IAV polymerase leads to a gradual accumulation of mutations that causes antigenic drift, resulting in new seasonal strains. An estimated 300,000-500,000 deaths are attributed annually to these seasonal influenza epidemics <sup>189</sup>. The constant emergence of new strains necessitates annual vaccination programs and leads to the resistance of strains to anti-viral therapeutics. Antigenic shift is caused by the reassortment of RNPs from two diverse strains, creating a strain with a new combination of antigens. These antigenically unique strains can have pandemic potential due to the lack of circulating antibodies that recognise them. There have been six major influenza pandemics in the last 140 years, including the 1918 'Spanish Flu', which caused ~20 million deaths <sup>190</sup>.

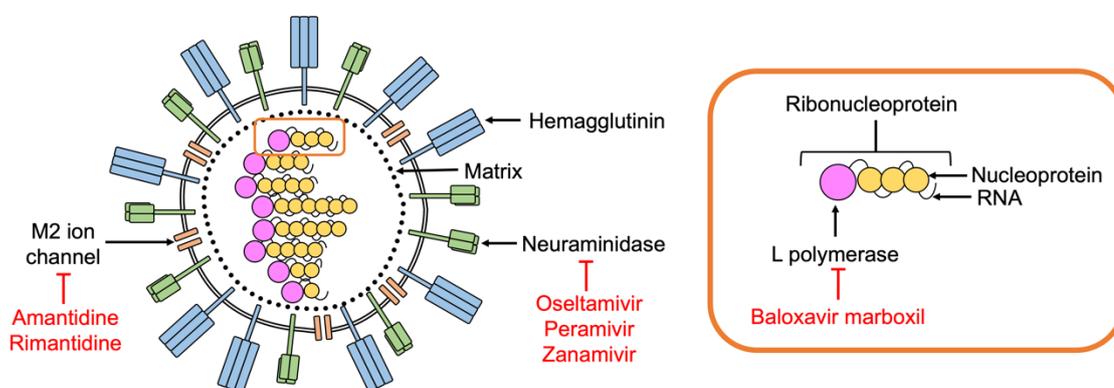
#### **3.1.3.2 IAV lifecycle**

The molecular biology and intracellular lifecycle of IAV is similar to CCHFV with a few key differences. For example, IAV contains eight RNPs rather than three, contains additional proteins like the M2 ion channel (Figure 3.4), and IAV replicates in the nucleus rather than in the cytoplasm. HA is one of two spike glycoproteins along with

neuraminidase (NA) that sit in the viral envelope. HA mediates cell entry by binding sialic acid sugars on receptors prior to receptor-mediated endocytosis, then using the acidity in the endosome to trigger conformational changes that induce fusion of the viral and endosomal membranes<sup>67</sup>. Furthermore, along with NA, HA is also involved in viral assembly and budding during egress<sup>191</sup>.

### 3.1.3.3 Influenza treatments

There are several small molecules approved for the treatment of IAV but all have problems with resistance (Figure 3.4). The first to be approved were the M2 inhibitors amantadine and rimantadine, but widespread resistance to the drugs means they are no longer recommended<sup>192</sup>. Currently, the first-line treatments are NA inhibitors such as zanamivir, peramivir and oseltamivir, which are effective at reducing illness severity and duration<sup>193</sup>. However, some circulating strains in the UK are starting to show resistance to these compounds<sup>194</sup>. Baloxavir marboxil is a polymerase inhibitor that has been approved for use in Japan and the USA but not yet in the UK. Unfortunately, a study found that 10% of patients with influenza that were treated with baloxavir marboxil developed a mutation associated with resistance<sup>195</sup>. As of yet, there are no approved inhibitors of HA but various compounds targeting HA have demonstrated inhibition of IAV infection in mice, indicating that HA is a promising target that could be investigated further<sup>196,197</sup>.



**Figure 3.4 Example of influenza a virus structural proteins and small molecule drugs that target them.** Structural proteins labelled in black and drugs highlighted in red.

### 3.1.4 Affimer inhibitors of hemagglutinin

Affimers represent a promising modality for targeting HA, since their size makes them efficient at blocking target-receptor interactions like HA-sialic acid, and the rapid pipeline for Affimer development could be applied to new, dangerous IAV strains. To test this, Affimers were isolated against HA from model IAV strain A/Aichi/68 (H3N2) by Dr Oliver Debski-Antoniak<sup>157</sup>.

### 3.1.4.1 Development

A monomer of the HA trimer from model IAV strain A/Aichi/68 (H3N2) was used in three rounds of phage display to identify Affimer binders<sup>157</sup>. Of the 34 unique Affimers that bound HA in an enzyme-linked immunosorbent assay (ELISA), seven showed inhibition of IAV infection in tissue cultured cells and two named A5 and A31 were characterised further. A5 and A31 showed inhibition of virus infection using a 50% tissue culture infectious dose (TCID<sub>50</sub>) assay, with inhibitory TCID<sub>50</sub> values of 1.96 nM and 0.44 nM. They also bound HA with high affinity, with K<sub>D</sub> values of 2.8 nM and 5.95 nM for A5 and A31, as measured by SPR. Both Affimers bound exclusively to the HA head domain (HA1) when tested in an ELISA, and they blocked receptor binding in a hemagglutination assay, indicating they bind at the receptor binding domain (RBD) on HA1<sup>157</sup>. However, when the work presented in this chapter began, there was no structure of either Affimer bound to HA, so residue-level information about how and where the Affimers bind HA was lacking.

### 3.1.4.2 Use in drug discovery

A5 and A31 could potentially become therapeutics themselves since HA is exposed on IAV at all times, unlike CCHFV NP. However, much of the IAV lifecycle occurs intracellularly, and since Affimers cannot permeate the cell membrane, A5 and A31 could only bind in the extracellular fluid. Developing small molecule mimics of A5 or A31 would allow for IAV inhibition to occur both intracellularly and extracellularly, increasing their opportunities for binding. However, without high resolution structural information to narrow down the A5 and A31 binding loops to the core site of interactions, it might be challenging to develop mimics under the 500 Da limit of drug-like small molecules.

### 3.1.5 Objectives

The aim of this chapter was to develop a ligand-based virtual screening (LBVS) method for identifying small molecule mimics of Affimer binding regions using NP-Affimer, A5 and A31 as templates. The high-resolution NP:NP-Affimer crystal structure was used to focus the search to the main region of interactions. However, generating high-resolution A5- and A31-bound HA structures was unsuccessful, so mimics were developed without an experimentally-derived Affimer-target structure. The small molecule mimics were screened for binding their targets *in silico* using a structure-based virtual screening (SBVS) program, highlighting how the complementary VS techniques can be employed together.

## 3.2 Methods

### 3.2.1 Overview of key methods

The methods employed in LBVS and SBVS are crucial to this chapter, so they are explained in more detail here. The workflows require various programs and the Schrödinger and OpenEye suites each contain all of the programs required<sup>198,199</sup>. Schrödinger uses a graphical user interface (GUI), with all programs run through Maestro, whereas OpenEye programs run through the command line<sup>200</sup>. Highlighted in Figure 3.5 are some of the programs used in this chapter.

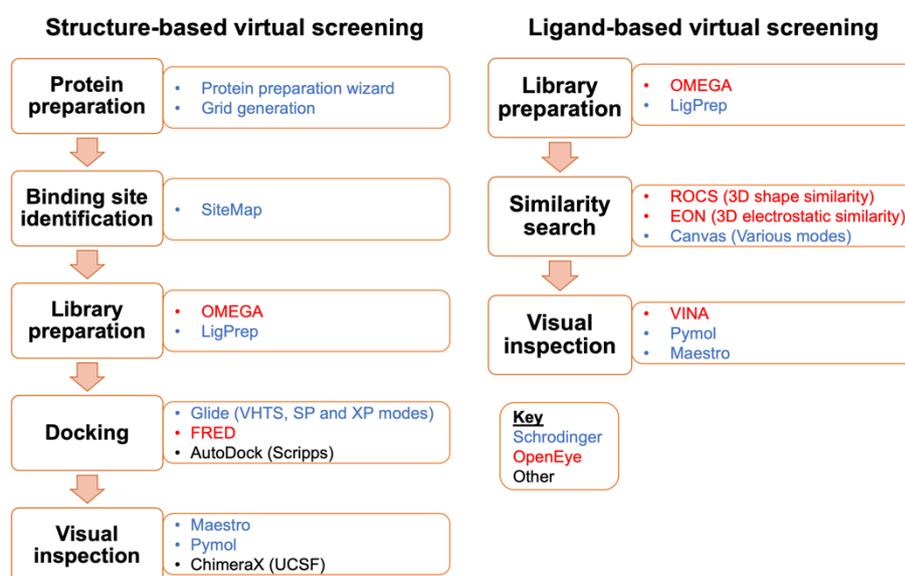


Figure 3.5 The structure-based and ligand-based virtual screening pipelines and example programs used for this chapter

#### 3.2.1.1 Protein preparation

SBVS requires an accurate, complete protein structure. This is generated during the protein preparation stage by completing any information missing from the structure file. Common issues include missing side chains, loops or hydrogen atoms due to weak density or low resolution, unspecified or ambiguous protonation states, and flipped side chains due to flexibility. The protein preparation wizard (Schrödinger) combines tools to address all of the aforementioned problems<sup>201,202</sup>.

The search region of the prepared protein is then extracted into a grid for compound screening. This leads to faster ligand screening by masking out unnecessary regions and allowing the prior calculation of the binding contributions of the target protein pocket. The protein is static within the grid, which is not representative of its true conformational landscape, but this is a trade-off for faster ligand screening. The search space can be widened by enlarging the grid, but this increases the screening time.

### 3.2.1.2 Binding site prediction

Predicted ligand binding sites can be useful when there is no known ligand site or the protein search space requires expansion. Various programs exist to predict binding sites within proteins such as SiteMap which searches through an entire protein structure and measures solvent exposure, degree of site enclosure, the balance of hydrophobic and hydrophilic character and the ability to donate or accept hydrogen bonds<sup>24,203,204</sup>. Regions which fulfil these criteria are identified, scored and ranked. The SiteMap scoring function was trained using a test set of protein-ligand complexes and the scoring was normalised so that the average score for the known ligand sites was 1.0<sup>204</sup>. Since >96% of the ligands in the test set had affinities of <1  $\mu\text{M}$ , a predicted site with a score of >1.0 is not just 'above average' but it is above average for sites that bind small molecules with high affinity.

### 3.2.1.3 Library preparation

Ligands exist in three dimensional conformations but commercial suppliers often supply their compound libraries as two dimensional files. Efficient and accurate prediction of their 3D conformations is essential for accurate virtual screening. Small molecules can be flexible and occupy multiple conformations in solution. Representing all of these conformations would require many discrete conformers to be generated for each compound. However, the more conformers that are screened, the more computational power or screening time is required. In order to speed up screening, the small molecule conformational landscape is limited to a maximum number of conformations with the lowest predicted energy in solution that summarise the full landscape. LigPrep and OMEGA are two programs to generate a 3D library of small molecules<sup>205-207</sup>.

### 3.2.1.4 Ligand-receptor docking

SBVS refers to the docking, scoring and ranking of a library of small molecules to identify those most likely to bind a target. Docking searches over the static protein 'grid' with a flexible small molecule and predicts the small molecule pose with the best fit and the greatest energy release upon binding. The poses are then scored for their predicted binding energy. Each program has its own unique scoring function, but generally the scoring balances positive binding contributions from hydrogen bonds, van der Waals interaction energies, lipophilicity complementarity and solvation effects, and negative binding contributions like steric clashes and electrostatic repulsions<sup>208</sup>. Usually, a library is first docked using a faster but less exhaustive searching mode and the top-scorers are re-docked using a slower but more accurate mode. For example, Glide had three different modes of increasing accuracy: high throughput virtual screen (HTVS), standard precision (SP) and extra precision (XP)<sup>208-210</sup>. For comparison, on a single CPU processor, Glide HTVS mode docks compounds at a rate of ~2

seconds/compound, whereas SP mode takes ~10 seconds/compound and Glide XP mode takes ~2 minutes/compound <sup>211</sup>.

### 3.2.1.5 Ligand-based virtual screening

LBVS involves scoring compounds in a library for similarity to a template molecule in terms of 2D chemical fingerprint or 3D shape, with both methods being employed in this chapter. 2D fingerprint similarity identifies common chemical features between a set of compounds. This was used in this chapter to reduce purchasing costs by trimming a set of commercially available compounds down to the most chemically diverse ones using the Canvas programs within Maestro (Schrödinger) <sup>200</sup>. 3D shape similarity is used in this chapter to develop the Affimer-guided small molecules using the program ROCS<sup>212,213</sup>. ROCS aligns compounds to a template molecule to maximise the volume overlap, then scores the similarity of the shapes of the aligned molecules. After shape similarity, further screening can be done to identify electrostatic similarity using the related program EON<sup>214</sup>.

### 3.2.1.6 Visual inspection of compounds

The outputs of VS programs still require visual inspection of the top ranked compounds. Commercial libraries contain some non-druglike compounds and compounds that are synthetically inaccessible or incompatible with downstream medicinal chemistry optimisation. Pan assay interference compounds (PAINS) which generate false positive results in a broad range of assays can also be found in commercial libraries <sup>215</sup>. The four visualisation programs used here were VIDA, Maestro<sup>200</sup>, Pymol<sup>216</sup> and ChimeraX<sup>149</sup>. They were each used for different purposes, and the advantages and disadvantages are uses are detailed in Table 3.1.

**Table 3.1 Visualisation programs**

Program	Distributor	Advantages	Disadvantages	Main uses
<b>VIDA</b>	OpenEye	<ul style="list-style-type: none"> <li>Displays ROCS outputs, scores and rank</li> </ul>	<ul style="list-style-type: none"> <li>Simple, unattractive lighting</li> <li>Cannot display proteins for SBVS</li> </ul>	<ul style="list-style-type: none"> <li>Visualising and analysing ROCS output</li> </ul>
<b>Maestro</b>	Schrödinger	<ul style="list-style-type: none"> <li>Intuitive GUI for running Schrodinger programs for SBVS</li> <li>Easy to sort through docked ligands</li> <li>Displays more interaction types e.g., Pi bonds and salt bridges</li> </ul>	<ul style="list-style-type: none"> <li>Only one basic lighting option</li> <li>Complex selection</li> <li>Limited colour scheme</li> </ul>	<ul style="list-style-type: none"> <li>Running Schrodinger processing</li> <li>Viewing Glide-docked ligands</li> </ul>
<b>Pymol</b>	Schrödinger	<ul style="list-style-type: none"> <li>Good all-rounder used for protein structure analysis and results from SBVS or LBVS</li> <li>Plugin for Autodock</li> </ul>	<ul style="list-style-type: none"> <li>Limited colour scheme and lighting</li> </ul>	<ul style="list-style-type: none"> <li>Protein structure analysis and manipulation</li> <li>Viewing non-Glide docking outputs</li> </ul>
<b>ChimeraX</b>	UCSF	<ul style="list-style-type: none"> <li>Superior lighting and colour options for making publication-ready images</li> <li>Best display of cryo-EM density</li> </ul>	<ul style="list-style-type: none"> <li>Poor at visualising VS outputs</li> <li>Only displays H bonds</li> <li>Misses some H bonds</li> </ul>	<ul style="list-style-type: none"> <li>Cryo-EM protein structure analysis</li> <li>Figure making</li> </ul>

## 3.2.2 NP-Affimer mimic development

### 3.2.2.1 Analysis of NP:NP-Affimer crystal structure

In order to identify the key Affimer residues for mimic development, the Affimer-bound CCHFV NP crystal structure (PDB: 6Z0O) was used as an input for the Protein Interfaces, Surfaces and Assemblies' service (PISA) hosted at the European Bioinformatics Institute<sup>217</sup>. The interfaces between NP-Affimer (chain E) loops 1 and 2 and CCHFV NP (chain A) were analysed and the accessible surface area ( $\text{\AA}^2$ ), buried surface area ( $\text{\AA}^2$ ) and solvation energy effect (kcal/mol) was calculated at each residue. The structure was also visually inspected in PyMOL 2.3.2<sup>216</sup>. Polar interactions between the Affimer and NP were identified as well as the orientation of binding. Following this analysis, the sequence DFWKD at NP-Affimer positions 106-110 was extracted into a new model using PyMOL.

### 3.2.2.2 Ligand-based virtual screening

The DFWKD structure from NP-Affimer was used as a query molecule in the shape similarity search program ROCS 3.2.1.4<sup>212,213</sup>. The compound library was the eMolecules small molecule library, which had been converted by Dr Katie Simmons to a maximum of 10 conformers/compound using OMEGA 2.2.0.5<sup>206,207</sup>. ROCS was split across 8 CPU cores on the high-performance computing cluster ARC3 and took 5 hours and 42 minutes (613.6 molecules/second).

The top scoring 10,000 molecules were then used as input for the electrostatic similarity search program EON (OpenEye)<sup>214</sup>, using the DFWKD structure as the template. This screen was performed on local chemistry computer CHMIBM1073 using 4 cores and took 19 minutes and 31 seconds (8.54 molecules/second).

The top ~150 scoring compounds from EON were visually inspected using VIDA (OpenEye). Promising compounds were identified based on the following criteria:

- Good mimic of the Affimer backbone shape
- Good mimic of at least one side chain
- Mimic of at least one hydrogen bond donor or acceptor present in the NP:NP-Affimer crystal structure
- A central aromatic group to introduce rigidity
- Lack of long flexible chains in the molecule

Promising compounds were screened for availability to purchase, then clustered by similarity. A 2D chemical fingerprint was assigned to each compound using Canvas (Schrödinger) and they were reduced to the eight compounds with the maximum structural diversity according to the Tanimoto coefficient. 5 mg of the eight compounds were purchased from various suppliers through eMolecules and the compounds were given IDs of AF1-01 – AF1-08.

### **3.2.2.3 Investigating the binding modes of the AF1 series using docking programs**

From the NP:NP-Affimer crystal structure (6Z0O), a single NP monomer was extracted and prepared using the protein preparation wizard in Maestro. Two grids were generated of the prepared structure, both centred on the NP residue E333 which sits in the centre of the Affimer-binding site. One grid contained only the Affimer-binding site and had the default grid dimensions of 10Å x 10Å x 10Å, and the other also contained the proposed RNA-binding site and had the longest possible grid dimensions of 40Å x 40Å x 40Å. The AF1 molecules were prepared using LigPrep with ten maximum conformers per molecule<sup>205</sup>. The prepared molecules were docked against both grids with Glide XP and the top scoring pose for each compound was identified.

### **3.2.2.4 Prediction of small molecule binding sites**

The CCFHFV NP structure (PDB: 6Z0O) was prepared using the protein preparation wizard. The small molecule binding sites were predicted in this structure using SiteMap, which is described in section 3.2.1.2<sup>24,204</sup>.

### **3.2.3 Structure-based virtual screening to identify modulators of the CCHFV NP RNA site**

A CCHFV NP structure (PDB: 6Z0O) was prepared using the protein preparation wizard in Maestro. A grid was prepared around I448, a central residue in the proposed RNA-binding site, with ligand diameter midpoint box dimensions of 20Å x 20Å x 20Å. A subset of the Chembridge DiverSET library totalling 100,000 compounds was docked against the NP grid using Glide (Schrödinger) HTVS mode. The top 10% of compounds which scored the best were then redocked using Glide SP mode and the top 5% were then redocked using Glide XP mode.

The Glide XP poses were visually inspected and compounds were discarded which had a docking score worse than -8.0, contained large, fused ring structures, contained long, flexible regions, were docked in a potentially high energy conformation, made one or fewer hydrogen bonds with CCHFV or made interactions in a single localised area of the compound. This left 103 compounds which were screened for availability through Chembridge. The 90 available compounds were given 2D chemical fingerprints by Canvas (Schrödinger) and reduced to the 15 compounds with the maximum structural diversity. Canvas used a statistical term termed the Tanimoto coefficient to identify the compounds with the lowest overlap of their 2D chemical fingerprints<sup>218</sup>. Fifteen compounds were purchased from Chembridge at 1 mg per compound and given the IDs AF2-01 – AF2-15.

### **3.2.4 Determining Influenza A virus HA-Affimer structures by cryo-EM**

### 3.2.4.1 Cryo grid preparation

For the HA-A31 complex, 3.0 mg/ml (16  $\mu$ M) HA trimer (Wilson Lab), 1.25 x molar excess of A31 and 0.2%  $\beta$ -octyl glucoside (OG) were mixed and incubated for 15 minutes at 4°C. The mixture was vitrified on self-wicking grids that were supplied by SPT Labtech and glow discharged in a Cressington 208 Carbon coater with glow-discharge unit for 80 seconds at 0.1 mbar air pressure and 15 mA current. The vitrification used a custom-built setup for rapid cryo-EM grid preparation, which has been described in detail elsewhere<sup>65</sup>.

For the HA-A5 complex, on-grid aggregation caused by Affimer binding meant the set up was modified to allow for rapid mixing and freezing, which froze Affimer-bound HA particles prior to aggregation. 6.0 mg/mL (32  $\mu$ M) HA trimer was mixed in-flow with an equal volume of A5 pre-mixed with 0.4% OG<sup>66</sup>. This gave final concentrations in the mixer of 3.0 mg/mL HA trimer, 1.25 x molar excess of A5 and 0.2% OG. The sample was vitrified on self-wicking grids that were supplied by SPT Labtech and glow discharged in a Cressington 208 Carbon coater with glow-discharge unit for 80 seconds at 0.1 mbar air pressure and 15 mA current. The method for rapid mixing and freezing has been described in more detail previously<sup>66</sup>. The time delay of 600 ms between mixing of HA and A5 and spraying was optimized to generate non-aggregated A5-bound particles.

### 3.2.4.2 Cryo-EM data collection

Movies were collected using a Titan Krios Cryo-TEM (Thermo Fisher Scientific) operating at 300 keV and equipped with a Falcon 4 Direct Electron Detector (Thermo Fisher Scientific). All cryo-EM data were acquired using the EPU 2 software (Thermo Fisher Scientific). All movies were collected in electron counting mode, over a defocus range of -2 to -4  $\mu$ m and at a nominal magnification of 96,000x which corresponded to a calibrated pixel size of 0.83  $\text{\AA}/\text{pix}$ .

### 3.2.4.3 Image processing

Image processing for both datasets began with the same steps. Movies were imported into RELION 3.1<sup>140</sup> and underwent beam-induced motion correction using RELION's own implementation of MotionCor2<sup>137</sup>. Then, the contrast transfer function (CTF) of each micrograph was estimated and corrected for using CTFFIND-4.1<sup>138</sup>. Particles were automatically picked on micrographs by CrYOLO 1.6.1 (Sphire) using the weights from its general model and a picking threshold of 0.1. 200,759 particles were picked from the H3-A31 dataset and 275,510 particles were picked from the HA-A5 dataset. Particles for both datasets were extracted into a 280-pixel box re-scaled to 100 pixels.

From here, the datasets were treated differently. HA-A31 particles were imported into cryoSPARC<sup>104</sup> and underwent one round of 2D classification. Classes containing HA

were taken forward for *ab initio* 3D model generation and particles classified into models containing HA were taken back into RELION and re-extracted in a 400 pixel box without re-scaling. The unbinned particles were then processed by cryoSPARC using the algorithm for 3D non-uniform refinement without applied symmetry and then with C3 symmetry<sup>106</sup>. Particles were imported into RELION for Bayesian polishing, then back to cryoSPARC for a second round of non-uniform refinement and global sharpening. This resulted in final map with a global resolution of 3.41 Å according to the gold standard half-map criteria at a 0.143 cut-off.

Relion 3.1 was used for all HA-A5 processing<sup>140</sup>. Particles were sorted by 2D classification and particles in classes containing HA were taken forward for 3D classification using a hypothetical map of HA generated from an apo HA model (PDB: 5HMG) using the molmap function in ChimeraX<sup>149</sup>. Particles in the one class containing HA with an Affimer bound were taken forward for 3D refinement in C3 symmetry, then re-extracted in a box of 280 pixels re-scaled to 200 pixels and 3D refined again. The particles went through two cycles of per-particle motion correction termed Bayesian polishing and CTF refinement, with 3D refinement and post-processing between each step. This gave a final map with global resolution of 4.4 Å according to the gold standard half-map criteria at a 0.143 cut-off.

#### 3.2.4.4 HA-A31 model building

Model building was performed by Dr Oliver Debski-Antoniak. AlphaFold models of Affimer molecules A5 and A31 were first generated<sup>45</sup>. HK68 (PDB:4FNK) HA and the AlphaFold Affimer models were rigid-body fitted into the cryo-EM maps produced using the UCSF Chimera ‘fit in map’ tool<sup>219</sup>. H3-A31 was then modelled by first improving the fit in Coot<sup>147</sup>, before utilizing Namdinator<sup>220</sup>. To aid model fitting around the Affimer region, the final H3-A31 map was sharpened using DeepEMhancer as implemented in COSMIC2<sup>221,222</sup>. This map was used alongside cryoSPARC maps to improve confidence in A31 modelling. The final model was refined in Phenix and iteratively improved in Coot.

### 3.2.5 Developing Affimer A5 mimics

#### 3.2.5.1 Affimer sequence alignment

Affimer protein sequences were aligned using the Clustal Omega server (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) for multiple sequence alignment. Structures of Affimers against various targets (Table 3.2) were aligned in Pymol (Schrödinger).

**Table 3.2 Affimers used for structural alignment**

Affimer target	PDB code
----------------	----------

Sumo2 (S2B3 isoform)	5ELU
BclxL	6HJL
p300	5A0O
K33 diUb	5OHV
Alpha actinin 2 CH domains	6SWT
Fc gamma receptor	5MN2
KRAS	6YR8
Bcl-2	6ST2
Sumo1 (S2D5 isoform)	5ELJ

### 3.2.5.2 IAV HA Affimer homology modelling

Homology models of the HA Affimers were generated using the I-TASSER server with the default parameters <sup>223</sup>. The template for the modelling was a crystal structure of an Affimer isolated against the Bcl-xL protein which had the highest sequence similarity to the Affimers <sup>224</sup> (PDB: 6HJL). The loop region was extracted from the homology models as previously described and used as inputs for ROCS <sup>212,213</sup>.

### 3.2.5.3 Ligand-based virtual screen using 3D shape similarity

Small molecule mimics of the A5 and A31 loops were identified using shape similarity program ROCS using the same workflow outlined in section 3.2.2.2, except only the top 1000 conformers from the ROCS run were used as input for EON <sup>213</sup>. The rules for identifying promising compounds were more general than for the AF1 mimics of NP-Affimer due to the lack of high-resolution HA-Affimer structures. The rules were based on the following criteria:

- Good mimic of the Affimer backbone shape
- Good mimic of at least one side chain
- A central aromatic group to introduce rigidity
- Lack of long, flexible chains in the molecule
- Inclusion of at least one potential hydrogen bond donor or acceptor matching the amino acid chain from NP (since the exact groups involved in hydrogen bonds were unknown)

Promising compounds were grouped according to their pharmacophore and one compound was picked from each group to purchase. Of the nine mimics of loop 1, only six were available to purchase, whereas for the six mimics of loop 2, only two were available to purchase. Of the eight purchasable compounds, 5 mg were purchased

from various suppliers through eMolecules and the compounds were given the IDs OA1-OA8.

### 3.3 Results

The objectives of this chapter were to develop a LBVS method for Affimer-guided small molecule discovery based on a 3D shape similarity program. An Affimer inhibitor of CCHFV NP termed NP-Affimer and two Affimer inhibitors of IAV HA named A5 and A31 were used as test systems for this novel method.

#### 3.3.1 Small molecule mimics of NP-Affimer were developed using a NP:NP-Affimer crystal structure

##### 3.3.1.1 NP-Affimer binds to CCHFV NP through loop 2 residues 106-110

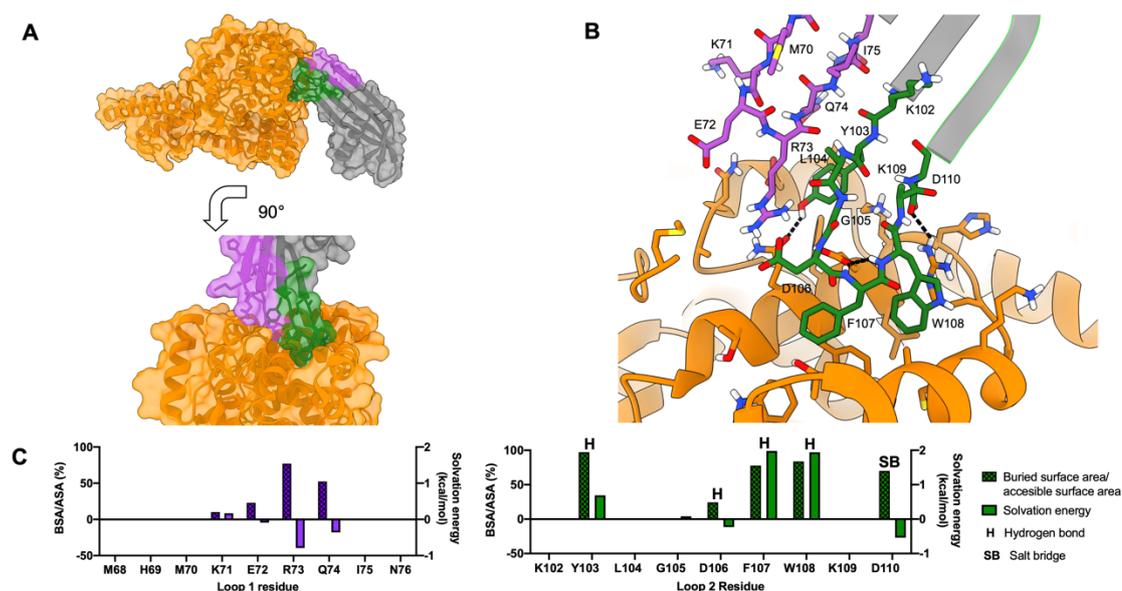
The loops in NP-Affimer are much larger than the drug-like molecules in small molecule libraries, meaning small molecules will not accurately mimic the whole loop region. The residues in NP-Affimer loops 1 and 2 have combined masses of 1,186 Da and 1,171 Da respectively, but the mass of a typical drug-like small molecule is <500 Da according to Lipinski's rule-of-5<sup>225</sup>. Since Alvarez-Rodriguez *et al.* solved a crystal structure of CCHFV NP:NP-Affimer, this was analysed for the core site of NP:NP-Affimer interactions to reduce the template mass and focus virtual screening efforts.

As well as visual analysis of the structure in ChimeraX (Figure 3.6A-B), the binding at each residue was also quantified using PDBePISA (Figure 3.6C). The metrics calculated by PDBePISA were buried surface area/accessible surface area (BSA/ASA), the percentage of the surface area that is buried upon binding, and the solvation energy, a measurement of the effect of polar and hydrophobic interactions. A positive solvation energy indicates interactions were based on the hydrophobic effect and a negative value indicates the interactions were due to the polarity between residues.

NP-Affimer binds to a shallow groove on NP's head domain that was proposed to be the site of NP oligomerisation, with loop 2 positioned close to the surface but loop 1 pointing away (Figure 3.6A). In particular, the residues in the second half of loop 2, from positions 106-110, are positioned closest to NP, with F107 and W108 sitting in a groove between two helices (Figure 3.6B). Four residues in loop 2 (Y103, F107, W108 and D110) were buried >70% upon binding, but only R73 was buried as much in loop 1 (Figure 3.6C). Therefore, residues 106-110 in loop 2 form the core interaction site.

Solvation energies calculated by PDBePISA highlight that the chemical basis for interactions varies throughout loop 2. Consistent with the visual inspection, binding at Y103, F107 and W108 is driven by the hydrophobic effect but binding at D106 and D110 is driven by polarity (Figure 3.6C). Considering the hydrogen bonds, which are among the strongest non-covalent bonds between proteins, ChimeraX identified that

loop 1 does not make any with NP but loop 2 makes five, with four of them located between positions 106-110 (D106, F107, W108, D110) (Figure 3.6B and Figure 3.6C). PDBePISA identified that the polar interaction at D110 is actually a salt bridge, which is even stronger than a hydrogen bond. The binding energies for a salt bridge and a hydrogen bond in a protein in aqueous solution have been calculated as 3-5 kcal/mol<sup>226</sup> and 1-2 kcal/mol<sup>227</sup>, respectively. Overall, a combination of interactions driven by polarity and hydrophobicity was important for NP:NP-Affimer binding and should be mimicked by small molecules.



**Figure 3.6 The X-ray crystal structure of Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) bound to an inhibitory Affimer (NP-Affimer). Residues 106-110 of the second Affimer loop is the core site of NP binding.** A) NP-Affimer bound to the CCHFV NP head region. NP shown as orange surface and ribbons, Affimer shown as surface and ribbons with scaffold, loop 1 and loop 2 coloured as grey, purple and green, respectively. Same colour scheme used throughout figure. B) Affimer loop 2 is the core site of NP interactions. Affimer loop and interacting NP residues highlighted with sticks. C) PDBePISA analysis of NP-Affimer:NP crystal structure showing buried surface area/accessible surface area and solvation energy for each NP-Affimer residue. Residues where hydrogen bonds or salt bridges were identified are labelled with H and SB respectively. Panels A and B were generated using ChimeraX and panel C was generated in GraphPad Prism.

The crystal structure of NP bound to NP-Affimer identified that the second half of loop 2 from D106-D110 was the core site of binding. This DFWKD pentapeptide was excised from the structure and used as the template for mimic development. Even though there were interactions outside of this region, they were most concentrated here and the lower mass of this region of 710 Da could be better mimicked by a small molecule than the full loops.

### 3.3.1.2 Small molecule mimics of the NP-Affimer were identified using shape and electrostatic similarity programs

The section of loop 2 from D106-D110 was excised from the NP-Affimer:NP structure and used as a template for mimic development. The side chain of K109 had been truncated to a methyl group by Alvarez-Rodriguez *et al.* due to weak electron density in the crystal structure (Figure 3.6B). So that the DFWKD structure matched the true Affimer sequence, the full lysine side chain was added to K109 *in silico* (Figure 3.7A).

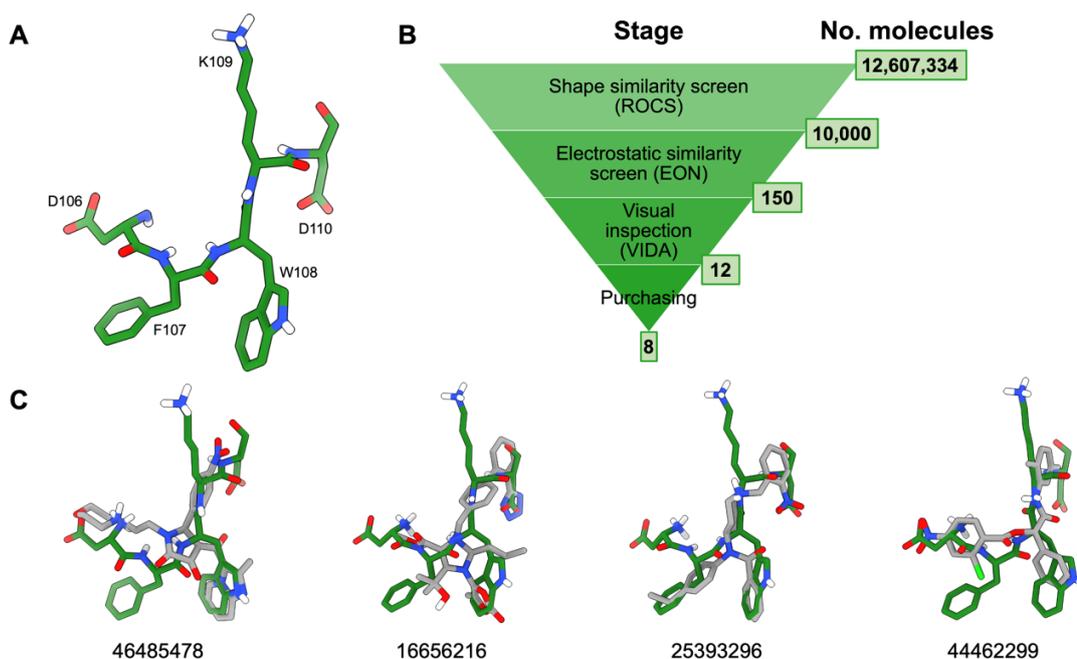
To find small molecules that mimicked NP-Affimer binding, the DFWKD structure was used as a template in the shape similarity search program ROCS (OpenEye)<sup>213</sup> and the electrostatic similarity search program EON (OpenEye)<sup>214</sup>. ROCS and EON both function by best aligning each molecule from a chemical library against the template

structure, scoring this pose for its similarity to the template, then ranking the compounds in order of score. Therefore, as well as needing a template molecule, ROCS and EON also require a three dimensional chemical library to screen against.

From the wide range of commercially available chemical libraries, the eMolecules database was chosen because by compiling seventeen different suppliers into one easy-to-use database, it covers a larger chemical space than any one supplier<sup>228</sup>. The eMolecules screening compound library was downloaded and converted from 2D to 3D conformations by Dr Katie Simmons using OMEGA (OpenEye)<sup>206,207</sup> with 10 maximum conformations per molecule. This gave 12.6 million different conformers in the 3D library. LBVS programs like ROCS screen compounds faster than SBVS programs due to the smaller search space they cover, meaning 12 million conformers is within the range for ligand-based virtual screen to triage within a day, whereas this is not the case for a SBVS program.

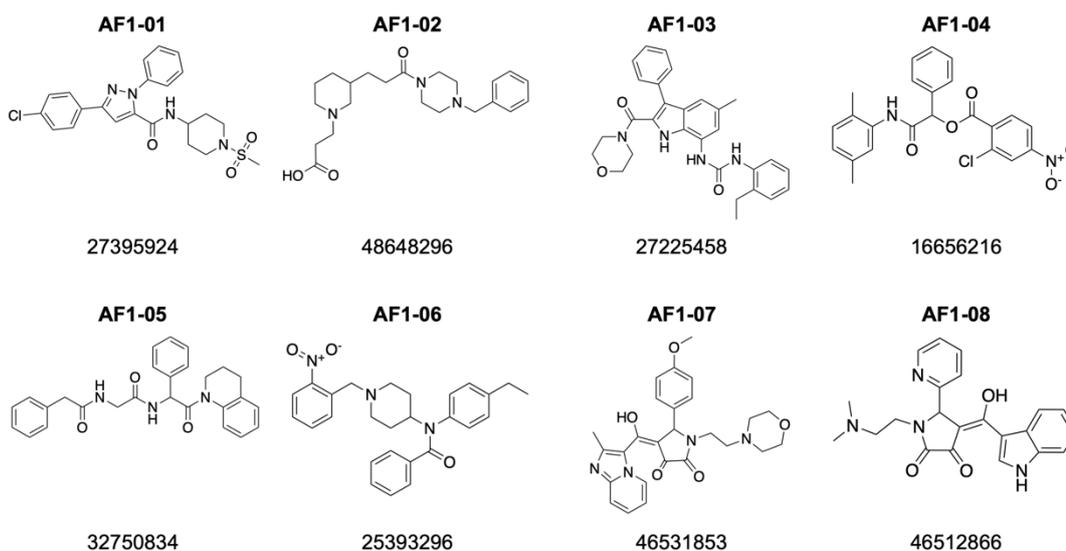
The eMolecules library was then screened using ROCS for compounds similar in shape to the DFWKD pentapeptide, then the 10,000 top scoring compounds were used as input for EON to search for compounds with similar electrostatics (Figure 3.7B). The Approximately the 150 top scoring compounds from EON were manually inspected and promising compounds identified using general and rules specific to NP-Affimer. In general, Affimer mimics should mimic some of the main chain and at least one side chain, be in the same orientation as the loop and not contain hydrophobic fused ring structures or flexible chains. Specifically for NP-Affimer, the high resolution structure of NP-Affimer bound to NP means at least one of the four hydrogen bonding groups in the template should be mimicked by promising compounds.

Overall, small molecules identified via this LBVS did not mimic the entirety of the template pentapeptide well. For example, whereas compound 46485478 mimicked W108 well with a nitrogen-containing double ring structure, it mimicked D106 and D110 poorly with a slightly unaligned nitro group and a non-planar morpholine ring, respectively (Figure 3.7C). Compound 16656216 mimicked D106 well with a well aligned nitro group but mimicked W108 poorly with a benzene ring. Similarly, compound 25393296 mimicked D110 well with a well aligned nitro group but mimicked W108 poorly with single ringed benzene. Finally, compound 44462299 mimicked D110 well with a tetrazole group, but mimicked W109 poorly with a dioxolone ring. There were no compounds that mimicked K109 well, likely because it is long and flexible and drug-like small molecules do not typically contain these groups. However, this was not a problem because K109 pointed away from NP in the structure and was not buried upon binding (Figure 3.6B and Figure 3.6C).



**Figure 3.7 A ligand-based virtual screening workflow identified small molecules that mimic the nucleoprotein (NP) -binding region in an inhibitory Affimer (NP-Affimer).** A) The DFWKD pentapeptide that forms the core site of interactions and was used as template for the virtual screen. B) The virtual screening workflow used to identify small molecule mimics of the Affimer loop from the eMolecules library of 1.2 million compounds. C) Example promising compounds from the virtual screen. NP-Affimer DFWKD template is shown as green stick and the compounds are shown as grey sticks. The eMolecules ID number is shown for each compound. All images generated using ChimeraX.

Twelve compounds were selected for purchasing but only eight were available for a reasonable price, so these were purchased and named AF1-01 - AF1-08 (Figure 3.8). However, AF1-02 did not arrive and so was not screened in downstream assays.



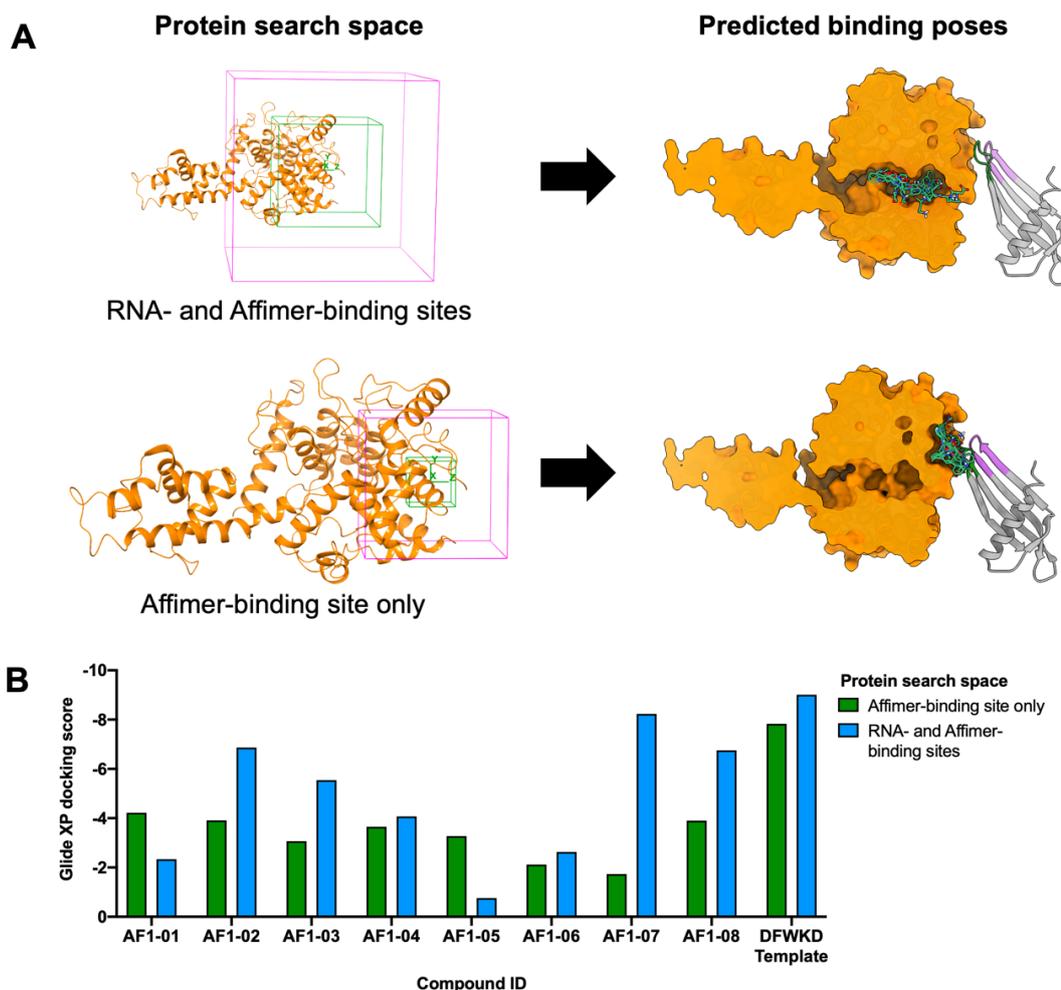
**Figure 3.8 The AF1 set of small molecules generated using a novel Affimer-guided approach.** eMolecules ID for each compound is labelled. Images generated using Chemdraw.

### 3.3.1.3 *In silico* predictions of Affimer mimic binding identified the RNA-binding tunnel as a potentially druggable site on CCHFV NP

To probe whether the AF1 NP-Affimer mimics were similar enough to their template to still bind at the NP Affimer-binding site *in silico*, the compounds were docked into the NP protein structure using Glide (Schrödinger)<sup>208</sup>. Most docking programs require the region of interest to be extracted into a grid. This reduces the search space and time by masking out unnecessary regions of the protein. However, narrowly focussing on an area can bias the experiment by forcing the docking program to reveal the expected poses.

NP from the NP:NP-Affimer structure (PDB: 6Z0O)<sup>86</sup> was extracted and processed through the protein preparation wizard in Maestro (Schrödinger). Since there were only eight mimics to dock, the dimensions of the grid, which was centred around the Affimer-binding site, was maximised to reduce bias (Figure 3.9A). When the AF1 compounds were docked into this grid along with the DFWKD template as a control, all the poses were located not at the Affimer-binding site, but in an adjacent region buried within the protein, proposed to be the RNA-binding site<sup>229</sup>. Since the Affimer template does not bind at this proposed RNA-binding site, the binding of Affimer mimics to this site was unexpected.

It is possible that the compounds were docked to the Affimer-binding site, but that the poses in the RNA-binding site scored higher. To test this, the AF1 compounds were docked against a smaller grid which excluded the RNA-binding site (Figure 3.9A). The resulting poses were all found at the Affimer-binding site, predicting that the AF1 compounds do bind like their Affimer template. However, the docking scores for poses in the RNA-binding site were typically better (more negative) than in the Affimer-binding site (Figure 3.9B). Therefore, Glide predicted that the AF1 compounds have two binding sites, with them binding more tightly in the RNA-binding site. This could be identified experimentally by binding assays or by obtaining ligand-bound NP structures.

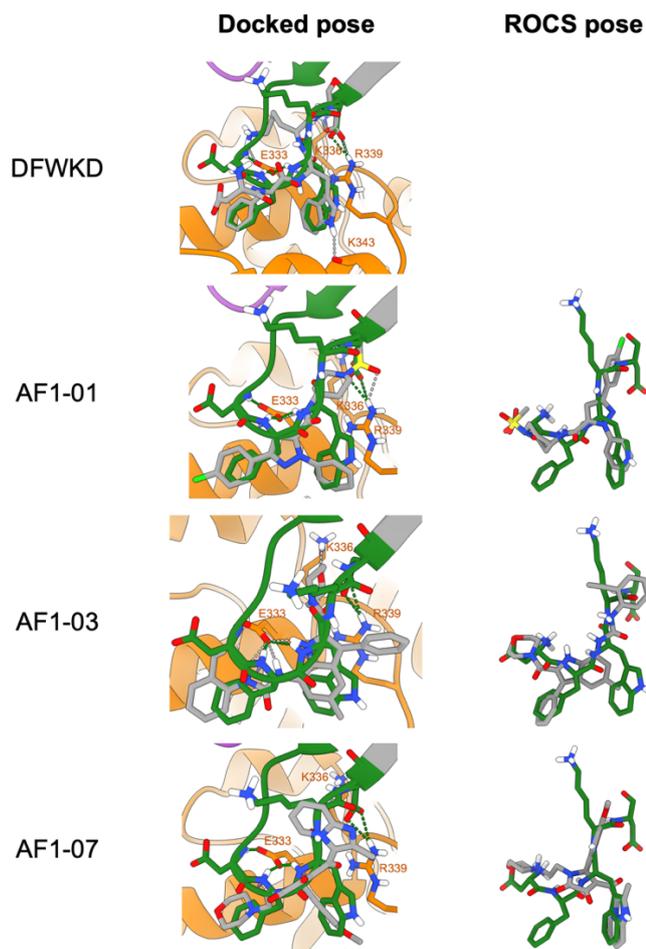


**Figure 3.9 Docking experiments identified the RNA binding tunnel in Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) as a potentially drugable site.** The protein grid used for docking using Glide (Schrödinger) and the resulting poses of the docked compounds (green sticks). The pink cube represents the search space outer limit and the green cube represents the ligand diameter midpoint box. The protein search space images were generated using Maestro (Schrödinger) and the predicted binding pose images were generated using ChimeraX. B) Docking scores calculated by Glide XP for compounds when docked in the two protein search spaces shown in panel A: the larger search space containing the RNA- and Affimer-binding sites and in the smaller search space containing only the Affimer-binding site. Image produced using Prism (GraphPad).

To test how the Affimer-guided small molecules might bind to the Affimer-binding site and to identify whether the structure-based and ligand-based virtual screening methods generated the same poses for the AF1 compounds, the poses generated through the docking and shape-similarity programs were compared to the pose of the template Affimer in the NP:NP-Affimer crystal structure (Figure 3.10).

As a control, the DFWKD template was docked at the Affimer-binding site and showed the same pose as it forms in the crystal structure, indicating this SBVS method can be used to validate mimics generated by Affimer-guided LBVS. However, the pose for AF1-01, which was the mimic with the highest docking score in the Affimer-binding site, was flipped compared to the pose derived through shape similarity screening, casting some doubt as to the validity of the mimic's binding. Either way, the pose for AF1-01 derived through SBVS aligns well with the Affimer loop and AF1-01 mimics the

hydrogen bonds involving W108 and D110. This suggests that AF1-01 might be able to bind to CCHFV NP in either orientation. A similar effect was seen for AF1-03. Its SBVS-derived pose showed the compound flipped relative to its pose derived by shape similarity screening, and yet in the SBVS-derived pose, it recreated two hydrogen bonds involving F107 and W108 and formed a new hydrogen bond with K336. Finally, AF1-07, which generated the lowest docking score in the Affimer-binding site, was not flipped but was placed in a different orientation compared to its LBVS-derived pose and Glide predicted that it recreated none of the hydrogen bonds present in the crystal structure.

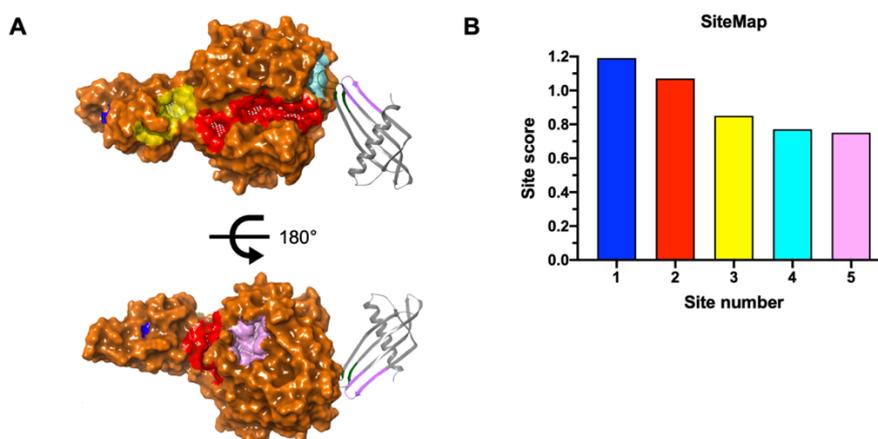


**Figure 3.10 Docked poses compared to ROCS-aligned poses for example compounds from the AF1 series of small molecule NP-Affimer mimics.** Docking was performed on a grid that was restricted to the Affimer-binding site. Affimer loop 1, loop 2 and scaffold shown as purple, green and grey ribbons. The DFWKD AF1 template is highlighted as sticks. AF1 compounds shown as grey sticks. Hydrogen bonds are coloured according to molecule they come from. For figures A and C, the Affimer was removed for the docking experiments but it is shown here for comparison.

In summary, validating the LBVS poses using SBVS was not successful. Not only did the SBVS predict that the Affimer mimics bind more tightly to the proposed RNA-binding site compared to the Affimer-binding site, all three SBVS-derived poses highlighted here showed the compound in a different orientation to the pose predicted

by LBVS. Taken at face value, SBVS predicts that the small molecule Affimer mimics bind differently to the Affimer. On the other hand, it is not surprising that the SBVS program scored compounds in the RNA-binding site higher than in the Affimer-binding site, because the RNA-binding site more closely resembles the deep, charged enzyme-substrate pockets that scoring methods for SBVS programs like Glide have been trained on. Furthermore, *in silico* binding predictions are not a substitute for real structural binding data and determining how the Affimer mimics bind to NP is key to validating the Affimer-guided approach.

Whether the SBVS or LBVS poses are correct, SBVS identified the RNA-binding site as a likely spot for drug binding. So, to further probe the druggability of the RNA-binding site, the most likely small molecule binding sites in NP were predicted, scored and ranked using SiteMap (Schrödinger). Five binding sites were predicted on NP (Figure 3.11A-B), with the top scoring site being a small pocket located in the stalk region (Figure 3.11A, blue surface), followed by the RNA-binding site in second (Figure 3.11A, red surface), with the Affimer-binding site ranked fourth (Figure 3.11A, cyan surface).



**Figure 3.11 SiteMap (Schrödinger) analysis of Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) for potential ligand-binding sites.** NP (orange surface) with the top five potential small molecule binding sites (blue, red, yellow, cyan and lilac surfaces) according to SiteMap. Affimer was removed from the structure for analysis but it is included here for reference. B) SiteScores for the top five ligand binding sites in NP as predicted by SiteMap.

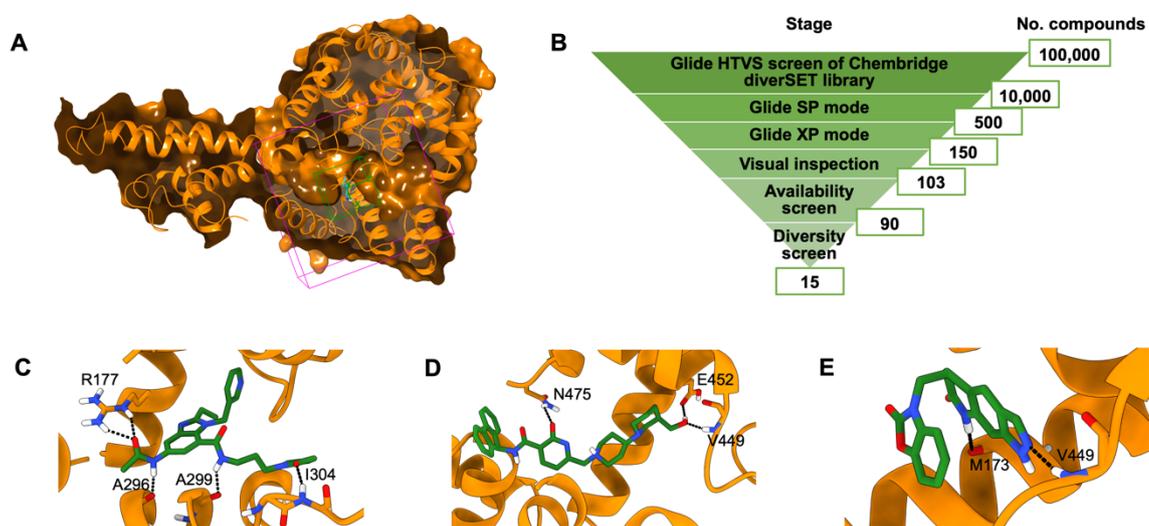
Like the SBVS experiments described previously, the binding site prediction program also predicted the RNA-binding site to be more amenable to targeting with small molecules than the Affimer-binding site. This indicated that CCHFV inhibitors could be generated by targeting the RNA-binding site by a SBVS workflow. This would also allow the comparison of the novel Affimer-based approach with a more traditional approach.

### 3.3.2 A traditional protein structure-guided virtual screening workflow identified potential small molecules binders of CCHFV NP

Given that different in silico approaches predicted the RNA-binding site to be a druggable region, to find potential small molecule at this site, a SBVS workflow was performed there. The Chembridge diverSET library was selected for the screen because despite its smaller size (100,000 compounds) compared to the eMolecules library, its high structural diversity means it samples a broader chemical space that can be searched through more quickly. This is particularly important for SBVS, since SBVS programs screen at a slower rate than LBVS programs.

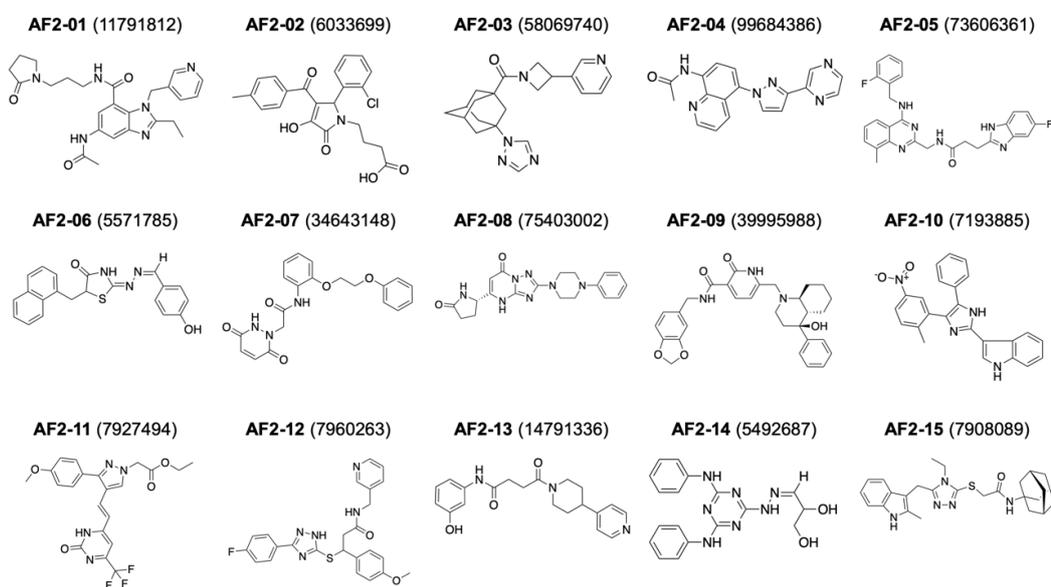
The full Chembridge diverSET library was docked against a grid centred at the RNA-binding site (Figure 3.12A). This was done using a virtual screening workflow, meaning all the compounds were first docked using Glide HTVS mode, which favours docking speed over sampling breadth (Figure 3.12B). The 10% top scoring compounds were then re-docked using the more balanced Glide SP mode and finally 5% top scoring compounds were then re-docked using Glide XP mode. Glide XP employs a more exhaustive sampling approach and is efficient at removing false positives, but takes more time per compound <sup>210</sup>.

The 150 top scoring compounds with scores  $< -8.0$  were visually inspected and undesirable compounds were removed. An example of a drug-like molecule is the top scoring compound ZINC11968242 with a score of  $-10.263$  (Figure 3.12C). This compound has a good mixture of hydrophobic and polar groups and makes H bonds with NP across the molecule, introducing rigidity into the binding pose. Undesirable compound properties include it containing large hydrophobic portions of fused aromatic rings which lower its solubility, such as that found in ZINC20600420 (Figure 3.12D), it binding in energetically unfavourable conformations such as ZINC67447014 (Figure 3.12E), which is folded over on itself or it containing flexible portions not bonded to NP.



**Figure 3.12 A virtual screening workflow identified 15 potential binders of the Crimean-Congo Haemorrhagic Fever Virus (CCHFV) nucleoprotein (NP) RNA-binding site.** Grid used for docking experiments against NP RNA-binding site. Pink cube represents the search space outer limit and the green cube represents the ligand diameter midpoint box. NP is clipped to highlight the RNA-binding tunnel. Image produced in Maestro (Schrödinger). B) Virtual screening workflow used to screen the Chembridge DiverSET library for binders of the RNA-binding site. C) Top-scoring compound ZINC11968242 showing a favourable conformation. D) Compound ZINC20600420 that was not taken forward because of its large, hydrophobic fused-ring portion. E) Compound ZINC67447014 that was not taken forward because of its energetically unfavourable folded conformation. Figures C-E were produced in ChimeraX.

103 promising compounds were then screened for availability to purchase through Chembridge at that time. The 90 available compounds were reduced to the 15 with the maximum structural diversity or lowest 2D chemical fingerprint similarity as calculated by Canvas (Schrödinger) according to the Tanimoto index statistical term (Figure 3.12B). These 15 compounds were purchased from Chembridge and named AF2-01 – AF2-15 (Figure 3.13).



**Figure 3.13 Fifteen AF2 compounds selected using a traditional structure-based virtual screening workflow against the CCHFV NP RNA-binding site.** Compounds Chembridge ID numbers are shown. Images produced using ChemDraw.

### 3.3.3 Mimics of the HA-Affimer were developed without prior structural information

In order to further test the applications of this new Affimer-guided method for small molecule discovery, the approach was applied to a different set of Affimers. Affimers A5 and A31 were isolated against the hemagglutinin (HA) spike protein from influenza A virus (IAV) by Dr Oliver Debski-Antoniak and were shown to inhibit live IAV infection in a TCID<sub>50</sub> assay and bind with nM K<sub>D</sub> binding affinities to the sialic acid receptor-binding domain of the HA1 subunit<sup>157</sup>. However, the exact binding mode was not verified via a 3D Affimer-HA structure.

During the development of the AF1 series of NP-Affimer mimics, information from the Affimer-bound NP structure was helpful for identifying the core site of binding, allowing the search space to be reduced by 75% and the virtual screen to be focussed on this region. This was likely a reason that the resulting small molecules mimicked the template so well. Furthermore, the high resolution in the structure even allowed the identification of the exact H bond donors and acceptors that were to be mimicked. Therefore, obtaining an Affimer-bound HA structure would aid in the development of small molecules against HA, as well as helping to verify the basis for HA inhibition by A5 and A31.

Cryo-EM was selected over X-ray crystallography as the method for obtaining this structure. Although both methods have led to high-resolution HA structures previously, the sample requirements for X-ray crystallography are much higher than that for cryo-EM, meaning there was only enough purified HA protein for cryo-EM. However, HA is a difficult protein to image by cryo-EM due to its preference for orienting in top/bottom views, caused by interactions HA makes with the air-water-interface (AWI) during vitrification<sup>230</sup>. The number of HA side views can be increased by the inclusion of detergents such as  $\beta$ -octyl glucoside (OG) in the sample buffer<sup>67</sup>, the use of a continuous carbon support on the grid and rapid sample application and vitrification<sup>231</sup>. The latter works by reducing the time given for HA to make those interactions with the AWI. A bespoke sprayer for rapid cryo-EM grid preparation at the Astbury Biostructure Laboratory, in combination with detergents, was considered the most appropriate for generating the Affimer-bound HA structures<sup>65,66</sup>.

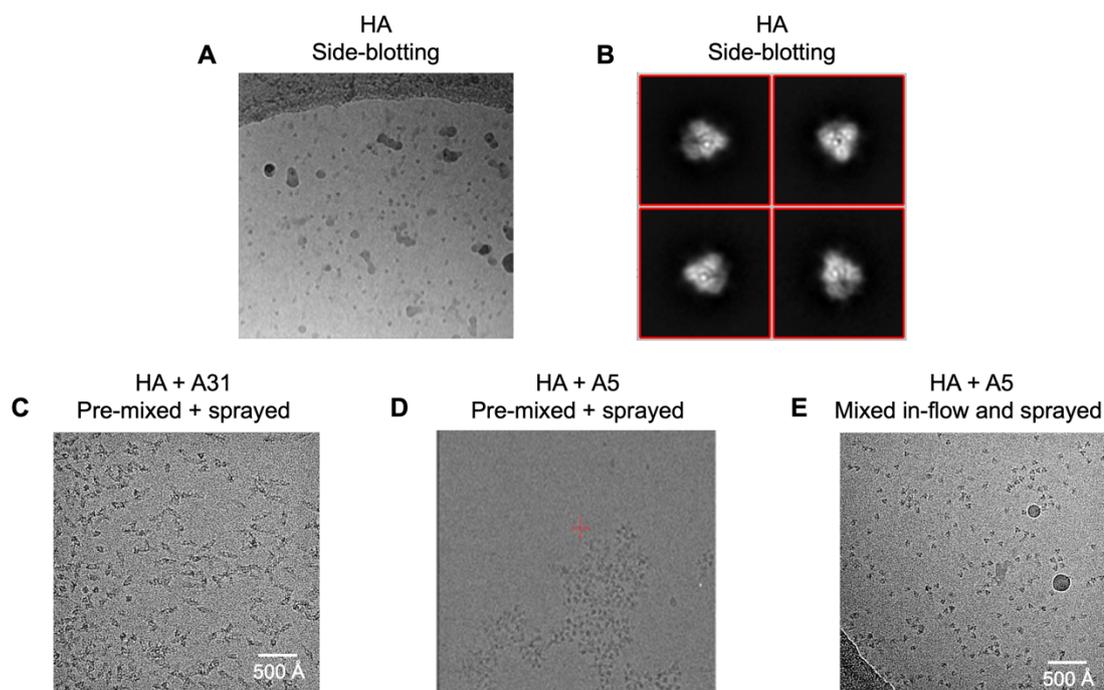
### **3.3.3.1 Initial attempts identified issues with standard cryo-EM analysis of HA**

To test that the HA sample provided by Prof Ian Wilson's group showed preferred orientation like HA samples in literature, Dr Oliver Debski-Antoniak and Dr David Klebl prepared grids of HA using the traditional side blotting method, whereby protein was directly applied to a glow-discharged cryo grid, then over the course of several seconds, blotted with filter paper and plunged into liquid ethane<sup>129</sup>. As expected, Dr Debski-Antoniak and Dr Klebl found HA was oriented in top/bottom views, with no side views of HA visible in the micrographs (Figure 3.14A) or in the 2D classes (Figure 3.14B). Therefore, rapid (ms timescale) cryo grid preparation using a bespoke sprayer was tested for the sample.

Dr Klebl and Dr Debski-Antoniak pre-mixed HA with A5 and A31 separately, then sprayed the mixtures onto grids which were rapidly travelling into liquid ethane. When these grids were imaged, the two samples behaved very differently. Whereas the HA-A31 sample showed a good mixture of orientations and mono-disperse particles (Figure 3.14C), the HA A5 sample had formed large aggregates on the grid (Figure

3.14D). HA aggregation was not visible in the sample tubes prior to grid application, indicating that the aggregation was occurring on the grid.

To prevent the on-grid aggregation, myself, Dr Debski-Antoniak and Dr Klebl used a mixer within the sprayer to mix HA and A5 in-flow before spraying onto the grid rather than pre-mixing HA and A5. This rapid mixing and spraying reduced the time available for HA to aggregate upon A5-binding<sup>66</sup>. The resulting HA particles were monodispersed, indicating that this was an effective way at preventing A5-induced HA aggregation (Figure 3.14E). Although the HA particle orientations were more variable than when side blotting was used, they were still dominated by top/bottom views. Optimisation found that 600ms was a good compromise between allowing time for A5 to bind to HA but enough for HA to adopt a preferred conformation.



**Figure 3.14 Issues of hemagglutinin (HA) preferred orientation and on-grid Affimer A5-induced aggregation were solved by rapid mixing and rapid cryo grid preparation.** A) Grids of HA prepared using a standard side-blotting protocol. HA is completely oriented as top/bottom views. B) 2D classes of HA particles processed from the grid shown in figure A. C) HA pre-mixed with Affimer A31 and sprayed rapidly onto a cryo grid during vitrification. The rapid grid preparation increased the number of HA side views present on the grid. D) HA pre-mixed with Affimer A5 and sprayed rapidly onto a cryo grid during vitrification. HA has aggregated on-grid due to A5 binding. E) HA and A5 mixed in-flow within the sprayer with a mixing time of 600 ms and sprayed rapidly onto the grid before rapid vitrification. This has prevented aggregation and improved side views compared to the standard protocol.

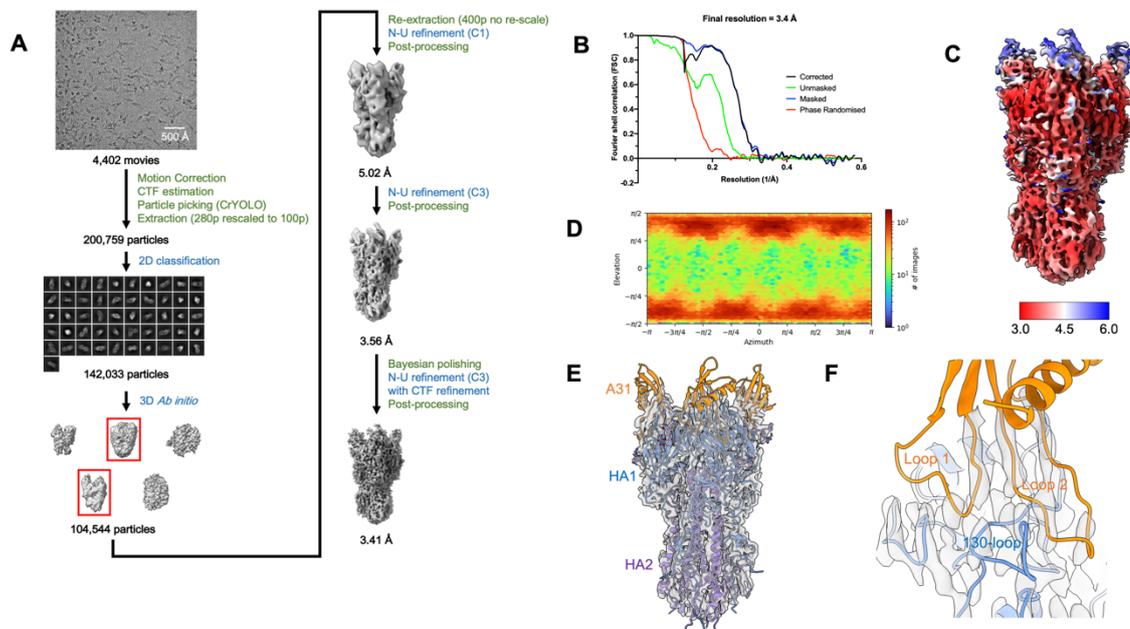
One final issue was that the resolution of the reconstructions from sprayed grids were limited by the thickness of the ice in the sprayed droplets, and so to overcome this, myself, Dr Debski-Antoniak and Dr Klebl sprayed samples onto self-wicking grids which use wires on the grid bars to wick away sample and thin the sample layer prior to freezing<sup>232</sup>. It was through imaging these grids that we were able to obtain higher resolution reconstructions of A31 and A5 bound to HA.

### 3.3.3.2 Rapid cryo-EM grid preparation generated the first A31-bound HA structure at 3.4 Å resolution

HA was pre-mixed with A31 and sprayed onto hairy cryo-EM grids by Dr Klebl and myself. The frozen grids were imaged by cryo-EM by Dr Klebl and the data processing was performed by myself and Dr Klebl. Initially the data was processed in Relion, but this generated a short, unrealistic HA particle that was likely caused by remaining preferred orientation. CryoSPARC possesses a unique non-uniform (N-U) refinement algorithm<sup>106</sup>, which applies different weighting during 3D refinement for maps of spatial variability, as was the case for HA + A31, which was a protein complex and suffered preferred orientation. So, the HA + A31 dataset was processed using both Relion and

cryoSPARC, leading to a 3.41 Å map of A31 bound to HA (Figure 3.15A and Figure 3.15B).

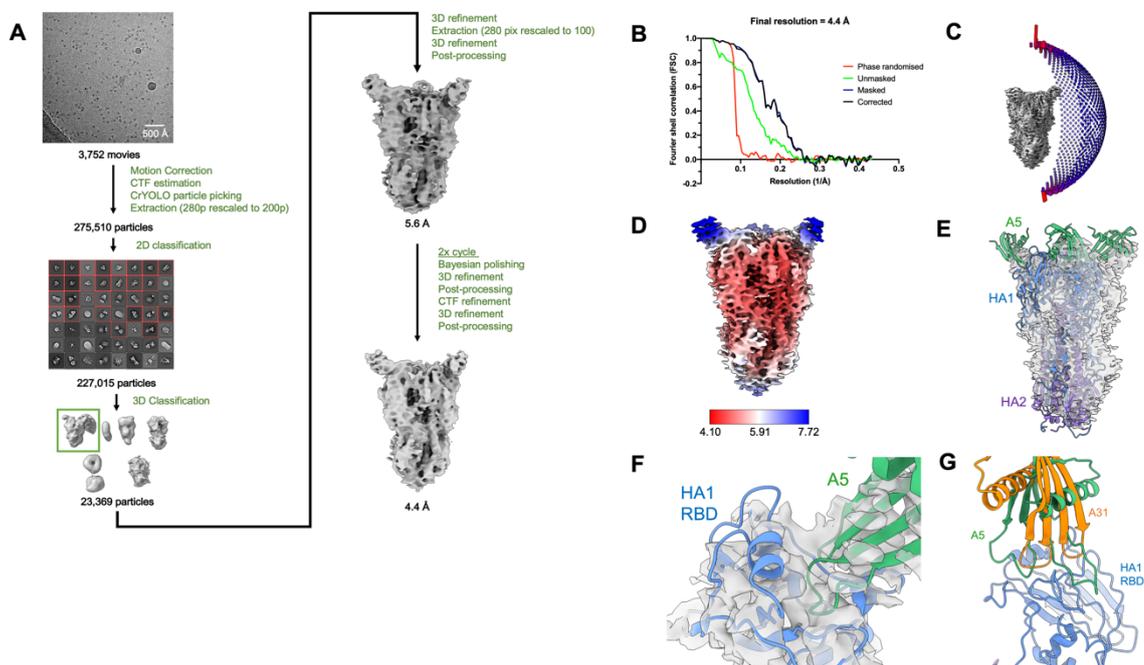
Although the local resolution was close to 3 Å for HA, it was 5-6 Å for the Affimer region, likely indicating the flexibility of the Affimer when bound to HA (Figure 3.15C). The distribution of angles used in the reconstruction shows that there was still a clear preference for top/bottom views which likely limits the resolution and contributes to the visible anisotropy in the map (Figure 3.15D). Dr Debski-Antoniak built atomic models of HA and A31 into the cryo-EM density (Figure 3.15E). Although the cryo-EM density is weak for the Affimer, the structure confirms that A31 binds to the RBD of HA. The two A31 loops can be seen straddling the 130-loop which is directly involved in binding to sialic acid on cell-surface receptors (Figure 3.15F). Overall, although we obtained the first Affimer-bound HA structure and uncovered which site A31 blocks on HA, the poor resolution in the Affimer prevented us narrowing down which residues or even which of its loops was most important for binding.



**Figure 3.15 Cryo-EM map of inhibitory Affimer A31 bound to hemagglutinin (HA) generated by using rapid grid preparation.** A) Image processing workflow showing how movies of HA and A31 were processed into a final map with global resolution of 3.41 Å. Blue text indicates a process was performed in cryoSPARC and green text indicates it was performed in Relion. B) Fourier shell correlation of the final HA + A31 map. C) Local resolution of the final HA + A31 map. D) Angular distributions of particles for the final HA + A31 reconstruction. E) Model of HA and A31 model overlaid with cryo-EM map. HA1 and HA2 subunits shown as blue and purple cartoons, respectively. A31 shown as orange cartoon. Colour scheme is used throughout. F) A31 binding site at the HA1 receptor binding domain. Loops 1 and 2 straddle the 130-loop used for receptor binding.

### 3.3.3.3 Time-resolved cryo-EM generated the first A5-bound HA structure at 4.4 Å resolution

Grids of HA and A5 were prepared by mixing the samples in-flow and spraying the mixture onto a self-wicking grid after 600ms delay time. These were imaged and processed into a 4.4 Å resolution map using Relion (Figure 3.16A and B). CryoSPARC was initially used to process the data because of its success with A31, but it generated a map of HA-A31 where HA was unrealistically short. The visible anisotropy in the map was likely due to the imbalance of top/bottom views over side views, which was even more pronounced than in the A31 dataset (Figure 3.16C). The final reconstruction was formed from only 20% of the number of particles that generated the HA-A31 map. The lower particle number and greater anisotropy meant that this map had a lower global resolution, with local resolution of 4-5 Å in the HA region and 6-8 Å in the Affimer region (Figure 3.16D). Due to the weak density throughout the map, during the model building by Dr Debski-Antoniak, a previously determined HA structure and AlphaFold-generated A5 structure were simply rigidly fit into the map (Figure 3.16E). The location of the Affimer in the structure confirms that the Affimer binds at the RBD, blocking HA-receptor binding, but further detail is unclear at this resolution (Figure 3.16F). Overlaying the A5- and A31-bound structures based on the HA coordinates indicates that the two Affimers bind at the same region and likely inhibit HA-receptor binding by blocking similar interactions (Figure 3.16G).

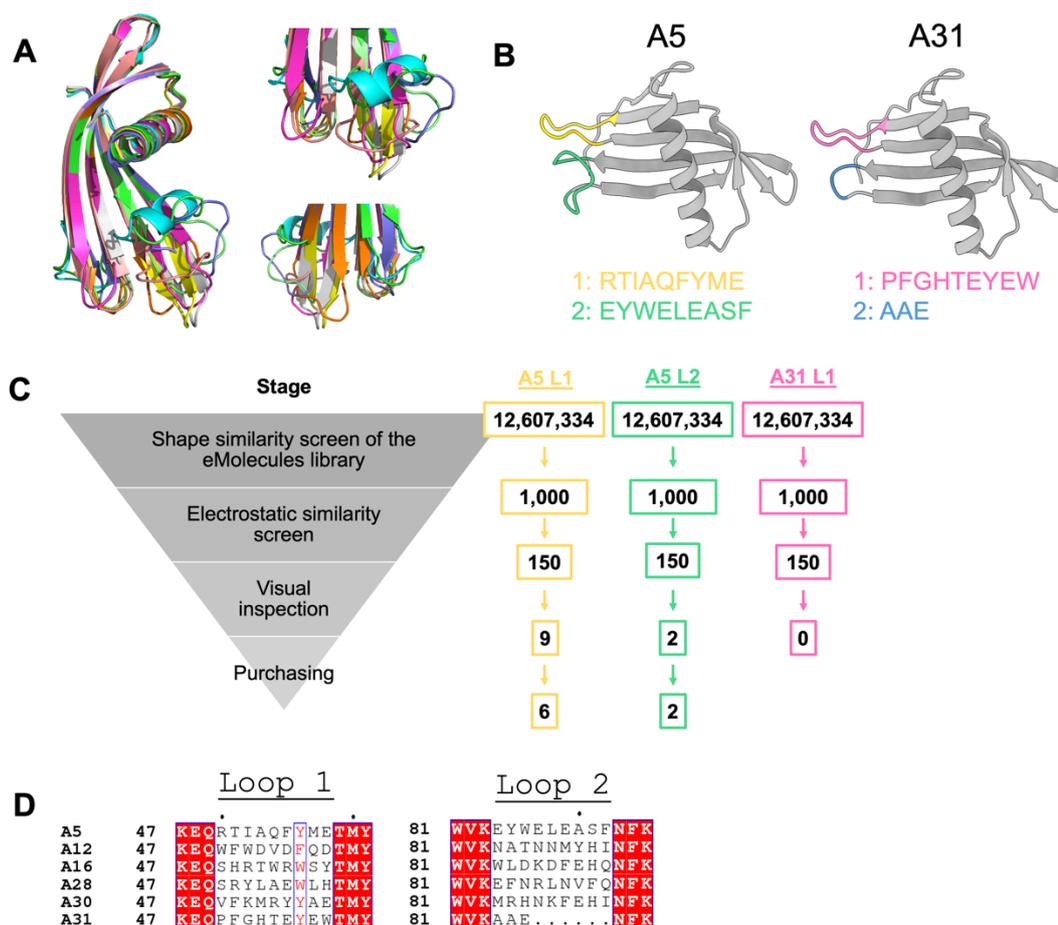


**Figure 3.16 Structure of inhibitory Affimer A5 bound to hemagglutinin (HA) generated by cryo-EM using rapid grid preparation.** A) Image processing workflow performed in Relion showing how movies of HA and A5 were processed into a final map with global resolution of 4.4 Å. B) Fourier shell correlation of the final HA + A5 map. C) Angular distributions of particles for the final HA + A5 reconstruction. D) Local resolution of the final HA + A5 map. E) Model of HA and A5 generated by rigid body fitting into the cryo-EM density map. HA1 and HA2 subunits shown as blue and purple cartoons, respectively. A5 shown as green cartoons. The same colour scheme is used throughout the figure. F) A5 binding site at the HA1 receptor binding domain. G) A5 and A31 overlaid based on an alignment of HA co-ordinates using the Matchmaker function in ChimeraX. A31 shown as an orange cartoon.

### **3.3.3.4 Software to predict shape and electrostatic similarity identified small molecule mimics of HA-Affimer**

The low resolution in the Affimer-HA interface in both the A5- and A31-bound HA maps meant that unlike for NP-Affimer, the exact region of the Affimer loops responsible for HA binding could not be determined. In order to generate mimics of A5 and A31, the Affimer structures were first generated via homology modelling using an anti-Bcl-XL Affimer<sup>224</sup> (PDB: 6HJL) as a template, since this Affimer had the highest sequence similarity to both A5 and A31. Homology models were chosen over the models generated from the low-resolution cryo-EM maps to show that this Affimer-guided method could even be applied to projects without an Affimer-bound structure. All Affimers have identical scaffold structures, with the variation in sequence and structure found in their hypervariable loops (Figure 3.17A). Therefore, any Affimer structure could be used as the homology model template as long as it puts the loop is modelled within the constraints of an Affimer scaffold.

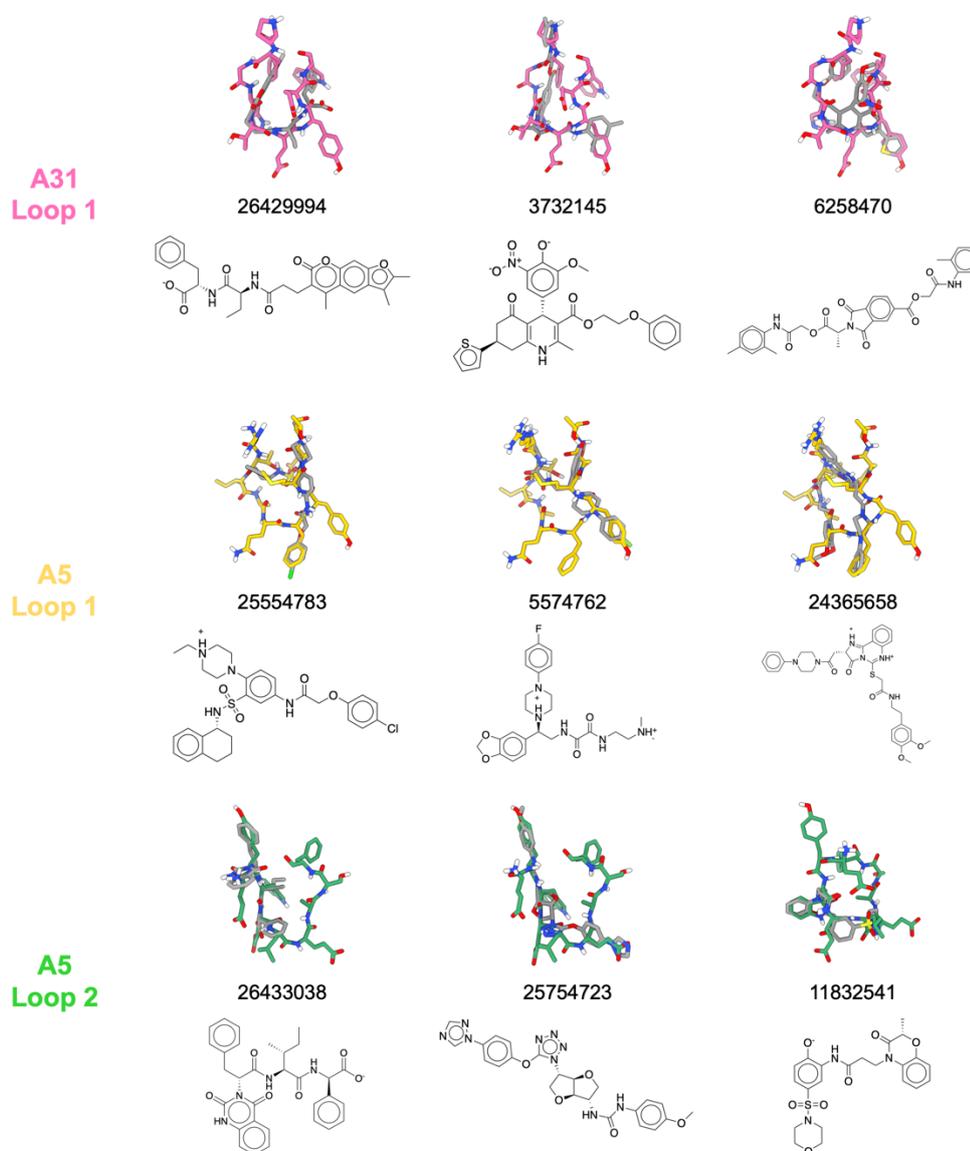
The loop regions of the A5 and A31 homology models (Figure 3.17B), except for A31 loop 2 which is a short, likely non-functional loop, were excised and used as templates for the same virtual screening workflow used for NP-Affimer (Figure 3.17C). The lack of structural information on the key Affimer residues responsible for HA binding, including the H bond donors and acceptors to mimic, meant the rules used to identify promising compounds were more general than for the NP-Affimer mimics. The only residue-level information came from the sequence alignment of the six Affimers identified as binders following phage display, including A5 and A31. The alignment identified that loop one position seven was conserved in all Affimer clones as a hydrophobic residue, which is likely important for binding and so should be mimicked by downstream small molecules (Figure 3.17D). More generally, good compounds should have the same shape and orientation as the loop, mimic at least one side chain and potential hydrogen bond donor or acceptor, not contain large hydrophobic regions, not contain flexible chains and have a central aromatic group to stabilise the conformation of the compound.



**Figure 3.17 Development of small molecule mimics of Affimers which bind and inhibit influenza A virus hemagglutinin (HA).** Structural alignment of Affimers isolated against SUMO1 (bright green), SUMO2 (cyan), p300 (pink), KRAS (yellow), K33 DUB (peach), Fc gamma receptor (grey), Bcl-XL (purple), Bcl-2 (orange) and CH alpha actinin 2 (pale green). Image generated in PyMol<sup>216</sup>. B) Homology models and loop sequences of A5 and A31 generated based on a template Affimer that binds Bcl-XL. A5 loop 1, A5 loop 2, A31 loop 1 and A31 loop 2 are coloured in yellow, green, pink and blue, respectively. This colour scheme is used throughout. Images generated in ChimeraX<sup>149</sup>. C) Ligand-based virtual screening workflow used to identify compounds from the eMolecules library which are structurally similar to either A5 loop 1, A5 loop 2 and A31 loop 1. D) Sequence alignment of the two hypervariable loops in six Affimers which bound HA. The six Affimers were identified as binders following screening by phage display, but only A5 and A31 were taken on for further characterisation. Sequence identity indicated by a filled red box and sequence similarity indicated by red text. Image generated using the eSPRIPT server<sup>233,234</sup>.

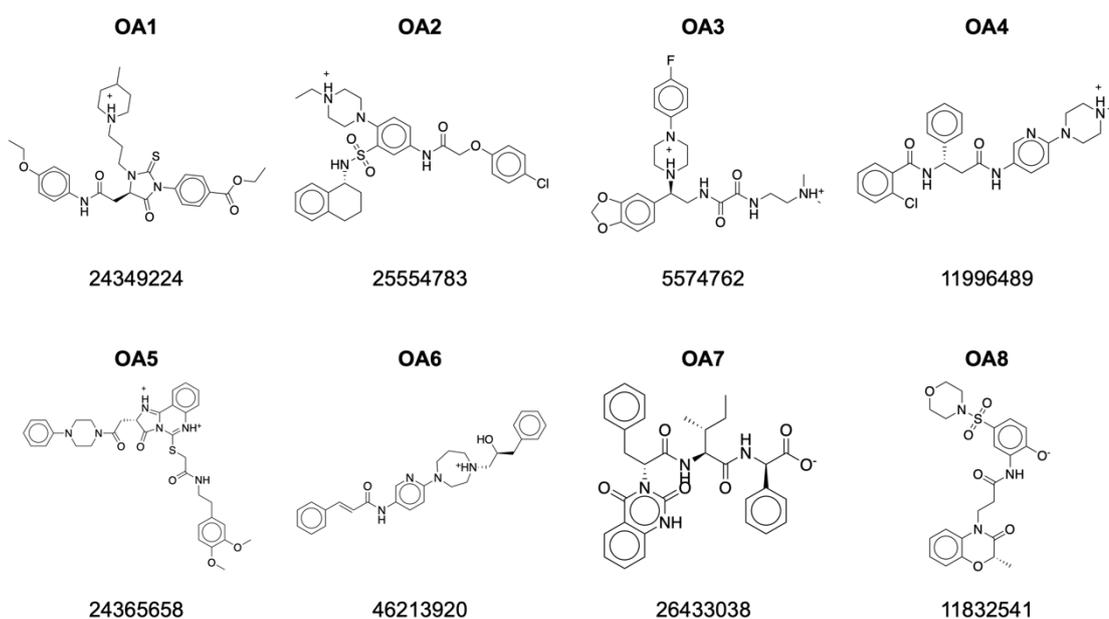
Visual inspection of the top scoring compounds for A31 loop 1 showed that few mimicked the loop well (Figure 3.18). This is likely because the loop was composed of large hydrophobic residues like tyrosine, tryptophan and phenylalanine, which cannot be easily mimicked by small molecules <500 Da. Most high scoring compounds like 26429994 accurately mimicked the main chain shape but did not extend into the side chains, and those which did mimic side chains did not mimic them accurately. For example, of the most promising A31 loop 1 mimics, 6258470 mimics the conserved Y56 with a di-methyl phenyl group without any polar group to mimic the hydroxyl, and although 3732145 mimics E57 well with a nitro group, its thiophene ring does not mimic Y56 well. Overall, this suggests the chemical space did not exist in the eMolecules library to accurately mimic this loop and so no A31 loop 1 mimics were purchased.

Conversely, compounds screened against A5 loop 1 and loop 2 mimicked the template more accurately. For example, 25554783 mimicked A5 loop 1, with a sulphonyl group that aligned with E58's carboxyl group and a fluorinated aromatic ring overlaid with F55. 5574762 also had a fluorinated aromatic ring which mimicked the conserved Y56 well. Finally, 24365658 has a non-conjugated benzene ring that overlaid well with F55. For A5 loop 2, the best mimics covered less of the loop but mimicked it well. For example, 26433038 only covered the first third of the compound, but it mimicked Y51 with a benzene ring and consecutive amide groups overlaid with the main chain between Y51 and E53. 25754723 mimicked T51 and E55 well with a methoxy benzene and 1,2,4-triazole groups, respectively. Finally, 11832541 mimicked W52 almost identically with a conjugated double ring structure and also has amide groups that mimic the main chain well.



**Figure 3.18** Example small molecule mimics of Affimers A5 and A31 isolated against the hemagglutinin (HA) spike protein from influenza A virus. A5 loop 1, A5 loop 2, A31 loop 1 and small molecules shown as yellow, green, pink and grey sticks, respectively.

Because there was no available structural information about the binding of A5 to HA, compounds were purchased so that together they gave maximum coverage of each loop. This approach tried to ensure that at least one compound would cover the region responsible for binding. For example, whereas 25554783 mimics F55, 5574762 mimics F56 instead; and whereas 26433038 covers the first third of loop 2, 11832541 covers the second third. Compounds that were too expensive we also discarded, and so this resulted in eight compounds, six mimicking loop one and two mimicking loop two, which were purchased from eMolecules and renamed OA1-OA8 (Figure 3.19).



**Figure 3.19** Eight small molecules identified by ligand-based virtual screen that mimic an Affimer isolated against hemagglutinin (HA)

### 3.4 Discussion

In this chapter, a novel method for small molecule discovery that uses Affimers as templates in a LBVS workflow was developed and applied to two Affimer molecules. Regarding NP-Affimer, which inhibits CCHFV NP, a previously solved crystal structure was used to focus the search space to a core binding region of five Affimer residues which mediate the majority of NP:Affimer interactions. This pentapeptide was used as a template for LBVS using programs for 3D shape similarity and electrostatic similarity. From the top scorers, promising mimics were identified based on a set of rules guided by features in the crystal structure, leading to the AF1 compound series. The binding modes were predicted for these compounds using a SBVS workflow which, along with a program for binding site prediction, suggested that the adjacent RNA-binding site was more appropriate than the Affimer-binding site for targeting with small molecules.

Therefore, a traditional SBVS workflow was performed at the RNA-binding site, leading to the AF2 compound series.

To determine the structures of Affimers A5 and A31, two Affimer inhibitors of the IAV spike protein HA, bound to HA, a bespoke method for rapid cryo-EM grid preparation was employed. However, the low resolution in the reconstruction, particularly at the HA-Affimer interface, meant the structures could not be used to focus the search space down. Instead, the Affimer structures were generated by homology modelling as a way to demonstrate the application of the method to projects with no experimental structures available. The entirety of the A5 and A31 binding loops were fed into the LBVS workflow, which led to the OA compound series.

Overall, this chapter demonstrates the broad range of *in silico* work that can be employed for generating small molecule inhibitors. It shows how protein structures like the NP:NP-Affimer structure can guide LBVS and conversely, how SBVS can help to validate compounds generated by LBVS. Additionally, generating the AF1 and AF2 series of compounds independently through LBVS and SBVS, respectively, allows for the direct comparison of the two complementary techniques. Subsequent compound testing (described in chapter 4) will identify whether the RNA site is a useful target for inhibiting with small molecules and whether there are differences in the properties of hits from the two series.

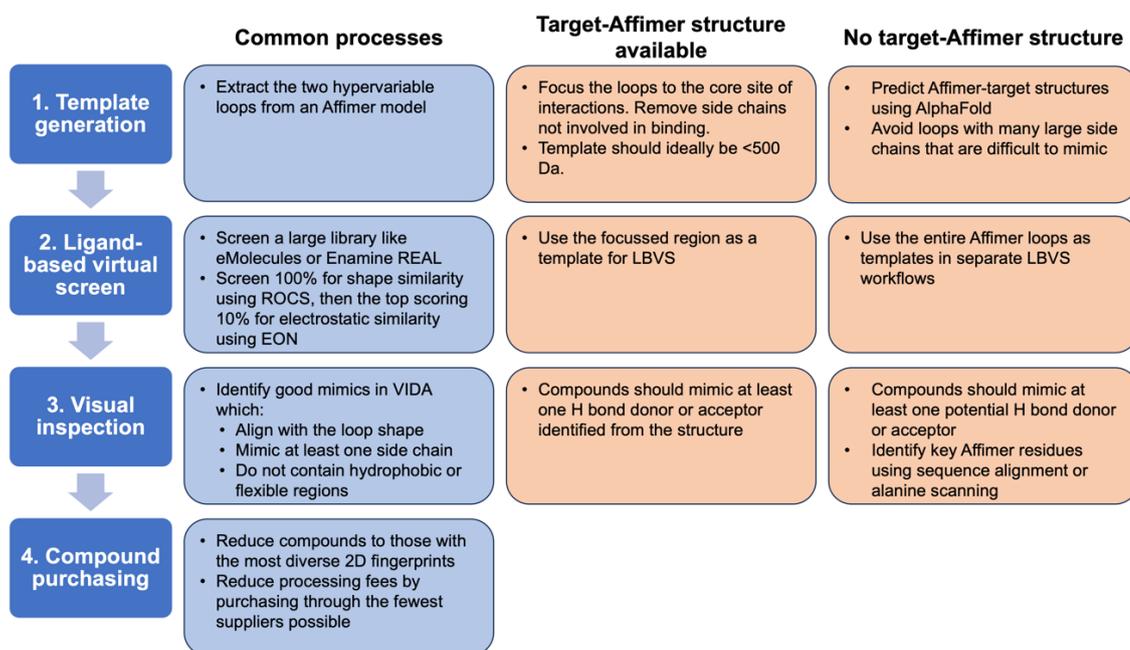
The NP:NP-Affimer structure was useful for focussing the search space, but it could have been reduced further. Although K109 is within the region of NP-Affimer responsible for binding, it points away from NP and does not interact with it. Therefore, the side chain of K109 did not require restoration *in silico* and could have been removed entirely prior to the LBVS. This was not so important here because none of the AF1 series mimicked K109, but in future Affimer-guided LBVS, non-binding residues should be replaced with a glycine to focus mimic development even further.

Although the new method for rapid cryo-EM grid preparation had been applied to samples with preferred orientation previously<sup>235</sup>, it had not been applied to samples which aggregate upon binding to a partner protein. Here we show that through time-resolved cryo-EM, you can capture a protein after its partner has bound but before it aggregates, which can be a common problem when determining protein complex structures by cryo-EM. The global resolution of the A31- and A5-bound HA maps was ~4 Å and lower at the Affimers, meaning it was not high enough to guide mimic development. Since HA has been recently determined to 2.6 Å by cryo-EM<sup>236</sup>, the resolution of the HA:Affimer complexes was limited by certain factors, likely the ice thickness, since thick ice has been noted for several samples processed using the sprayer<sup>65,235</sup>. The self-wicking grids helped to thin the sample prior to freezing, but in the future, a longer wicking time between spraying and freezing might lead to greater thinning and a higher resolution HA-Affimer structure, but potentially at the expense of

more HA aggregation and preferred orientation. Another way to improve map resolution is to collect more cryo-EM data or to improve the quality of the data. For the HA + Affimer samples here, the total area of useable ice generated by the sprayer was small despite replicate grids being produced, so simply collecting more data was not possible. However, the data quality could be improved since a Selectris energy filter, which helps to boost the signal to noise ratio in micrographs, was not part of the cryo-EM set up but could be employed in future data collections. Nonetheless, even if the cryo-EM data quality was improved, the flexibility of Affimers bound at the HA exterior would likely limit the resolution here, so methods to reduce protein flexibility may be required, such as cross-linking.

The small molecules developed here were not perfect mimics of the Affimer templates, even when the template size was reduced to the core site of binding. This indicates that the chemical space was not broad enough in the eMolecules library to include accurate Affimer mimics. In future LBVS, the chemical space should be increased by screening a larger library like the Enamine REAL library of >6 billion make-on-demand compounds. This library is ~500X larger than the eMolecules library used here, so screening it would require significantly more computational power to screen, but such a scale-up could be achieved by further utilising high performance computing clusters like ARC or branching out into cloud computing systems. Alternatively, highly accurate mimics could be synthesised by a medicinal chemist, removing the reliance on the chemical space found in commercially available libraries.

For future reference, Figure 3.20 shows an example workflow to follow in order to generate small molecule mimics of an Affimer using LBVS, whether or not there is an Affimer-bound target structure to guide this process.



**Figure 3.20 Workflow for generating small molecule mimics of Affimer proteins depending on whether there is a target:Affimer structure to guide mimic development.**

# **Chapter 4 : Characterising small molecule Affimer mimics in vitro**

## 4.1 Introduction

In chapter 3, it was shown how protein structures can inform ligand-based and structure-based virtual screening to develop small molecule inhibitors of proteins. A LBVS workflow was developed to identify small molecules termed the AF1 series which mimic the binding region of NP-Affimer, an Affimer which inhibits NP from the tick-borne virus CCHFV<sup>86</sup>. Computational modelling predicted that the RNA-binding site, which is adjacent to the Affimer-binding site, was a better target for small molecules, so SBVS was performed at the RNA-binding site, leading to the AF2 series. The same LBVS workflow that led to the AF1 series was applied to the Affimer A5 which inhibits the influenza virus HA spike protein<sup>157</sup>. Lack of a high-resolution HA-A5 structure meant the exact site of A5 binding was unknown, so the mimics were developed based on the entirety of the two Affimer binding loops derived from a homology model, leading to the OA compound series.

VS predicts whether a compound binds to a target, but these predictions are based on simplified models. For example, most SBVS uses a static protein structure and a limited number of compound conformations to speed up binding calculations. Therefore, all three compound series require testing to evaluate whether they bind to their viral targets and inhibit the corresponding viral processes.

In this chapter, I will demonstrate how hits were identified from the compound series using a pipeline of screening assays. For the AF1 and AF2 compounds developed through LBVS and SBVS, respectively, this allowed for direct comparison of the two *in silico* methods for one system. It also allowed for comparison of the Affimer-guided method when the mimics were guided by a target-Affimer structure or not.

## 4.2 Materials and methods

Section 4.2.1 describes the basis of the techniques employed in this chapter and sections 4.2.3-4.2.16 provide detailed methods used in this chapter.

### 4.2.1 Screening assays background

#### 4.2.1.1 Cytotoxicity assays

Measuring compound toxicity is important because cell death may present a false positive or negative result in a downstream cellular assay. In this chapter, cytotoxicity is measured using the ATPlite kit (Perkin Elmer) which measures the concentration of adenosine triphosphate (ATP) in cells. The intracellular ATP concentration declines rapidly when cells die, making it an accurate marker for cell viability. The concentration of ATP is measured by the ATP-dependent conversion of the substrate luciferin to oxyluciferin by the enzyme luciferase (Equation 4:1)<sup>237</sup>. This reaction generates light

with a peak emission at 560nm as a by-product, meaning luminescence is used as a surrogate for ATP concentration.



**Equation 4:1 Conversion of luciferin to oxyluciferin by luciferase**

During compound screening, conditions in the cytotoxicity assay were kept consistent with the downstream cellular assay, by growing the same cell line in media containing compound at the same concentration. After the cells were incubated for a length of time also consistent with the downstream cellular assay, the intracellular ATP concentration was analysed. Cells were lysed in cell lysis buffer, releasing intracellular ATP into the growth media, then a substrate solution containing the luciferase enzyme and substrates described in Equation 4:1 was added following the manufacturer's instructions. After incubation, the measured luminescence for each well containing compound was normalised to a control well that was incubated with an equivalent DMSO concentration, because DMSO causes low levels of toxicity that must be accounted for.

#### **4.2.1.2 TCID<sub>50</sub> assays**

50% Tissue culture infectious dose (TCID<sub>50</sub>) assays are used to quantify the viral titre, meaning the concentration of infectious viruses in a sample<sup>238</sup>. Varying virus dilutions are added to cells and incubated until a time that cytopathic effects (CPEs) are detected, which are the consequence of cell death upon virus infection. A TCID<sub>50</sub> value is calculated which is the virus dilution required to induce CPEs in 50% of cells. For compound screening, compounds are added to a dilution of virus just higher than its TCID<sub>50</sub>, so that virus control wells present CPEs, but inhibition of viral infection by a compound prevents CPEs, leading to cell survival. TCID<sub>50</sub> assays are high-throughput, since they can be performed in a 96-well plate, and show a direct impact on virus infection but they do not reveal which stage of the virus lifecycle is blocked by the compound or any information on how compounds bind to the protein of interest.

#### **4.2.1.3 Mini-genome system assay**

CCHFV is a biosafety level (BSL) 4 pathogen, meaning that live CCHFV can only be studied in high containment facilities. A CCHFV mini genome system (MGS) was developed for the studying of CCHFV replication without live CCHFV, in which cells are transfected with three DNA plasmids which code for the NP, the RNA-dependent RNA L polymerase and an RNA segment containing the negative sense of the open reading frame (ORF) for green fluorescent protein (GFP) (Figure 4.1)<sup>86</sup>. When all three plasmids are transcribed and translated, the GFP RNA is packaged into RNPs with NP and L. The L polymerase uses the RNA as a template to make new RNA with the GFP ORF in the positive sense, which is translated by ribosomes to make GFP. Therefore,

GFP expression is a surrogate for CCHFV gene expression without making any live CCHFV, so this system can be used under more accessible BSL2 tissue culture conditions. Compounds can then be tested against this system, with a reduction in green fluorescence corresponding to a reduction in viral infection. The cells used for the mini-genome system are BSR-T7 cells, which is a stable cell line of BHK-21 cells that express the T7 polymerase, since the three mini-genome system plasmids are under the control of the T7 promoter.

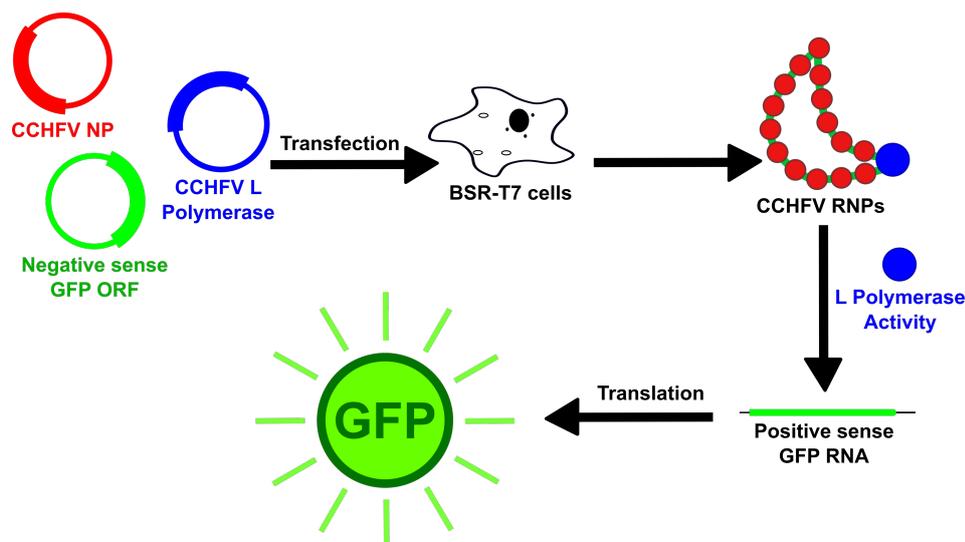


Figure 4.1 A schematic for the CCHFV mini-genome system.

#### 4.2.1.4 Surface Plasmon Resonance spectroscopy

Surface plasmon resonance (SPR) is an optical effect that occurs when plane-polarized light is shined at a gold surface<sup>239</sup>. SPR spectroscopy is a biophysical technique to study protein-ligand binding whereby one molecule is conjugated to the gold surface and another molecule termed the analyte is flowed over the surface<sup>239,240</sup>. If the analyte binds to the immobilised molecule, then the refractive index of the chip changes, altering the SPR signal. The binding signal is calculated by subtracting the signal in a reference flow cell from the signal for the flow cell containing the immobilised target. The change in binding signal over time can be used to measure the binding kinetics through the  $k_{on}$  and  $k_{off}$  rates of association and dissociation and the  $K_d$  binding affinity for the ligand-analyte complex can be calculated by measuring signals as a function of analyte concentration.

A common method for immobilisation of a protein onto the gold surface is to use a layer of streptavidin molecules which bind to a biotinylated target. The biotin can either be chemically added or added biologically by the recombinant introduction of a biotin acceptor protein (BAP) tag with the amino acid sequence of GLNDIFEAQKIEWHE into

the protein <sup>241</sup>. The central lysine residue is biotinylated by BirA during protein expression in specific *E. coli* cells named CVB101 cells.

#### **4.2.1.5 Thermal shift assays by differential scanning fluorimetry**

Thermal shift assays (TSAs) can identify target-ligand binding by measuring the melting temperature ( $T_m$ ) of a target and its shift caused by a ligand binding and stabilising the target. One way to study protein unfolding is differential scanning fluorimetry (DSF), whereby changes in fluorescence are measured as a function of temperature <sup>242</sup>. Fluorescence can be intrinsic or extrinsic to the protein target. Extrinsic fluorescence requires binding of a fluorescent dye to the target and the most commonly used fluorescent protein dye being SYPRO Orange, which fluoresces when bound to the hydrophobic portions of proteins that are exposed upon protein unfolding. SYPRO orange fluorescence can be detected using filters in standard real-time PCR machines, increasing the method's accessibility. Typically, the derivative is taken for the protein melting curve which generates a melting peak from which the  $T_m$  is taken. The  $T_m$  in a ligand-containing well is compared to the  $T_m$  in a control well which contains a matched DMSO concentration.

#### **4.2.1.6 Thermal shift assays by Nano Differential Scanning Fluorimetry**

NanoDSF is similar to DSF but measures intrinsic protein fluorescence <sup>242</sup>. Tryptophan residues intrinsically fluoresce at 350 nm when natively folded but this shifts to 330 nm when the protein is unfolded <sup>242</sup>. Therefore, the tryptophan emission wavelength can be used as a measure of protein unfolding by plotting the ratio of (emission at 330nm/emission at 350nm), which generates a melting curve similar to that created using DSF. This is also usually converted to the derivative to calculate the  $T_m$ .

#### **4.2.1.7 <sup>19</sup>F-NMR**

The lack of fluorine atoms in proteins makes <sup>19</sup>F Nuclear Magnetic Resonance (NMR) relatively simple to interpret, as signals are only observed for the fluorine-containing ligand, compared to that of proton NMR where protein background noise often overwhelms any ligand signals <sup>243</sup>. Furthermore, the naturally >99% abundant fluorine-19 nuclei have a spin of 1/2, a wider chemical shift range than protons, and a nuclei sensitivity of 83% of that of a proton. Therefore, in a mixture of protein, buffer and fluorinated compound, the only <sup>19</sup>F-NMR peak will derive from the compound, making it a highly sensitive method with limited routes for interference. When a fluorinated compound binds to a protein, its <sup>19</sup>F-NMR signal is broadened compared to when the compound is in isolation, meaning <sup>19</sup>F-NMR can be used as a screening technique to measure compound binding. However, NMR is performed in ~300  $\mu$ L capillary tubes, meaning the protein requirements are high and the throughput is low compared to other plate-based biophysical methods.

## 4.2.2 Materials

### 4.2.2.1 Plasmids

Table 4.1 Plasmids used in this chapter

Plasmid name	Plasmid amplification		Gene expression		
	Bacterial strain	Antibiotic resistance	Expression system	Promoter	Gene expressed
pETSUMO NP-Affimer	DH5α	Kanamycin	<i>E. coli</i> : Rosetta2	T7	6x His-SUMO- NP-Affimer
pET28a CCHFV NP	DH5α	Kanamycin	<i>E. coli</i> : Rosetta2	T7	6x His-SUMO- CCHFV NP
pET28a CCHFV NP-BAP	DH5α	Kanamycin	<i>E. coli</i> : CVB101	T7	6x His-SUMO- CCHFV NP-BAP
CCHFV Sseg UTRs-eGFP	DH5α	Ampicillin	Mammalian: BSR-T7	T7	eGFP (negative sense) flanked by CCHFV S seg UTRs
CCHFV-NP-support	DH5α	Kanamycin	Mammalian: BSR-T7	T7	CCHFV NP
CCHFV-L-support	DH5α	Spectinomycin	Mammalian: BSR-T7	T7	CCHFV L

Table 4.2 Purified proteins used in this chapter

Protein	Virus	Strain	Supplier	Uses
Trimeric HA	IAV	Hong Kong/1/1968 (H3N2)	Wilson Group	Cryo-EM
Trimeric HA	IAV	Aichi/2/1968 (H3N2)	My Bio Source	Screening assays
Monomeric HA	IAV	Aichi/2/1968 (H3N2)	Sino Biological	Screening assays
NP-BAP	CCHFV	Baghdad-12	In-house	SPR
NP	CCHFV	Baghdad-12	In-house	Thermal shift assays, crystallography

### 4.2.3 Cytotoxicity assay

In general the assay for assessing compound cytotoxicity followed the workflow shown in Figure 4.2.

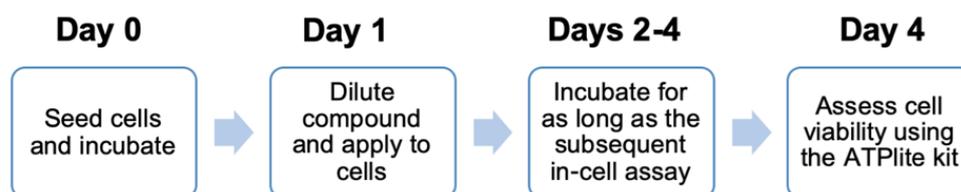


Figure 4.2 Workflow for small molecule cytotoxicity assay

#### 4.2.3.1 Screening OA compounds

MDCK cells were seeded at  $0.5 \times 10^5$  cells/well in a 96-well plate and incubated at 37°C and 5% CO<sub>2</sub> for 16-24 hours until they were ~80% confluent. The OA compounds were serially diluted 1 in 2 in infection media (DMEM, 0.1% TPCK trypsin) across a 96-well plate from 1 mM to 488 nM, and dispensed on to the MDCK cells. The cells were incubated with the compounds for 3 days, after which the cell viability was measured using the ATPlite kit (Perkin Elmer). Cells were lysed using the lysis buffer and then ATP-dependent luminescence was initiated by adding the substrate buffer. The luminescence in each well was analysed using a FLUOstar OPTIMA fluorescent plate reader (BMG Labtech).

#### 4.2.3.2 Screening AF compounds

BSR-T7 cells were seeded in a 96-well plate and incubated at 37°C and 5% CO<sub>2</sub> for 16-24 hours until they were ~80% confluent. Compounds were diluted to 10 μM in 100 μL growth media (DMEM + 1% Pen/Strep + 10% FBS) and applied to cells. The cells were incubated with the compounds for 3 days. Finally, cell viability was measured using the ATPlite kit (Perkin Elmer) and a FLUOstar OPTIMA fluorescent plate reader (BMG Labtech).

#### 4.2.4 CCHFV mini-genome system assay

The workflow for the CCHFV mini-genome system (MGS) assay is described in Figure 4.3:

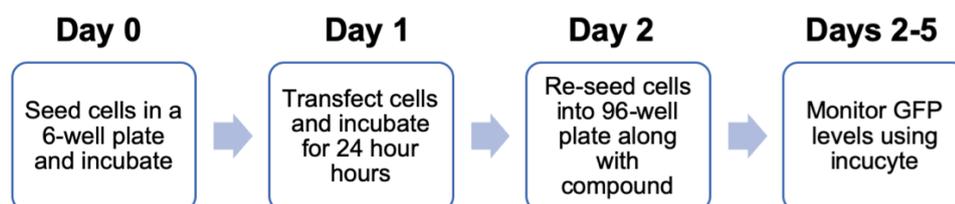


Figure 4.3 Workflow for using the CCHFV mini-genome system assay to screen compounds

BSR-T7 cells were seeded at  $9.0 \times 10^5$  cells/well in a 6-well plate and the plate was incubated at 37°C and 5% CO<sub>2</sub> for 16-24 hours until they were ~80% confluent.

For each well in a 6-well plate, 1.125 μg CCHFV L support plasmid, 375 ng CCHFV N support plasmid and 375 ng S seg UTR-eGFP plasmid were mixed gently in Gibco

Opti-MEM reduced serum media (Thermo Fisher Scientific). 2.5  $\mu$ L room-temperature Trans-IT LT1 reagent (Mirus) per  $\mu$ g plasmid DNA was mixed in with the DNA to give a final volume of 250  $\mu$ L. For the negative control wells, the L support plasmid was excluded and for the mock transfection, no DNA or Trans-IT was added. The mixture was incubated for 30 minutes at room temperature.

After the incubation, cells were washed in PBS and 1.75 mL fresh growth media (DMEM + 10% FBS + 1% PenStrep) was dispensed onto each well. The transIT-LT1 Reagent:DNA complexes were added drop-wise to different areas of the wells and the wells were mixed by rocking the plate back-and-forth. The plate was incubated at 37°C in 5% CO<sub>2</sub> for 24 hours in an Incucyte S3 live-cell imager (Sartorius) to measure GFP expression.

After 24 hours, the cells were trypsinised and re-seeded into wells of a 96-well plate at  $3 \times 10^4$  cells/well in 50  $\mu$ L growth media. 50  $\mu$ L of compound at 20  $\mu$ M concentration was dispensed onto the cells to give a final compound concentration of 10  $\mu$ M, with triplicate wells per compound. A DMSO control and media control was included. The plate was incubated at 37°C in 5% CO<sub>2</sub> for 72 hours in an Incucyte S3 live-cell imager (Sartorius) to measure GFP expression which was normalised to the DMSO control.

#### **4.2.5 IAV TCID<sub>50</sub> assay**

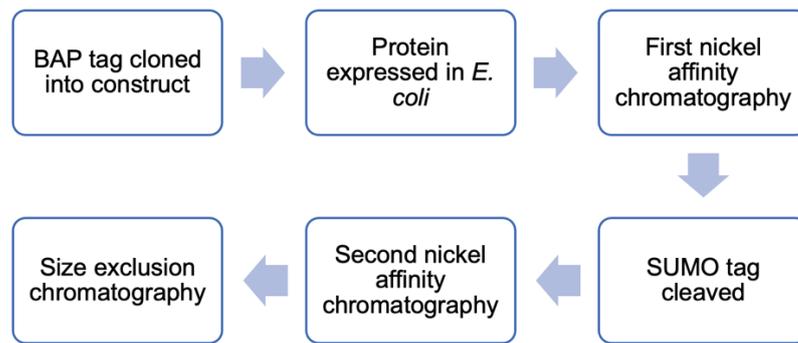
MDCK cells were seeded at  $5 \times 10^4$  cells/well in a 96-well plate and incubated at 37°C in 5% CO<sub>2</sub> for 16-24 hours until they were ~90% confluent. Starting at 2X the highest compound concentration which led to a >95% cell viability in the cytotoxicity assay, the OA compounds were serially diluted 1 in 2 in infection media (DMEM, 0.1% TPCK trypsin) across a 96-well plate, leaving 50  $\mu$ L 2X final concentrations. 50  $\mu$ L IAV (A/Aichi/1968/(H3N2)) at 200x the TCID<sub>50</sub> value for the virus stock was added to each well and mixed 1:1 with the drug dilutions.

Compound and virus were incubated for 45 minutes then applied to the MDCK cells which had been washed with PBS. Cell only wells containing no compound or virus were included as a control along with virus only wells, which contained virus but no compound.

After 72 hours, the drug/virus mixture was aspirated off the cells and the cells were washed with PBS, fixed with 4% paraformaldehyde then stained with crystal violet solution for 15 minutes. Finally, the cells were then washed with water and imaged.

#### **4.2.6 CCHFV NP-BAP protein production**

CCHFV NP-BAP was produced to facilitate immobilisation in SPR experiments. The general workflow for its production is shown in Figure 4.4.



**Figure 4.4 Workflow for generating BAP-tagged CCHFV NP**

#### 4.2.6.1 Cloning

A BAP tag with protein sequence GLNDIFEAQKIEWHE was cloned at the C-terminus of the pET-28a His-SUMO-NP expression plasmid by Dr Amy Turner. Site directed mutagenesis primers were designed to insert the full BAP sequence. The primers were as follows, with regions complementary to the NP sequence highlighted in yellow and regions coding for the BAP tag highlighted in blue:

Forward: 5' ACC AGC GCC AAC ATC ATC GGC CTG AAC GAT ATT TTT GAA GCG CAG AAA ATT GAA '3

Reverse: 5' — GTG GTG GTG CTC GAG TTA TTC ATG CCA TTC AAT TTT CTG CGC TTC AAA AAT ATC -3'

PCR was performed and the correct insertion was verified by sequencing the plasmid.

#### 4.2.6.2 NP-BAP protein expression

CVB101 *E. coli* cells, which are similar to BL21 *E. coli* cells but also contain a BirA ligase for more efficient biotinylation, were transformed with the pET28a His-SUMO-NP-BAP plasmid. The transformed cells were incubated overnight on LB agar containing 50 µg/mL kanamycin at 37°C, then stored at 4°C for no more than 2 weeks. A single colony was picked from the plate and added to a starter culture of LB media containing 50 µg/mL kanamycin. This was incubated at 37°C until the optical density (OD) at 600 nm reached 0.6-0.8 absorbance units. The expression of protein was induced at this time by addition of 500 µM isopropyl β-d-1-thiogalactopyranoside (IPTG) and the induced cultures were incubated overnight for ~16 hours at 18°C. The bacterial cells were harvested by centrifugation at 4,500g at 4°C for 30 minutes and the cells were stored at -80°C until use.

#### 4.2.6.3 Protein purification buffers

The buffers for the purification of NP-BAP included:

- Lysis buffer: 20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 0.1% (v/v) Triton X-100, 1 mM MgCl<sub>2</sub>, 5% glycerol, 1 mg/mL lysozyme from chicken

egg white (Sigma Aldrich), 1X Halt Protease Inhibitor Cocktail, 10 units/mL omnicleave.

- Nickel affinity chromatography wash buffer: 20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 5% glycerol.
- Nickel affinity chromatography elution buffer: 20 mM Tris-HCl pH 7.4, 500 mM NaCl, 300 mM Imidazole, 5% glycerol.
- Dialysis buffer: 20 mM Tris-HCl pH 7.4, 300 mM NaCl, 5% glycerol
- Size exclusion chromatography (SEC) buffer: 20 mM Tris-HCl pH 7.4, 300 mM NaCl, 5% glycerol.

#### **4.2.6.4 Nickel affinity chromatography**

Bacterial cell pellets from a 500 mL expression culture were resuspended in 10mL pre-chilled lysis buffer and incubated at 4°C for 30 minutes. Cells were lysed by sonication on ice using a Vibra cell sonicator (Sonics). Cells were sonicated at 25% amplitude for 6 minutes cycling between 10 s on and 20 s off to avoid the cells heating up. The lysed cells were centrifuged at 50,000 g for 30 mins using an Avanti J-20 XP centrifuge (Beckman Coulter) with a JA-25.50 rotor.

In a 50 mL conical tube, 3 mL washed, equilibrated Ni sepharose fast flow beads (Cytiva) were mixed with the soluble fraction and left to incubate overnight at 4°C. The bead mixture was poured into an Econo-Pac chromatography column (Bio-Rad) and the flow-through was collected. The beads were then washed with buffer containing increasing concentrations of imidazole including 20 mM, 50 mM, 100 mM and 150 mM. For every concentration, the beads were washed twice with 5 mL buffer. Four elution fractions were taken by incubating the beads with 2 mL 300 mM imidazole buffer for 5 minutes and taking the elution fraction. Fractions were analysed by SDS-PAGE and the protein concentration was estimated using the absorbance at 280nm ( $A_{280}$ ) as measured by a DS-11 nano spectrophotometer (DeNovix).

Fractions containing NP were pooled and dialysed overnight with Ulp1 protease to remove the SUMO tag. 0.5 mg Ulp1 protease (produced in house) per 10 mg His-SUMO-NP-BAP was added to the fractions. The mixture was placed into rinsed SnakeSkin dialysis tubing (Thermo scientific), sealed with clips and dialysed overnight at 4°C in 2 L dialysis buffer on a magnetic stirrer.

Following dialysis, the cleaved NP-BAP was separated from the His-SUMO tags and His-tagged Ulp1 protease using a second nickel affinity chromatography step. 2.5 mL washed and equilibrated Ni sepharose fast flow beads (Cytiva) were added to the dialysed mixture, incubated for 30 minutes at room temperature (20°C), then poured into an Econo-Pac chromatography column. The flow-through was collected, then the beads were washed with 6 mL wash buffer and bound protein was eluted using 6 mL

elution buffer. In order to identify those that contained cleaved NP-BAP, fractions from the second nickel affinity chromatography step were analysed by SDS-PAGE.

#### 4.2.6.5 Size exclusion chromatography

The fractions containing NP-BAP were pooled and concentrated for a final clean up step by size exclusion chromatography (SEC). Protein was concentrated using an Amicon ultra-15 3 kDa MWCO centrifugal concentrator (Merck) to 4.5 mL, so that the volume was less than half of the 10 mL capillary loop. Using an AKTA Pure system (Cytiva) at 4°C, the concentrated protein was injected onto a washed and equilibrated superdex 75 26/600 column (Cytiva) and eluted in SEC buffer. 2 mL fractions were collected and analysed by SDS-PAGE. Those containing NP-BAP were pooled and concentrated using an Amicon Ultra-15 10 MWCO centrifugal concentrator. The protein concentration was monitored using a nano spectrophotometer.

#### 4.2.7 CCHFV NP protein production

Wild-type CCHFV NP was produced for DSF and nanoDSF thermal shift assays and X-ray crystallography using the following workflow:



Figure 4.5 Workflow for producing wild-type CCHFV NP

##### 4.2.7.1 Protein expression

Protein expression used the same method as described in section 4.2.6.2 with a few differences. The pET His-SUMO-NP plasmid was transformed in Rosetta 2 *E. coli* cells. The starter culture was generated from glycerol stocks, with 20 mL starter culture used to inoculate 1 L of the expression culture.

##### 4.2.7.2 Protein purification buffers

The buffers for CCHFV NP purification included:

- Lysis buffer: 20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 0.1% (v/v) Triton X-100, 1 mM MgCl<sub>2</sub>, 5% glycerol, 1 mg/mL lysozyme from chicken egg white (Sigma Aldrich), 1X Halt Protease Inhibitor Cocktail, 10 units/mL Omnicleave.
- Nickel affinity chromatography wash buffer: 20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 5% glycerol.
- Nickel affinity chromatography elution buffer: 20 mM Tris-HCl pH 7.4, 500 mM NaCl, 300 mM Imidazole, 5% glycerol.

- 1<sup>st</sup> Dialysis buffer: 20 mM Tris-HCl pH 7.4, 300 mM NaCl, 5% glycerol, 2 mM DTT.
- 2<sup>nd</sup> dialysis buffer: 20 mM Tris-HCl pH 7.4, 300 mM NaCl, 5% glycerol.
- SEC buffer: 20 mM Tris-HCl pH 7.4, 300 mM NaCl.

#### 4.2.7.3 Nickel affinity chromatography

Bacterial cell pellets from 6L *E. coli* expression were resuspended in 120 mL pre-chilled lysis buffer and incubated at 4°C for 30 minutes. Cells were lysed by sonication on ice using a Vibracell sonicator (Sonics). Cells were sonicated at 35% amplitude for 8 minutes cycling between 15s on and 15s off, to avoid the cells heating up. The lysed cells were centrifuged at 50,000 g for 30 mins using an Avanti J-20 XP centrifuge (Beckman Coulter) with a JA25.50 rotor.

In 50 mL conical tubes, 12 mL washed, equilibrated Ni sepharose fast flow beads (Cytiva) were mixed with the soluble fraction and incubated overnight at 4 °C. The bead mixture was poured into an Econo-Pac chromatography column (Bio-Rad) and the flow-through was collected. The beads were washed twice with 10 mL buffer containing increasing imidazole concentrations including 20 mM, 50 mM, 75 mM and 100 mM imidazole. Four elution fractions were taken by incubating the beads with 2 mL 300 mM imidazole buffer for 5 minutes. Fractions were analysed by SDS-PAGE and the protein concentration was estimated in the fractions using the  $A_{280}$  as measured by a DS-11 nano spectrophotometer. Two extra elution fractions were taken due to the intensity of the NP band in the fourth elution.

Fractions containing NP were dialysed overnight with Ulp1 protease to remove the SUMO tag. 20 units Ulp1 protease (Sigma Aldrich) per mg His-SUMO-NP was added to the fractions. The mixture was placed into rinsed SnakeSkin tubing (Thermo scientific), sealed with clips and dialysed overnight at room temperature (20°C) in 2L dialysis buffer on a magnetic stirrer. DTT was included in the 1<sup>st</sup> dialysis buffer on the recommendation of the protease manufacturer to improve cleavage efficiency. The following morning, the DTT was removed from the buffer by dialysing the protein in the 2<sup>nd</sup> dialysis buffer without DTT.

Following dialysis, cleaved NP was separated from the His-SUMO tags and His-tagged Ulp1 protease using a second nickel affinity chromatography step. 2.5 mL (5 mL per L culture) washed and equilibrated Ni Sepharose fast flow beads (Cytiva) were added to the dialysed mixture, incubated for 30 minutes at room temperature (20°C), then poured into an Econo-Pac chromatography column. The flow-through was collected, then the beads were washed with 6 mL wash buffer and bound protein was eluted using 6 mL elution buffer. In order to identify those that contained cleaved NP, fractions from the second nickel affinity chromatography step were analysed by SDS-PAGE.

#### 4.2.7.4 Size exclusion chromatography

The relevant fractions were pooled and concentrated for a final clean up step by size exclusion chromatography. Protein was concentrated using an Amicon ultra-15 3 kDa MWCO centrifugal concentrator (Merck) to 3 mL, so that the volume was  $< \frac{1}{2}$  volume of the 10 mL capillary loop. Using an AKTA Pure system (Cytiva) at 4°C, the concentrated protein was injected onto a washed and equilibrated Superdex 75 26/600 column (Cytiva) and eluted at 1 mL/min in SEC buffer. 2 mL fractions were collected and analysed by SDS-PAGE. Those containing NP were pooled, concentrated using an Amicon Ultra-15 3 MWCO centrifugal concentrator and snap frozen in liquid nitrogen.

#### 4.2.8 NP-Affimer protein production

NP-Affimer was produced for use as a positive control for binding assays using the workflow outlined in Figure 4.6.

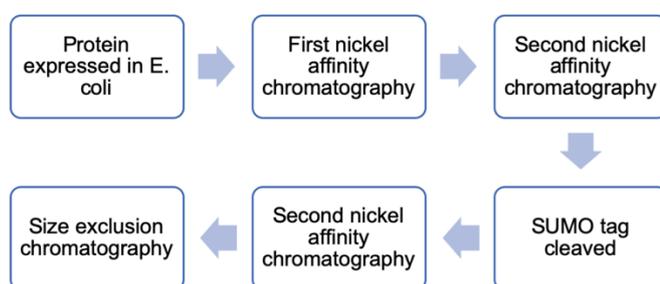


Figure 4.6 Workflow for generating NP-Affimer.

##### 4.2.8.1 Protein expression

NP-Affimer expression used the same method as described for CCHFV NP-BAP in section 4.2.6.2 except the plasmid transformed was pETSUMO NP-Affimer.

##### 4.2.8.2 Protein purification buffers

The buffers for NP-Affimer purification were as follows:

- Lysis buffer: 50 mM Na Phosphate pH 7.4, 500 mM NaCl, 20 mM Imidazole, 0.1 mg/mL lysozyme, 1X protease inhibitor cocktail, benzonase (10 U/mL), 0.1% (v/v) Triton X-100.
- First nickel affinity chromatography binding buffer: 50 mM Na Phosphate pH 7.4, 500 mM NaCl, 20 mM Imidazole.
- First nickel affinity chromatography elution buffer: 50 mM Na Phosphate, 500 mM NaCl, 300 mM Imidazole.
- Second and third nickel affinity chromatography wash buffer: 20 mM Tris pH 7.4, 500 mM NaCl, 5% glycerol, 20 mM imidazole.
- Second and third nickel affinity chromatography elution buffer: 20 mM Tris pH 7.4, 500 mM NaCl, 5% glycerol, 300 mM imidazole.
- Dialysis buffer: 20 mM Tris pH 7.4, 300 mM NaCl.

#### **4.2.8.3 First nickel affinity chromatography**

The Rosetta2 cell pellet from 1 L expression was resuspended in 20 mL lysis buffer and incubated for 30 minutes at 4°C. Cells were lysed by sonication at 30% amplitude cycling between 20 s, 20 s off for 2 minutes, then 40% amplitude cycling between 30s on 20s off for 3 minutes. The lysed cells were centrifuged at 50,000 g for 30 minutes at 4°C using an Avanti J-20 XP centrifuge with a JA-25.50 rotor.

1 mL HisPur cobalt resin (Thermo Fisher Scientific) was washed and equilibrated with binding buffer in an Econo-Pac chromatography column. The cell supernatant was incubated with the beads for 10 minutes at 4°C, then the flow-through was collected. The beads were then washed with buffer containing increasing concentrations of imidazole including 20 mM, 50 mM, and 100 mM imidazole. For each concentration, the beads were washed twice with 5 mL buffer. Four elution fractions were taken by incubating the column with 2 mL elution buffer for 5 minutes and taking the flow-through. The fractions were analysed by SDS-PAGE. Large quantities of NP-Affimer was visible in the flow-through and wash fractions, so these were snap frozen in liquid nitrogen for further purification.

#### **4.2.8.4 Second nickel affinity chromatography**

The flow-through, 20 mM wash and 50 mM wash fractions were thawed and centrifuged at 4,500g for 4°C for 10 minutes to remove aggregated protein. 10 mL washed and equilibrated Nickel fast flow sepharose beads were added to the protein and incubated overnight on a rotor mixer at 4°C. The following morning, the flow-through was taken and the beads were washed at 4°C with 20 mL binding buffer, then 20 mM binding buffer with 50 mM imidazole and binding buffer with 100 mM imidazole. Protein was eluted by incubating the beads with 5 mL elution buffer for 5 minutes and taking the flow-through. This was repeated to give four elutions.

Fractions were analysed by SDS-PAGE and the NP-Affimer concentration was measured in the fractions using a nano spectrophotometer. Fractions containing His-SUMO-NP-Affimer were pooled and placed with 20 units Ulp1 protease/mg NP-Affimer into a rinsed SnakeSkin 3500 MWCO dialysis tubing and dialysed overnight at 4°C against dialysis buffer.

#### **4.2.8.5 Third nickel affinity chromatography**

5 mL washed and equilibrated Nickel fast flow sepharose beads were put into the dialysed protein, incubated for 5 minutes at 4°C on a rotor mixer and poured into an Econo-Pac chromatography column before taking the flow-through. The beads were washed with 10 mL binding buffer and then protein was eluted with 5 mL elution buffer. Fractions were analysed by SDS-PAGE and the flow-through and wash were pooled

and concentrated with an Amicon Ultra-15 3 MWCO centrifugal concentrator (Merck) to 13.45 mg/mL before snap freezing in liquid nitrogen.

## **4.2.9 Surface Plasmon Resonance spectroscopy**

### **4.2.9.1 IAV HA screening**

All SPR analysis of IAV HA was performed using a Biacore 3000 (Cytiva) on a streptavidin-coated SPR chip (Cytiva) with a running buffer of PBS + 0.05% (v/v) Tween-20. Firstly, a flow cell was derivatised with A5 by injecting 1 nM chemically biotinylated A5 on to a flow cell at 5  $\mu$ L/min until an increase of 80 RUs was measured. Then, HA was titrated over the flow-cell to test for binding. HA was injected onto the A5 flow cell for 180 seconds, then protein was washed off for 650 seconds, with a flow rate of 20  $\mu$ L/min used for all steps.

50 nM HA + 100  $\mu$ M OA3 and 100  $\mu$ M OA3 alone were flowed over the flow-cell. When OA3 was tested, 1% DMSO was included in the running buffer to match the concentration in the sample and reduce buffer mismatch. The BiaEvaluation software (Cytiva) was used to subtract the reference flow cell sensogram from the A5 flow cell sensogram for each run.

### **4.2.9.2 CCHFV NP screening**

All SPR analysis of NP was performed using a Biacore T200 (Cytiva) on a streptavidin-coated SPR chip (Cytiva). The running buffer used was PBS + 0.05% (v/v) Tween-20 (PBST) + 1% DMSO. Firstly, a flow cell was derivatised with NP-BAP. The flow cell was derivatised to saturation by injecting 4000 response units (RU) worth of NP-BAP using 10 nM, 100 nM and 200 nM sample and a flow rate of 5  $\mu$ L/min

Then, NP-Affimer was titrated over the flow-cell as a positive control for binding. NP-Affimer at various concentrations was injected over the NP flow cell for 60 seconds and dissociated for 300 seconds at a flow rate of 30  $\mu$ L/min. Once NP:NP-Affimer binding was identified, compounds at 100  $\mu$ M were screened for binding to the NP flow cell. Sensogram analysis was performed automatically in the Biacore T200 software (Cytiva).

### **4.2.10 Thermal shift assays by differential scanning fluorimetry**

2  $\mu$ M CCHFV NP or 0.5  $\mu$ M monomeric HA (Sino Biological), 200  $\mu$ M compound and 1X SYPRO orange dye (Invitrogen) were diluted with buffer (20 mM Tris pH 7.4, 300 mM NaCl) for duplicate wells with 25  $\mu$ L volume. 20  $\mu$ M NP-Affimer and 10  $\mu$ M A5 were used as positive controls. A 2% DMSO control was included for normalising compound-containing wells. Ligand-only controls without protein were used to identify compound interference, with the threshold for interfering compounds of  $\pm$ 40,000 fluorescent counts. This was selected because it was 10% of the fluorescence peak caused by NP

melting, meaning fluorescence from compounds below this threshold was negligible compared to that from protein unfolding.

Well contents were dispensed into a white 96-well PCR plate (Thermo Fisher Scientific), which was sealed with ThermoSeal RT2RR film and placed in a QuantStudio 3 Real-Time PCR system (Thermo Fisher Scientific). The plate was heated at 1 °C/minute and the fluorescence from the SYPRO orange dye was recorded. Data were analysed in the Protein Thermal Shift software (Thermo Fisher Scientific). The raw data from duplicate wells were averaged and the negative derivative of fluorescence plots was calculated. The inflection points in the derivative plots which corresponded to the target protein with or without ligands was used to identify shifts caused by ligand binding.

#### **4.2.11 Thermal shift assays by nanoDSF**

5 µM NP or 1 µM monomeric HA, and 100 µM compound were diluted to 25 µL with buffer (20 mM Tris pH 7.4, 300 mM NaCl), which was enough to fill duplicate tubes. 3 µM NP-Affimer and 1.5 µM A5 were used as positive controls for binding. A 1% DMSO control was used for normalising compound-containing conditions.

Tubes were loaded, placed in a Prometheus NT.48 nanoDSF machine at the University of York. The tubes were heated from 20-95 °C at a rate of 1 °C/minute and the fluorescence was measured using the highest detector energy sensitivity of 100%. The raw fluorescence data from duplicate capillaries were averaged and the derivative of the fluorescence plots was calculated. The inflection points in the derivative plots which corresponded to the target protein with or without ligands was used to identify shifts caused by ligand binding.

#### **4.2.12 X-ray crystallography**

350 µM CCHFV NP was mixed with 350 µM AF1-03 and incubated for 30 minutes at 4 °C. Within wells in a sitting drop MRC crystallisation plate, using a mosquito, the protein:compound solution was mixed 1:1 with wells from the following crystallisation suites: JCSG suites I-IV, PACT, Morpheus. Undiluted crystallisation suite was dispensed into the reservoir well. The trays were sealed and incubated at 20 °C.

The crystal from a well containing 0.1 M SPG buffer pH 6.0 + 25 % (w/v) PEG 1500 was picked and shot at the Diamond Light Source. It was processed into a diffraction pattern using the Xia2 dials ispyb program. An electron density map was calculated based on the diffraction pattern via molecular replacement using MOLREP of the CCP4 suite aided by a previously solved structure of CCHFV NP (PDB: 4AKL).

#### **4.2.13 <sup>19</sup>F CPMG NMR**

Fresh OA3 compound was purchased and diluted to 10 mM in deuterated DMSO (Sigma Aldrich).

For the trimeric HA screen, a sample of 75.5  $\mu\text{M}$  OA3, 0.252  $\mu\text{M}$  trimeric HA (0.757 HA monomer) (My BioScience) and 10%  $\text{D}_2\text{O}$  was diluted to a final volume of 250  $\mu\text{L}$  in PBS. For the monomeric HA screen, a sample of 75.7  $\mu\text{M}$  OA3, 0.757  $\mu\text{M}$  monomeric HA (Sino Biological) and 10%  $\text{D}_2\text{O}$  were mixed and diluted to a final volume of 250  $\mu\text{L}$  PBS. Separate samples were prepared without HA as compound-only controls.

Tubes were scanned using a 4-channel Bruker AV4 NEO 11.75 T (500 MHz  $^1\text{H}$ ) NMR spectrometer (500-4C) fitted with a TBO probe. All experiments were run with 1024 scans at 298K. Values for chemical shifts ( $\delta$ ) were recorded in parts per million (ppm) and referenced to  $\text{D}_2\text{O}$ .  $^{19}\text{F}$  NMR were run with  $^1\text{H}$  decoupling. A paired t test was performed in Prism 9 (Graphpad) on the peak intensities.

#### **4.2.14 Cryo-EM analysis of HA and OA3**

##### **4.2.14.1 Grid preparation**

2.96 mg/mL (16  $\mu\text{M}$ ) Trimeric HA (A/Hong Kong/1968) (Produced by the Wilson lab)<sup>244</sup>, 100  $\mu\text{M}$  OA3, 0.2% Octyl  $\beta$ -D-glucopyranoside (OG) detergent were mixed and diluted to 20  $\mu\text{L}$  with TBS buffer. Self-wicking carbon cryo-EM grids (SPT Labtech) were glow discharged using a Cressington manual glow discharger for 80 seconds with a current of 15 mA.

The HA-OA3 sample was sprayed onto the glow discharged grids using a bespoke sprayer for rapid cryo-EM grid preparation outlined in Kontziampasis *et al.*<sup>65</sup>. TA volume of three clicks (3000 steps) of protein were taken up which was enough to make three grids. 800 steps of protein were sprayed per second with 1 second of spray and a pre-spray time of 0.8 seconds. The plunging arm pressure was 16 psi.

##### **4.2.14.2 Data collection**

The grids were screened and images collected using a Titan Krios TEM (Thermo Fisher Scientific) fitted with a Falcon 4 camera. In total, 5466 movies were taken at a nominal magnification of 96,000x, a pixel size of 0.83  $\text{\AA}/\text{pixel}$  and a defocus range of -1.7 to -3.2  $\mu\text{m}$ . A total dose of 56.2  $\text{e}^-/\text{\AA}^2$  was applied to the movies over an exposure time of 5.99 seconds, corresponding to a dose per  $\text{\AA}^2/\text{second}$  of 9.4. The movies were split into 43 frames, giving a dose per frame of 1.3  $\text{e}^-/\text{\AA}^2$ .

##### **4.2.14.3 Image processing**

All image processing was performed in RELION 3.1<sup>140</sup>. For pre-processing, beam-induced motion correction was performed on the movies using RELION's own implementation of MotionCor2<sup>137</sup>, then the contrast transfer function (CTF) of each

micrograph was estimated using CTFFIND-4.1<sup>138,141</sup>. The particles were picked on the motion corrected micrographs by crYOLO 1.6.1 (Sphire)<sup>139</sup> using the weights from its general model and a picking threshold of 0.1. In total 249,532 particles were picked and extracted from the CTF corrected micrographs and classified in 2D. Classes containing HA underwent 3D classification using a previously solved HA (A/Aichi/2/1968(H3N2)) structure (PDB: 5HMG) as a 3D reference filtered to 60 Å resolution to avoid introducing model bias. The best class which contained 64,682 particles was refined using RELION's auto-refine program with C3 symmetry to give a map with a global resolution of 4.4 Å. After two rounds of both Bayesian Polishing and per-particle CTF refinement, the resolution had improved further to 3.8 Å.

#### **4.2.14.4 Model refinement**

A model of HA (A/Hong Kong/1/1968 (H3N2)) (PDB: 6CEX) was docked into the final sharpened map in ChimeraX<sup>149</sup> and further refined using the all-atom refine in Coot<sup>147</sup>. Model refinement was performed using Refmac-Servalcat<sup>145</sup>.

#### **4.2.15 AF1-03 Analogue development**

The ligand-based shape matching software ROCS<sup>213</sup> and EON<sup>214</sup> were used to identify molecules with similar shape and electrostatics to the initial hit molecule AF1-03. Firstly, the eMolecules library was downloaded and converted to 38,771,944 3D conformers with a maximum of two conformers per compound by Dr Katie Simmons. This was split into 4 equally sized subsets for ease of screening and each subset was screened for structurally similar compounds to AF1-03 using ROCS<sup>213</sup>. The top 100,000 compounds from each screen were combined and screened for electrostatically similar compounds using EON<sup>214</sup>. The top scoring compounds were manually inspected using VIDA and compounds were purchased from eMolecules.

The EnamineHTS library of 1,059,517 unique molecules was screened to find mimics of AF1-03 using ROCS<sup>213</sup> and the 100,000 top scoring compounds were screened using EON<sup>214</sup>. The top scorers were manually inspected using VIDA (OpenEye) and compounds purchased from ENAMINE.

#### **4.2.16 Figure making**

Charts were produced using Prism 9 (GraphPad). Figures of compounds and protein structures were produced using ChimeraX 1.5<sup>149</sup>.

### **4.3 Results**

The aim of this chapter was to experimentally screen the small molecules that were developed using the VS workflows outlined in chapter three for binding to their viral targets and for inhibition of viral processes. This includes the small molecule mimics of NP-Affimer, which inhibits CCHFV NP, termed the AF1 series and the AF2 series of

small molecules which were generated by a traditional SBVS workflow at the RNA-binding site. The same LBVS was used to generate small molecule mimics of A5 which inhibits the HA spike protein from IAV.

### **4.3.1 NP-Affimer mimics AF1-01 and AF1-03 bind to and inhibit CCHFV NP**

#### **4.3.1.1 AF1-01, AF1-03 and AF2-05 inhibit CCHFV replication in a mini-genome system assay**

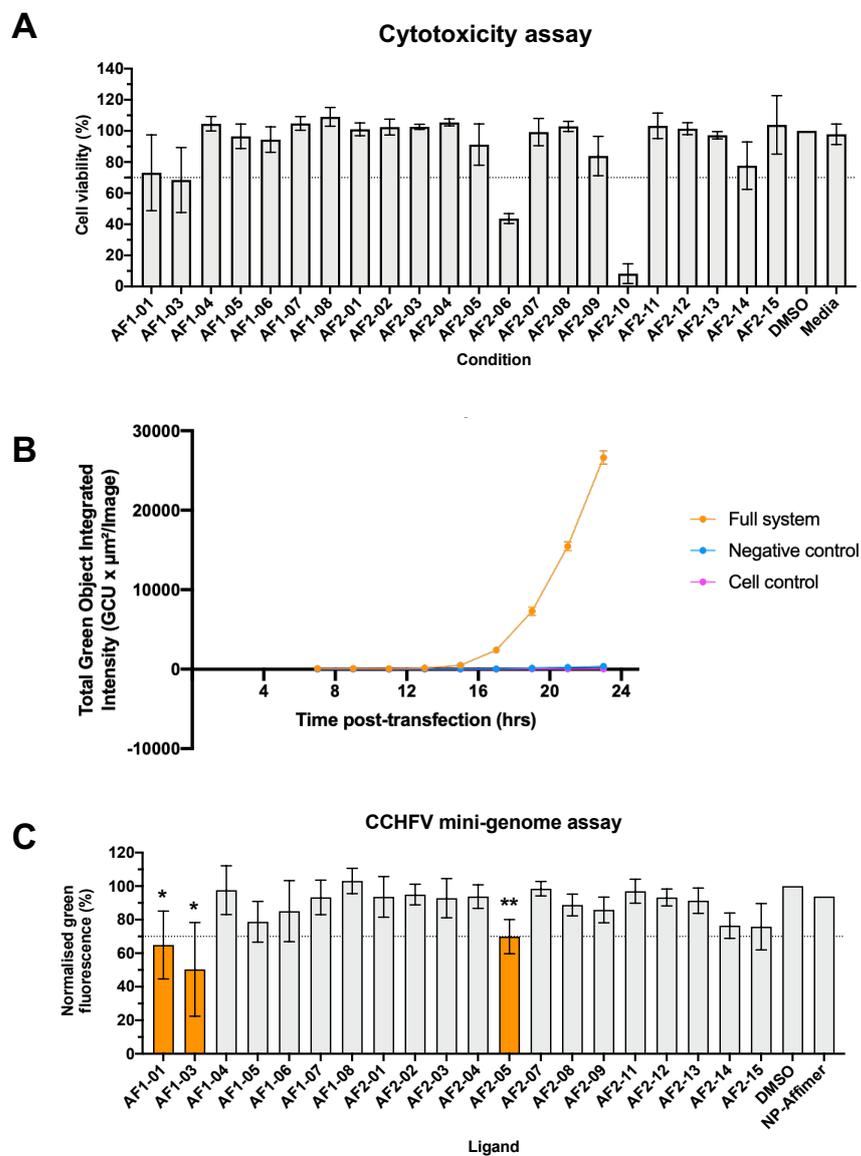
The AF1 and AF2 compounds were first screened for cytotoxicity in the BSR-T7 cell line used for the MGS, prior to screening them for inhibition of the CCHFV MGS, since cytotoxic compounds would present as false hits in the MGS assay by indirectly reducing GFP expression. Cell viability was assayed by measuring intracellular ATP concentrations. Compounds were initially screened at 100  $\mu$ M and 10  $\mu$ M, but the high cytotoxicity measured at 100  $\mu$ M meant this concentration was not progressed further. AF2-06 and AF2-10 at 10  $\mu$ M led to <70% cell viability, so they were not taken forward for screening against the CCHFV MGS (Figure 4.7A). AF1-03 led to a cell viability of 68.5%, which was above the threshold for cytotoxicity but deemed sufficiently close to the threshold to take AF1-03 forward for further screening.

Before compound screening against the MGS, the method for transfecting the CCHFV MGS was changed from that used to characterise NP-Affimer previously<sup>86</sup>. Previously, plasmids for the mini-genome system were co-transfected with a plasmid for NP-Affimer expression in a 12-well plate, which led to a reduction in GFP expression and therefore CCHFV replication relative to control wells. This method was not appropriate for screening the AF1 and AF2 compounds because the throughput in a 12-well plate was not high enough to screen all 19 non-cytotoxic compounds in triplicate. Furthermore, addition of the compound during transfection could negatively affect the DNA transfection and produce a false positive result.

Transfecting cells directly in a 96-well plate was attempted initially but this led to high variation in the transfection efficiencies between replicate wells, likely due to poor swirling and mixing of transfection complexes within the small wells. Instead, cells were first transfected in a 6-well plate, which allowed for better mixing of the transfection complexes (Figure 4.7B). After 24 hours, the transfected cells were trypsinised and re-seeded with 10  $\mu$ M compound in a 96-well plate.

When the compounds were tested at 10  $\mu$ M against the MGS using this new method, three compounds reduced the GFP fluorescence to below the 70% threshold (Figure 4.7C). These were Affimer-mimics AF1-01 and AF1-03, the latter of which showed the greatest effect with a normalised green fluorescence of 50.3%, and vHTS-based compound AF2-05, which was just below the threshold at 69.9%.

Although this re-seeding method reduced variation in fluorescence across the plate and removed any compound effects on the MGS DNA transfection, it meant that NP-Affimer no longer functioned as a positive control, as it is not cell permeable unlike the compounds. When NP-Affimer DNA was transfected when the compounds were added, there was no effect on the green fluorescence. This was attributed to three potential reasons: low Affimer transfection efficiency caused by poor mixing in a 96-well plate, NP levels being too high at 24 hours for inhibition by the Affimer and/or the transfection efficiency being low in cells following a second transfection within 24 hours. This lack of positive control in the MGS assay was unfortunate but deemed an acceptable trade off with the higher throughput that came from performing the assay in a 96-well plate.



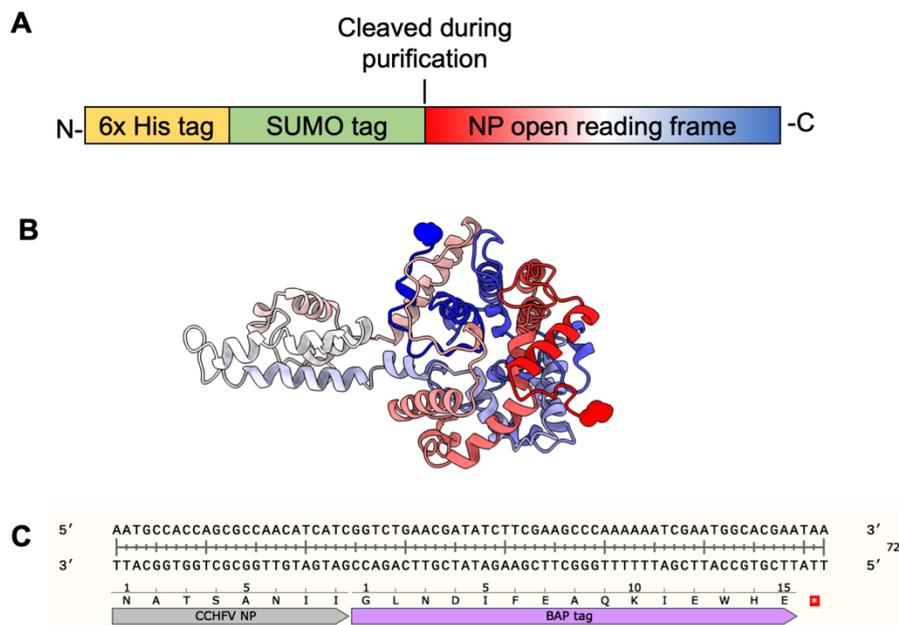
**Figure 4.7 Screening the AF1 and AF2 sets of small molecules for inhibition of the CCHFV mini-genome system (MGS) identified three active compounds.** A) Cytotoxicity assay in BSR-T7 cells with 10  $\mu$ M compound (N=3). Results normalised to the DMSO control. B) Green fluorescence following transfection of the CCHFV MGS in a 6-well plate. Negative control is cells transfected with the mini-genome system minus the L support plasmid. Cell control is non-transfected cells. C) AF1 and AF2 compound screen against the CCHFV MGS at 10  $\mu$ M (N=3). Results normalised to the DMSO control. Compounds that led to fluorescence below the 70% threshold are highlighted in orange.

By inhibiting the CCHFV MGS assay, AF1-01 and AF1-03 were the first examples of biologically active compounds derived via mimicking an Affimer structure. However, it is important to note that both compounds displayed cytotoxicity at 10  $\mu$ M. Therefore, if either of them were to be taken for hit-to-lead optimisation, removing this cytotoxicity would be key, perhaps by reducing the required dose via an increase in potency, or by making informed structural changes to reduce cytotoxicity.

#### **4.3.1.2 A new BAP-tagged NP construct was produced to facilitate immobilisation in SPR experiments**

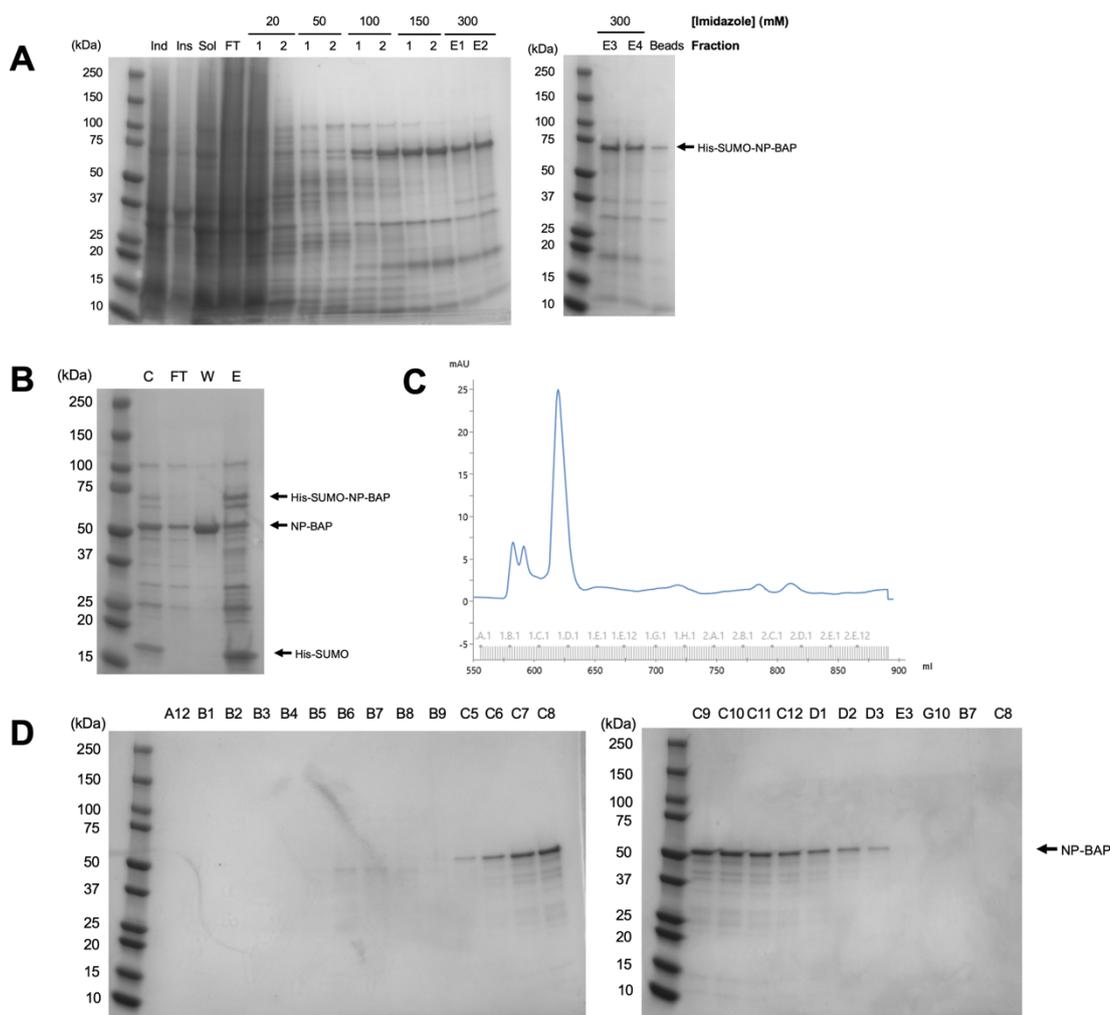
Once active AF1 and AF2 compounds were identified using the CCHFV MGS assay, it was important to verify that this activity was due to their binding to CCHFV NP. Surface plasmon response spectroscopy (SPR) was selected as the first method for this due to its high throughput, high sensitivity and its ability to determine the binding affinity ( $K_D$ ) as well as binding rates ( $k_{on}$  and  $k_{off}$ ) for compounds. During SPR, binding is detected by flowing compounds over an immobilised target and measuring changes in the optical properties of the surface. Target immobilisation can be a limitation of SPR, but streptavidin (SA) conjugated chips simplify the process, provided the target is biotinylated either chemically or during protein expression via the engineering of a biotin-acceptor peptide (BAP) into the target. Biotinylation via a BAP tag was selected for CCHFV NP since chemical biotinylation can impact target-ligand binding and it is difficult to regulate the location and intensity of the biotinylation.

During CCHFV NP purification, the N-terminal His and SUMO tags are cleaved off (Figure 4.8A, yellow and green regions). Since the BAP tag must remain on NP for SPR, the flexible C-terminus was chosen as the site for BAP tag introduction (Figure 4.8B, blue spheres). Introduction of the tag was performed by Dr Amy Turner. Following site directed mutagenesis, sequencing showed the BAP tag had been correctly inserted between the NP open reading frame and stop codon (Figure 4.8C).



**Figure 4.8 Introduction of a BAP-tag into CCHFV NP construct to generated biotinylated NP.** A) Schematic of the His-SUMO-CCHFV NP expressed using a pET28a expression construct. Cleavage performed by Ulp1 SUMO protease. B) Structure of CCHFV NP coloured by residue number. From N to C terminus, the colour gradient is red-white-blue. N and C terminal residues are highlighted as spheres. C) DNA and protein sequences following the cloning of a BAP tag into the C-terminus of CCHFV NP.

His-SUMO-NP-BAP was overexpressed in CVB101 cells and purified. Following the first purification by nickel affinity chromatography (NAC), a band corresponding to His-SUMO-NP-BAP was visible in the SDS-PAGE analysis of the 100 mM and 150 mM imidazole wash fractions and the elution fractions (Figure 4.9A), so these were pooled and incubated with Ulp1 SUMO protease to cleave off the N-terminal His and SUMO tags. After a second NAC purification to remove the cleaved tags and His-tagged protease, there was a shifted band corresponding to a cleaved NP in the flow-through, indicating the two tags were successfully removed (Figure 4.9B). Finally, the protein was cleaned up using size-exclusion chromatography (SEC) (Figure 4.9C), with NP found in the tallest SEC peak (Figure 4.9D). Fractions containing NP-BAP were concentrated to 1.63 mg/mL and snap frozen for SPR screening.

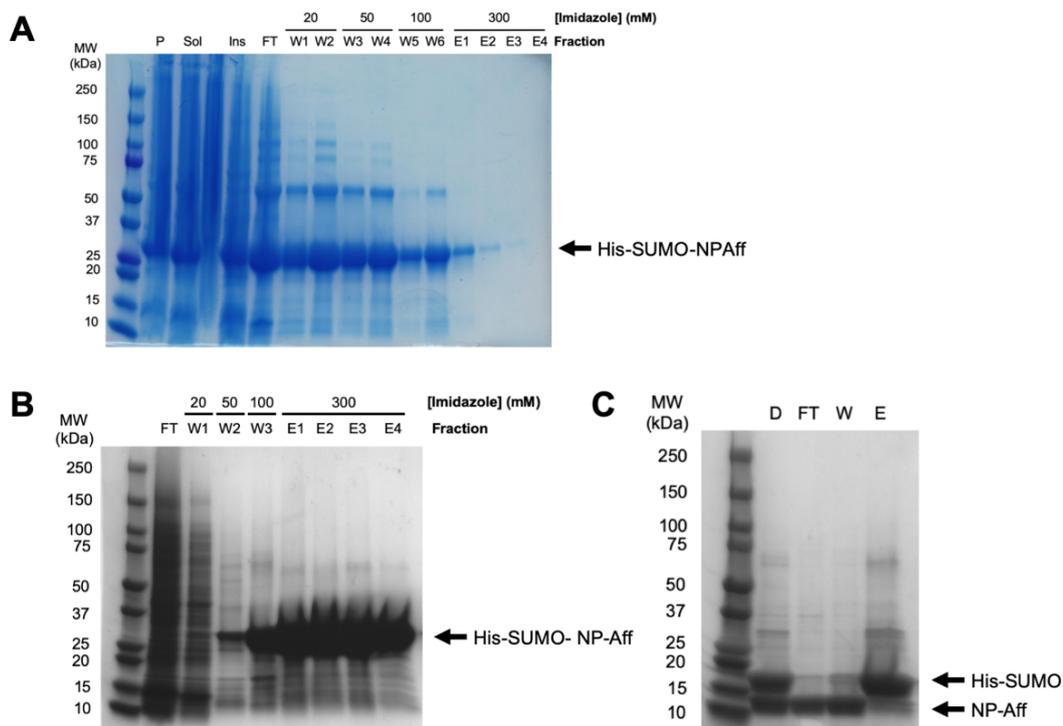


**Figure 4.9 Purification of a BAP-tagged CCHFV NP from *E. coli* cells.** A) SDS-PAGE analysis of fractions from the first nickel affinity chromatography run. Ind = induced *E. coli*, ins = insoluble fraction, sol = soluble fraction, FT = flow-through. B) SDS-PAGE analysis of fractions from the second nickel affinity chromatography run following cleavage by Ulp1 SUMO protease. C = Cleaved, FT = flow-through, W = wash, E = elution. C) Chromatogram of the size exclusion chromatography run to clean up the sample. D) SDS-PAGE analysis of peak fractions from the size exclusion chromatography run.

#### 4.3.1.3 NP-Affimer was produced for use as a positive control in binding assays

NP-Affimer represented a useful positive control for binding assays, so it was produced following the workflow explained in Álvarez-Rodríguez *et al.*<sup>86</sup>. Following a first nickel affinity chromatography step, His-SUMO-NP-Affimer was visible in the flow-through and all wash steps rather than the elutions, suggesting there was an issue with the purification (Figure 4.10A). Therefore, the flow-through, 20 mM and 50 mM imidazole wash fractions were pooled, the imidazole was dialysed out and the nickel affinity chromatography again with fresh buffers. This time, His-SUMO-NP-Affimer was in the 100 mM wash and elution fractions (Figure 4.10B), so these were pooled and the His-SUMO tags were cleaved off using Ulp1 SUMO protease. A third nickel affinity chromatography was performed to separate the His-SUMO tags and His-tagged Ulp1 protease from untagged NP-Affimer. NP-Affimer was identified in the flow-through and

wash fractions, so these were pooled and concentrated to 13.45 mg/mL and snap frozen (Figure 4.10C).

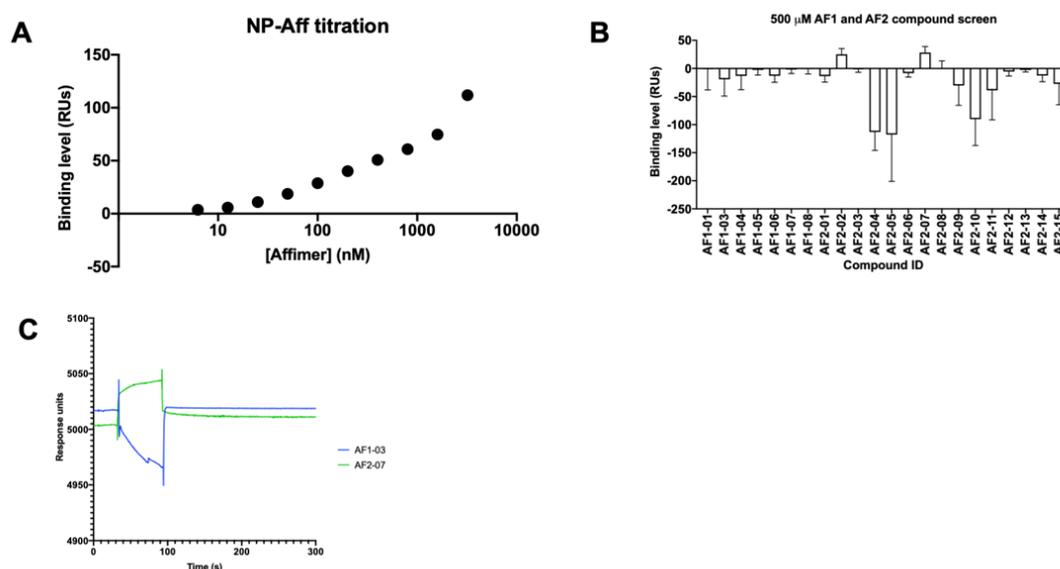


**Figure 4.10 Purification of the NP-Affimer which inhibits CCHFV NP for use as a control ligand in screening assays.** A) SDS-PAGE analysis of fractions from the first nickel affinity chromatography run. P = Cell pellet, sol = soluble fraction, ins = insoluble fraction and FT = flow-through. B) SDS-PAGE analysis of fractions from the second nickel affinity chromatography run. D) SDS-PAGE analysis of fraction from the third nickel affinity chromatography run following cleavage by Ulp1 protease. D = Dialysed mixture, FT = flow-through, W = wash and E = elution.

#### 4.3.1.4 Surface Plasmon Resonance failed to detect active compounds binding to CCHFV NP

Once biotinylated NP was produced, it was used in downstream SPR experiments to screen the AF1 and AF2 compounds. Initially, screening was performed with compounds at 100  $\mu$ M and a SA chip derivatised with 4000 response units (RUs) of NP-BAP. However, this setup generated low positive binding levels for the Affimer, which indicated weak binding to the NP-derivatised chip, and mostly negative binding levels for the compounds, which indicated they interacted non-specifically with the non-derivatised reference flow cell more than the NP flow cell. Therefore, the chip was saturated by further derivatising the chip with NP-BAP and increasing the compound screening concentration to 500  $\mu$ M. It was verified that the saturated NP flow cell could still bind NP-Affimer by titrating the Affimer over the chip, which generated positive binding signals that increased in a concentration-dependent manner (Figure 4.11A).

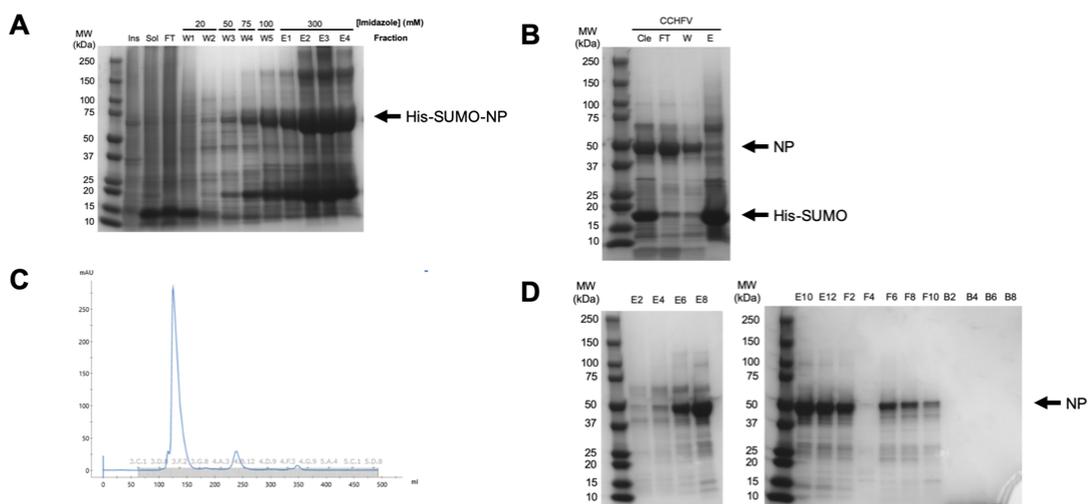
Following validation of the SPR setup, the AF1 and AF2 compounds were screened at 500  $\mu$ M (Figure 4.11B). Only two compounds identified using protein-based vHTS; AF2-02 and AF2-07, led to positive binding signals, with neither compound showing biological activity against the MGS. The majority of compounds, such as AF1-03, caused negative binding levels, which were just the inverse of the positive binding signals seen for compounds like AF2-07 (Figure 4.11C).



**Figure 4.11 Surface plasmon resonance spectroscopy identified some AF2 compounds, which were discovered through traditional structure-based virtual screening, binding to CCHFV NP.** Ligands were flowed over a flow cell derivatised with BAP-tagged NP. All binding levels were normalised to a reference flow cell. A) Titration of NP-Affimer as a positive control. B) Screen of the AF1 and AF2 compounds at 500  $\mu$ M (N=5). C) Sensograms for AF2-07 (green) and AF1-03 (blue) showing positive and negative binding signals.

#### 4.3.1.5 Wild-type CCHFV NP was produced for screening assays

Since SPR was unable to identify compound binding to NP, DSF-based thermal shift assays were used instead, since DSF does not require target immobilisation. Wild-type NP without the BAP tag was produced for the thermal shift assays. Following the first nickel affinity chromatography run, a band corresponding to His-SUMO-NP was visible in the SDS-PAGE analysis of the 75 mM and 100 mM imidazole wash fractions and the elution fractions (Figure 4.12B). These fractions were pooled and the His-SUMO tags were cleaved off. The subsequent nickel affinity chromatography run revealed shifted bands in the flow-through and wash, corresponding to cleaved NP (Figure 4.12C). These two fractions were pooled, concentrated, then purified by size exclusion chromatography (Figure 4.12D). Fractions containing NP were pooled and snap frozen for downstream screening assays.



**Figure 4.12 Purification of CCHFV NP for biophysical assays and structural studies.** A) SDS-PAGE analysis of fractions from the first nickel affinity chromatography run. Ins = insoluble fraction, sol = soluble fraction, FT = flow-through, W = wash, E = elution. B) SDS-PAGE analysis of fractions from the second nickel affinity chromatography run following cleavage by Ulp1 SUMO protease. Cle = Cleaved, FT = flow-through, W = wash, E = elution. C) Chromatogram of the size exclusion chromatography (SEC) used to clean up the sample. Absorbance at 280 nm shown as blue line. D) SDS-PAGE analysis of peak SEC fractions .

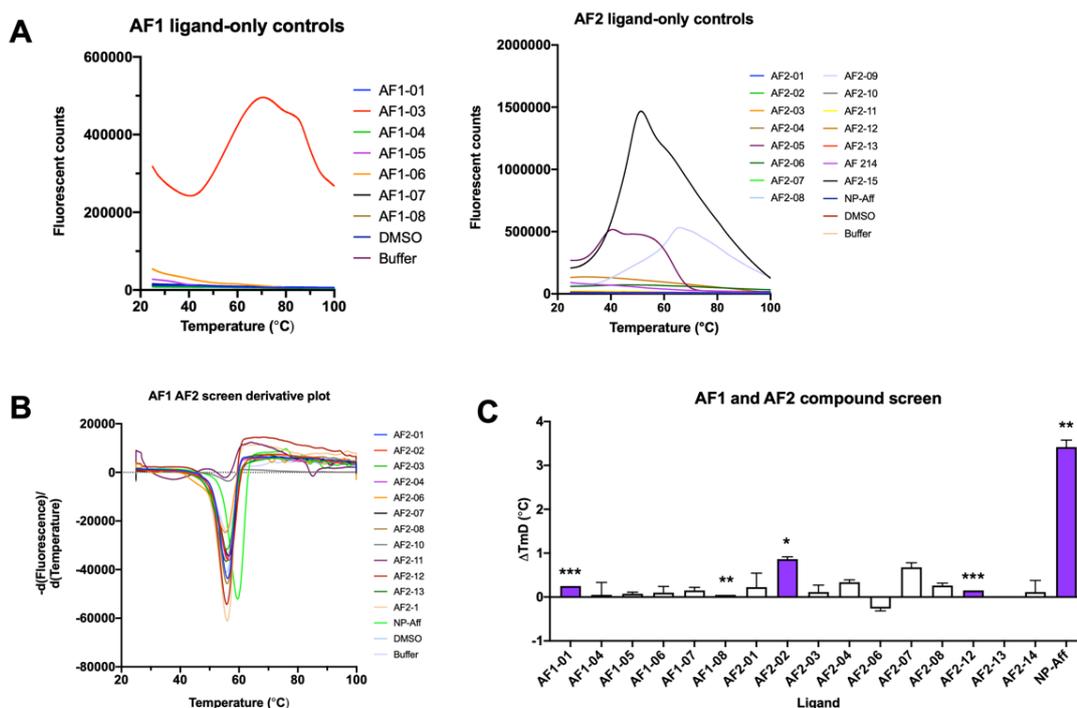
#### 4.3.1.6 DSF-based thermal shift assays identified several AF2 compounds as binders to CCHFV NP

DSF-based thermal shift assays were used to screen the AF1 and AF2 compounds for binding to NP. Ligand-only controls identified that many compounds caused high fluorescence even in the absence of NP (Figure 4.13A). These included AF1-03 and AF2-05, two active compounds in the MGS assay, as well as non-active compounds like AF2-09 and AF2-15. The fluorescence was likely caused by the compounds interacting with the SYPRO orange dye, since the dye binds to hydrophobic portions on proteins and small molecules tend to be hydrophobic. Heating up NP-Affimer alone generated a flat fluorescence curve, indicating it did not unfold even at 100°C and could be a useful positive control for compound screening.

The non-interfering compounds were heated up with NP and the derivative of the resulting fluorescence was plotted (Figure 4.13B). In both repeats, AF2-10 and AF2-11 showed little change in the derivative plot, indicating they potentially quenched the fluorescence emitted by the dye. This quenching could not have been identified in the ligand-only controls because there was no fluorescence generated by NP to be quenched.

From the derivative plots, the shift in the melting temperature ( $T_m$ ) of NP was calculated in the presence of each ligand compared to the control well (Figure 4.13C). Several compounds led to a positive NP  $T_m$  shift, including NP-Affimer mimic AF1-01 (0.3°C) and docking-based compounds AF2-02 (0.9°C) and AF2-07 (0.7°C), indicating they bind to NP. NP showed a relatively large  $T_m$  shift of 3.4°C in the presence of NP-

Affimer, which given its size and rounds of optimisation, likely produced the maximum shift that mimics could be achieved. Statistical significance was calculated but the overall significance was low because only two biological repeats were carried out due to the high level of compound interference, particularly the inhibitory compound AF1-03.



**Figure 4.13 DSF-based thermal shift assays identified binders of CCHFV NP from the AF1 and AF2 sets of small molecules.** A) Raw fluorescence for ligand-only controls to identify interfering compounds. B) Derivative plots for CCHFV NP with the AF1 and AF2 compounds. C) Thermal shifts for CCHFV NP in the presence of ligands (N=2). Compound wells normalised to a DMSO control and NP-Affimer wells normalised to a buffer control.  $T_m$  shifts marked with their statistical significance according to a paired t-test. Significance levels for  $T_m$  shift: ns = no significance, \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ , \*\*\*\* =  $p < 0.0001$ . Significant  $T_m$  shifts highlighted in purple. The overall significance is low due to only two biological repeats having been performed.

#### 4.3.1.7 Dye-free nanoDSF-based thermal shift assays detected AF1-03 binding to CCHFV NP

Several compounds including AF1-03 and AF2-05 could not be screened by DSF due to their interference with the dye, so a similar but dye-free method termed nanoDSF was used instead. Rather than detecting protein unfolding by measuring an extrinsic fluorescence source, nanoDSF measures changes in intrinsic protein fluorescence, since the wavelength of fluorescence emitted by tryptophan residues shifts from 330 nm to 350 nm when a protein unfolds.

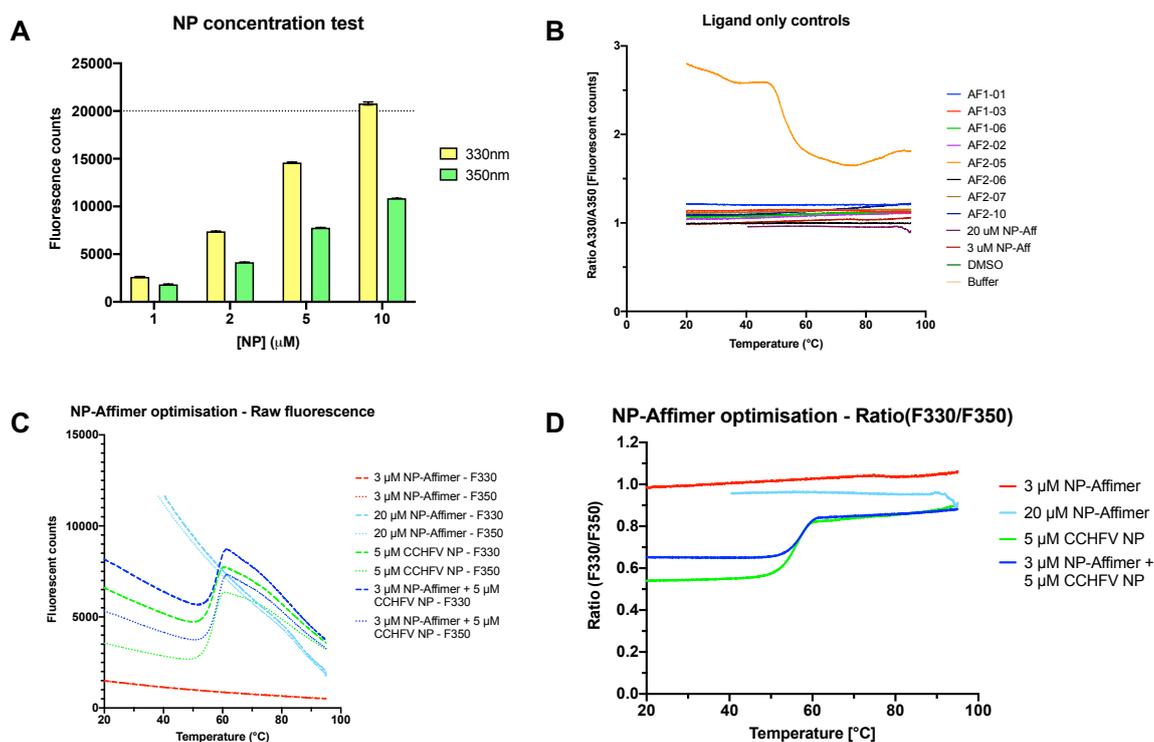
Prior to compound screening, a series of optimisation steps had to be performed. The optimum fluorescence at 330 nm for the Prometheus NanoDSF machine is  $>5,000$  units and ideally  $\sim 15,000$  units, but no higher than the 20,000 unit detector limit. The

fluorescence level is affected by changing the protein concentration and/or the excitation level. So, to identify the optimum conditions, a range of NP concentrations were scanned at 330 nm ( $F_{330}$ ) and 350 nm ( $F_{350}$ ) at various excitation levels. When 100% excitation was used, 10  $\mu$ M CCHFV NP generated a  $F_{330}$  of >20,000 and 5  $\mu$ M CCHFV NP generated a  $F_{330}$  of 14,601, which was close to the optimum level (Figure 4.14A). NP at 2  $\mu$ M led to a  $F_{330}$  of 7,387 which was high enough for screening if the target sample was limited, but this was not the case for NP, so 5  $\mu$ M was selected as the CCHFV NP screening concentration.

Even though nanoDSF circumvents one route of interference by being dye-free, interference is possible if a compound also fluoresces at 330 nm or 350 nm. Therefore, ligand-only controls were heated up to measure any fluorescence from the compounds (Figure 4.14B). The only compound that generated a high ratio( $F_{330}/F_{350}$ ) and interfered with the assay was AF2-05, so this was not taken forward for screening.

Although NP-Affimer is a useful positive control to show binding, it is a protein reagent and will also fluoresce in nanoDSF experiments thanks to its two tryptophan and 4 tyrosine residues. However, the strong affinity of NP-Affimer for CCHFV NP meant there was a potential balance to be struck where concentration of NP-Affimer was high enough to bind NP and shift its  $T_m$ , but not so high that it masks NP's fluorescence signal. To find an optimum NP-Affimer concentration, different concentrations of NP-Affimer were heated up with and without CCHFV NP.

Firstly, 20  $\mu$ M NP-Affimer caused the  $F_{330}$  and  $F_{350}$  to peak above the detection limit, meaning this concentration was too high (Figure 4.14C, green lines). However, 3  $\mu$ M NP-Affimer caused no peaking above the detection limit (Figure 4.14C, red lines). Furthermore, its ratio( $F_{330}/F_{350}$ ) was flat, indicating that NP-Affimer did not unfold even at 96°C, which was consistent with its DSF melting curve (Figure 4.14D, red line). When 3  $\mu$ M NP-Affimer was compared with CCHFV NP at 5  $\mu$ M, the concentration used for compound screening, the  $F_{330}$  and  $F_{350}$  generated by 3  $\mu$ M NP-Affimer was much lower than that generated by 5  $\mu$ M CCHFV NP (Figure 4.14C, red and dark blue lines). This meant the Affimer signal did not mask the NP signal, and 3  $\mu$ M NP-Affimer could be used as a positive control for binding despite being a protein reagent.



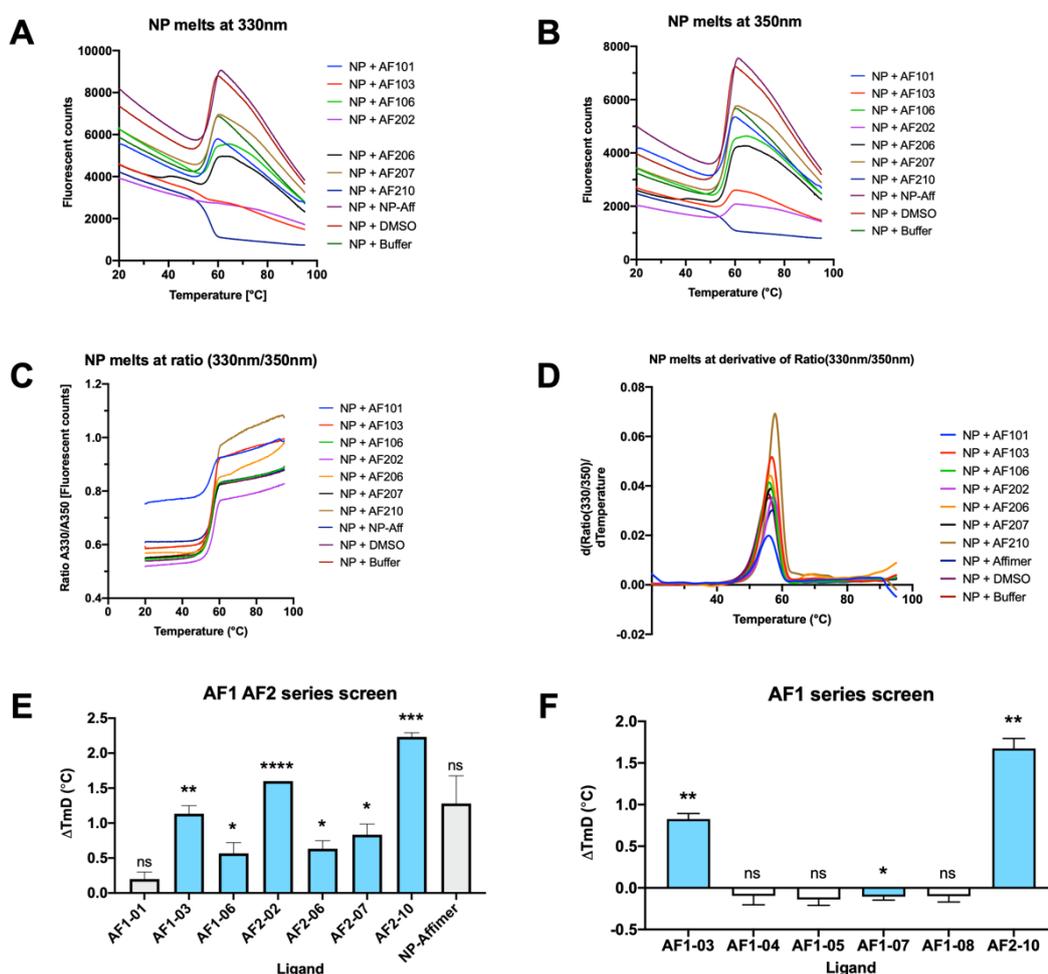
**Figure 4.14 Optimisation steps for NanoDSF thermal shift assays prior to screening the AF1 and AF2 compounds for binding to CCHFV NP.** A) NP concentration test to identify optimum [NP] for ligand screening. Yellow bars show fluorescence at 330nm (F330) and green bars show fluorescence at 350nm (F350). Y = 20,000 line represents physical limit of detection. B) Ratio(330nm/350nm) plotted for the ligand-only controls. C) Fluorescence emission at 330 nm (F<sub>330</sub>) and 350 nm (F<sub>350</sub>) for NP-Affimer and NP-Affimer + CCHFV NP. D) Ratio(F<sub>330</sub>/F<sub>350</sub>) for NP-Affimer and NP-Affimer + CCHFV NP.

When the selected compounds and NP-Affimer were screened for binding to NP by nanoDSF, the F<sub>330</sub> (Figure 4.15A) and F<sub>350</sub> (Figure 4.15B) were measured and the ratio (F<sub>330</sub>/F<sub>350</sub>) was plotted against the temperature (Figure 4.15C). The derivative of the ratio was plotted, generating melting peaks (Figure 4.15D). Finally, the inflection point of each curve was measured as the T<sub>m</sub> and the shift for each compound was calculated relative to the DMSO control or to the buffer control for NP-Affimer (Figure 4.15E).

As in the DSF thermal shift assays, NP-Affimer caused a positive shift in the T<sub>m</sub> for NP by nanoDSF, which was 1.28 °C in this case. This was not identical to the 3.4 °C shift seen by DSF, likely because DSF used different concentrations of CCHFV NP and NP-Affimer. Compounds that showed binding by SPR and DSF, such as AF2-02 and AF2-07, also led to positive binding shifts in nanoDSF, validating this experimental setup. The positive shift of 1.13 °C for AF1-03, which was the strongest inhibitor of CCHFV replication in the MGS assay, indicated that it bound to NP. This confirms that small molecules developed to mimic an Affimer can bind to the same target. AF1-01, which showed low but significant inhibition of CCHFV replication, led to a small binding shift with no significance, indicating AF1-01 is not a good example of an Affimer-based compound. AF2-10, which was too cytotoxic to screen against the CCHFV mini-genome system and interfered with SPR and DSF, generated a T<sub>m</sub> shift of 2.23°C by

nanoDSF, indicating it was potentially one of the tightest small molecule binders of the AF1 and AF2 series.

AF1-06 and AF2-06 generated positive  $T_m$  shifts of  $0.57^\circ\text{C}$  and  $0.63^\circ\text{C}$  by nanoDSF, whereas they induced a  $T_m$  shift of  $\sim 0^\circ\text{C}$  by DSF. This suggested that other compounds might show binding that was not identified in other assays. Therefore, the remaining AF1 compounds were screened, using AF1-03 and AF2-10 as positive controls (Figure 4.15F). This identified that AF1-04, AF1-05, AF1-07 and AF1-08 did not bind CCHFV NP, with the positive shifts detected again for AF1-03 and AF2-10, further increasing the validity of these results.



**Figure 4.15 Screening the AF1 and AF2 compound series at 200  $\mu\text{M}$  using thermal shift assays by nanoDSF.** A) Fluorescence emission at 330 nm ( $F_{330}$ ) for NP in the presence of ligands. B) Fluorescence emission at 350 nm ( $F_{350}$ ) for NP in the presence of ligands. C) Ratio( $F_{330}/F_{350}$ ) for NP in the presence of ligands. D) Derivative of the ratio( $F_{330}/F_{350}$ ) for NP in the presence of ligands. E) NP thermal shifts in the presence of ligands. NP-Affimer screened at 3  $\mu\text{M}$  ( $N=3$  for all small molecules,  $N=3$  for NP-Affimer). NP + compound normalised to NP + DMSO control and NP + NP-Affimer normalised to NP + buffer control. Error bars indicate mean  $\pm$  standard deviation.  $T_m$  shifts marked with their statistical significance according to a paired t-test. Significance levels for  $T_m$  shift: ns = no significance, \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ , \*\*\*\* =  $p < 0.0001$ . Significant  $T_m$  shifts highlighted in blue. This scheme is used throughout the figure. F) NP thermal shifts in the presence of ligands normalised to a NP + DMSO control ( $N=3$ ).

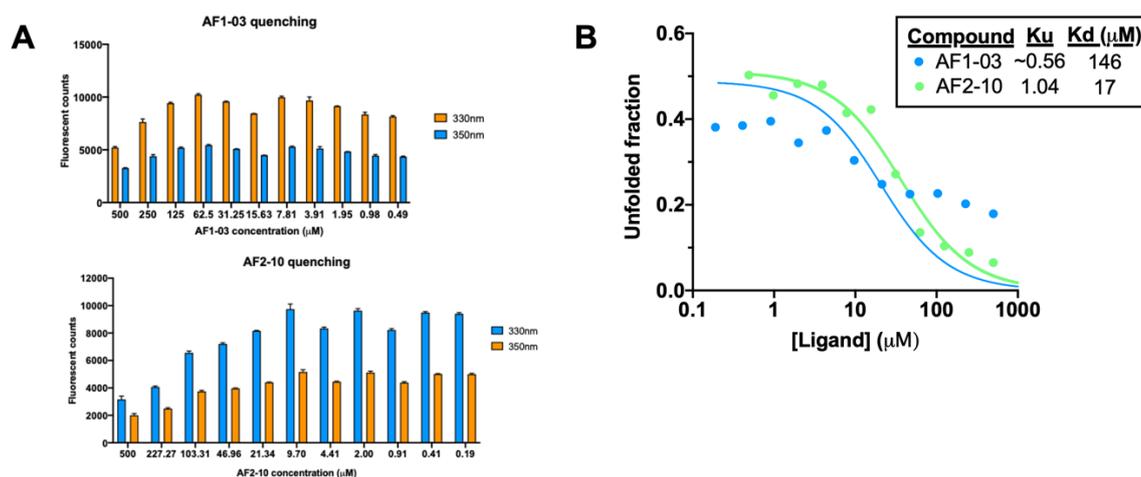
The success in nanoDSF measuring the binding of AF1-03 to NP allowed us to take it one step further and determine the  $K_D$  of this interaction. Previously, 'apparent  $K_D$ 's were calculated from nanoDSF data by plotting the  $\Delta T_m$  as a function of [Ligand].

However, recent work has shown a true  $K_D$  can be calculated through an alternative method through the FoldAffinity server<sup>245</sup>. First, the melting curves are fitted with a two-state folding model to calculate the amount of folded and unfolded protein at each ligand concentration. From these folding plots, an observed unfolding constant termed  $K_u$  is calculated and used in isothermal binding analysis to calculate the  $K_D$  binding affinity for the ligand at temperatures close to  $T_m$ . This method was tested on both AF1-03 and AF2-10, since the latter gave the highest  $T_m$  shift for all the compounds tested.

Firstly, in order to detect quenching at high ligand concentrations, the  $F_{330}$  and  $F_{350}$  was measured for a concentration series of AF1-03 and AF2-10 (Figure 4.16A). At high concentrations, both ligands showed some quenching, with 500  $\mu$ M AF1-03 generating 50-55% the maximum fluorescence seen at lower compound concentrations and 500  $\mu$ M AF2-10 generating 35-40% the maximum fluorescence. Advice from the FoldAffinity team was that their method has been successfully tested with compounds that showed a 40% decrease in fluorescence. So although the quenching was acceptable, the top compound concentrations were at the upper limit for testing.

NP was then melted in the presence of a concentration series of AF1-03 and AF2-10 and the data was processed using the FoldAffinity server to calculate a  $K_u$  and  $K_D$  (Figure 4.16B). In the initial AF1-03 test, the unfolded fraction did not reach zero even at the highest AF1-03 concentration, indicating that the concentration series needed to be shifted higher. However, this was not possible because 500  $\mu$ M was at the acceptable upper limit for both quenching and DMSO concentration in the sample. This meant that the estimated  $K_D$  at 54.3  $^{\circ}$ C of  $\sim$ 150  $\mu$ M was taken as a rough estimation. AF2-10 generated a sigmoidal shaped curve that got close to zero at the highest concentrations, so this was repeated three times to give an average  $K_D$  of 11.2  $\mu$ M at 54.3  $^{\circ}$ C.

In summary, AF1-03 has  $\sim$ 10-fold lower affinity for NP than AF2-10. Both of these  $K_D$ 's are several orders of magnitude lower than that for NP-Affimer which was low nM at 20  $^{\circ}$ C. It is not surprising that AF1-03 has a much lower affinity than NP-Affimer, because NP-Affimer went through several rounds of development and AF1-03 only went through one. Although AF2-10 was higher affinity compared to AF1-03, AF2-10 was too cytotoxic to test for biological activity, so it could not be confirmed whether it was also more potent. Either way, the affinities of AF1-03 and AF2-10 require improvement to nM  $K_D$  values for them to be considered lead molecules and potentially used in the clinic. This improvement is achieved through the development of more potent analogues.



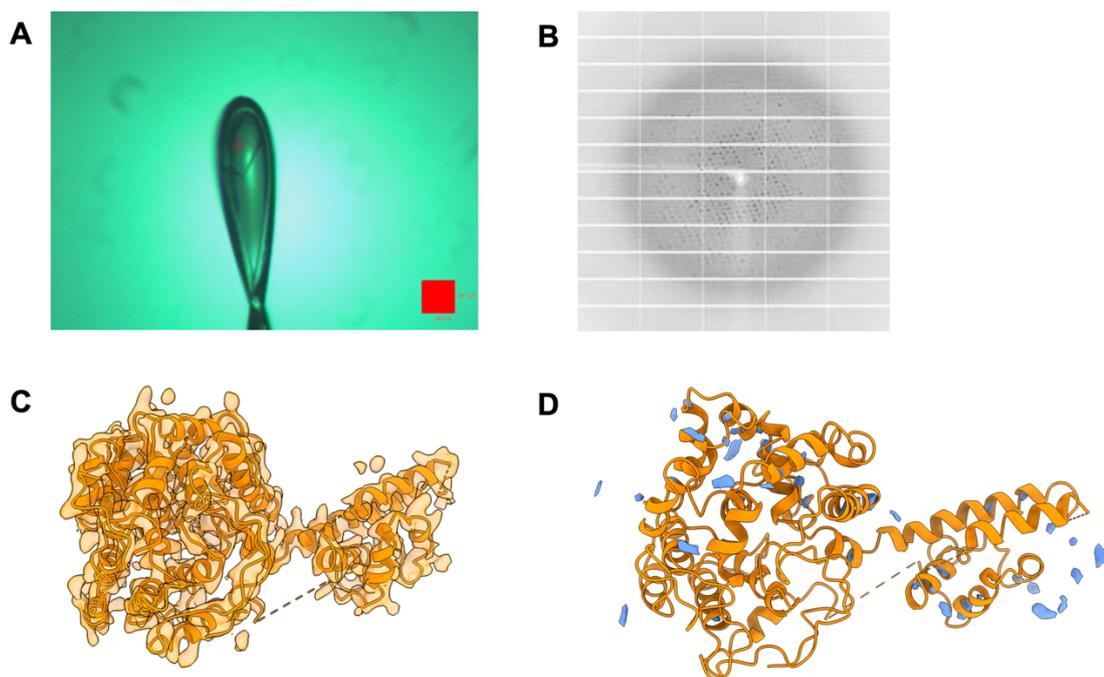
**Figure 4.16 Using nanoDSF to calculate  $K_D$ s of AF1-03 and AF2-10.** A) Quenching experiments at 330 nm and 350 nm ( $N=1$ ). B) Unfolded protein fraction plots for AF1-03 and AF2-10 at various concentrations of ligand. These were used to calculate the  $K_D$  for ligand binding.

### 4.3.1.8 X-ray crystallography did not identify the small-molecule binding sites

In order to identify the binding site of AF1-03 and whether this was the same as that for NP-Affimer, X-ray crystallography was employed to obtain a ligand-bound NP structure. Obtain a ligand-bound protein structure typically occurs through two methods, co-crystallising the small molecule with the protein or soaking the small molecule into a pre-formed protein crystal. Co-crystallisation was used because the location of the Affimer-binding site at a crystal contact point meant that if AF1-03 bound there, it was likely to disrupt the crystal during soaking.

NP and AF1-03 were mixed and trays of crystal trials were set up with six crystallisation suites. Crystals were picked and suitable crystals were collected on at Diamond Light Source. A crystal formed from the condition 0.1 M SPG buffer pH 6.0, 25% (w/v) PEG 1500 taken from the PACT suite generated the best diffraction pattern (Figure 4.17A and B). When this diffraction pattern was processed into a model using molecular replacement, there was no extra density that corresponded to a small molecule (Figure 4.17C-D, Table 4.3). This indicated that AF1-03 was not bound to the crystal and its binding site could not be elucidated.

Seeding was also tried to generate crystals for soaking. However this generated the wrong shaped crystals, such as needles and twinned crystals, where two or more separate crystal lattices grow from the same point.



**Figure 4.17 X-ray crystallography of CCHFV NP crystallised in the presence of AF1-03 generated an apo structure of NP.** Crystal grown in 0.1 M SPG buffer pH 6.0, 25% (w/v) PEG 1500 and shot at the diamond light source. B) Diffraction pattern generated by the crystal in panel A. C) Electron density (orange surface) and corresponding protein model D) Unaccounted electron density. No region was large enough to accommodate the small molecule.

**Table 4.3 Information on X-ray crystallography data collection and model refinement for CCHFV NP + AF1-03**

Data collection	
Unit cell parameters	72.8, 82.0, 311.2, 90.0, 90.0, 90.0
Space group	P 21 21 21
Low resolution limit	72.68
High resolution limit	3.50
Number unique reflections	24344
Overall, resolution estimate	3.5
Refinement	
R-work	0.314
R-free	0.411
RMS (angles)	2.69
RMS (bonds)	0.0282

Ramachandran outliers	1.07%
Ramachandran favoured	94.24%
C-beta outliers	51
Molprobiity Clashscore	113.28
Molprobiity Overall score	3.32

#### 4.3.1.9 Summary of screening assays

The aim of section 4.3.1 was to screen the AF1 and AF2 compound series for compounds which bound to and inhibited CCHFV NP. A number of assays were employed for this due to various examples of compound interference. AF1-03 was the most promising compound because it inhibited the CCHFV MGS and showed binding to NP in thermal shift assays by nanoDSF, with an estimated  $K_D$  of  $\sim 150 \mu\text{M}$  (Table 4.4). AF1-01 demonstrated weaker inhibition of the MGS and the low  $T_m$  shift it generated in nanoDSF thermal shift assays was not statistically significant, suggesting it is not a real inhibitor of CCHFV NP. AF2-05 showed activity against the CCHFV MGS but neither SPR, DSF nor nanoDSF were able to detect its binding to NP due to it interfering in the assays. Although it is likely that AF2-05 binds NP, it cannot be assumed.

Several compounds in the AF2 set showed binding but no activity. For example, AF2-02 and AF2-07 showed binding by SPR, DSF and nanoDSF but demonstrated no activity against the CCHFV MGS. AF2-06 and AF2-10 bound NP in multiple assays but were too cytotoxic to screen against the MGS. The  $K_D$  for AF2-10 of  $\sim 11 \mu\text{M}$  was even stronger than that for AF1-03. The lack of biological activity for these compounds indicated that the RNA binding site was not an effective site for blocking NP function, since even strong binding here did not lead to CCHFV inhibition. Such non-active binders were not common in the AF1 set. The only potential example was AF1-06, which demonstrated binding by thermal shift assays by nanoDSF but not those by DSF.

**Table 4.4 Summary of AF1 and AF2 compound screening assays.** Compound results are colour coded according to hit thresholds.

Ligand	Development	BSR-T7 Cell survival (%)	Mini-genome system (%)	SPR Binding level (RUs)	DSF-based thermal shifts ( $\Delta^{\circ}\text{C}$ )	NanoDSF thermal shifts ( $\Delta^{\circ}\text{C}$ )
AF1-01	Affimer-guided ligand-based virtual screening	73.08	64.88	-16.97	0.25	0.20
AF1-03		68.45	50.33	-12.10		1.13
AF1-04		104.57	97.61	-19.37	0.05	-0.10
AF1-05		96.53	78.72	1.74	0.08	-0.14
AF1-06		94.42	85.08	-9.79	0.10	0.57
AF1-07		104.79	93.28	2.20	0.15	-0.11
AF1-08		109.02	103.07	5.06	0.05	-0.10
AF2-01		Structure-based virtual screening	101.02	93.58	-7.48	0.23
AF2-02	102.46		94.97	28.92	0.86	1.60
AF2-03	102.61		92.80	0.50	0.11	
AF2-04	105.45		93.79	-129.15	0.34	
AF2-05	91.20		69.88	-110.07		
AF2-06	43.68			-6.01	-0.26	0.63
AF2-07	99.25		98.47	29.95	0.68	0.83
AF2-08	102.89		88.76	3.51	0.26	
AF2-09	83.87		85.80	-31.03		
AF2-10	8.29			-69.07	0.56	2.23
AF2-11	103.26		96.98	-45.12	-0.15	
AF2-12	101.44		93.26	-2.77	0.15	
AF2-13	97.23		91.29	-0.20	0.00	
AF2-14	77.67		76.46	-7.89	0.11	
AF2-15	103.89		75.76	-22.19		
NP-Affimer			28.75	3.42	1.28	
<b>Biological repeats</b>		3	3	5	2	3
<b>[Compound] (<math>\mu\text{M}</math>)</b>		10	10	500	200	200
<b>[NP-Affimer] (<math>\mu\text{M}</math>)</b>				0.1	20	3
<b>Hit threshold (Green)</b>		>70%	<70%	>10 RUs	>0.3 $^{\circ}\text{C}$	>0.3 $^{\circ}\text{C}$
<b>Weak hit threshold (Orange)</b>		<70%, >60%	>70%, <80%		<0.3 $^{\circ}\text{C}$ , >0.2 $^{\circ}\text{C}$	<0.3 $^{\circ}\text{C}$ , >0.2 $^{\circ}\text{C}$
<b>Non-hit threshold (Red)</b>		>70%	>80%	<10 RUs	<0.2 $^{\circ}\text{C}$	<0.2 $^{\circ}\text{C}$

### 4.3.2 Analogues of NP-Affimer mimic AF1-03 were inactive against CCHFV NP

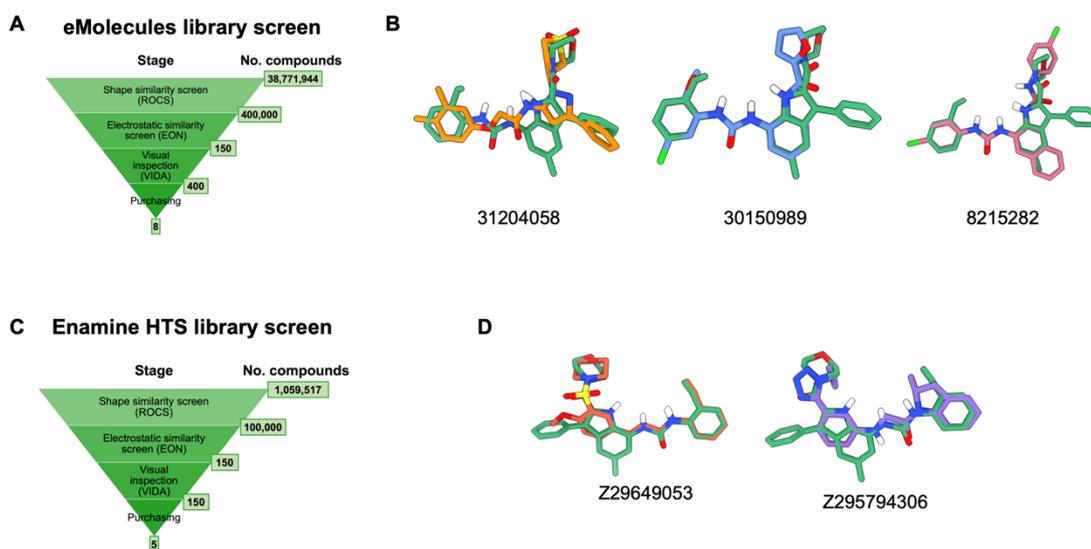
Once small molecule hits have been identified, their structure is refined to generate more active and drug-like lead molecules. Now that AF1-03 was identified as both a binder and inhibitor of CCHFV NP, its structure was optimised to try and improve its affinity and potency. This process is achieved through the development of analogues. In general, analogues are developed with the same scaffold as the original compound but minor changes in the functional groups which might alter their activity. An alternative method termed ‘scaffold hopping’ is where the functional groups are kept the same but the scaffold is changed<sup>246</sup>. Scaffold hopping can be useful to avoid intellectual property issues and to circumvent issues with off-target effects and toxicity, as was the case for AF1-03. Scaffold hopping can be performed using computational programs including the shape similarity searching program ROCS<sup>32</sup>.

#### 4.3.2.1 Structural optimisation of AF1-03

Analogue development is typically aided by a protein-ligand structure, which helps to identify where the compound could be grown into the binding site and which groups could be more complementary. However, X-ray crystallography proved unsuccessful at obtaining an NP:AF1-03 structure, so structural optimisation of AF1-03 was carried out by simply searching for small molecules with similar structures. Since AF1-03 originally came from the eMolecules library, this library was screened first to investigate whether it contained compounds with the same scaffold as AF1-03 but small variations in chemical groups. The latest version of the eMolecules library was converted to 3D conformations and screened for compounds similar in shape and electrostatics to AF1-03 using the programs ROCS and EON (Figure 4.18A).

Visual inspection of the top scoring compounds showed there were none with the same scaffold as AF1-03. This meant the traditional method of analogue development could not be used and scaffold hopping was employed instead. With that in mind, there were several compounds which had regions with similar structures as AF1-03 (Figure 4.18B). For example, compound 31204058 contained the same benzene and di-methyl benzene rings as AF1-03, but lacked its central indole ring. Half of compound 30150989 was almost identical to AF1-03 but it lacked the benzene ring in the template and contained a planar furan group rather than the saturated morpholine ring. Finally, compound 8215282 had a similar planar scaffold with an almost identical arrangement of the urea group, but it lacked AF1-03's benzene ring and contained different fluorobenzene rings.

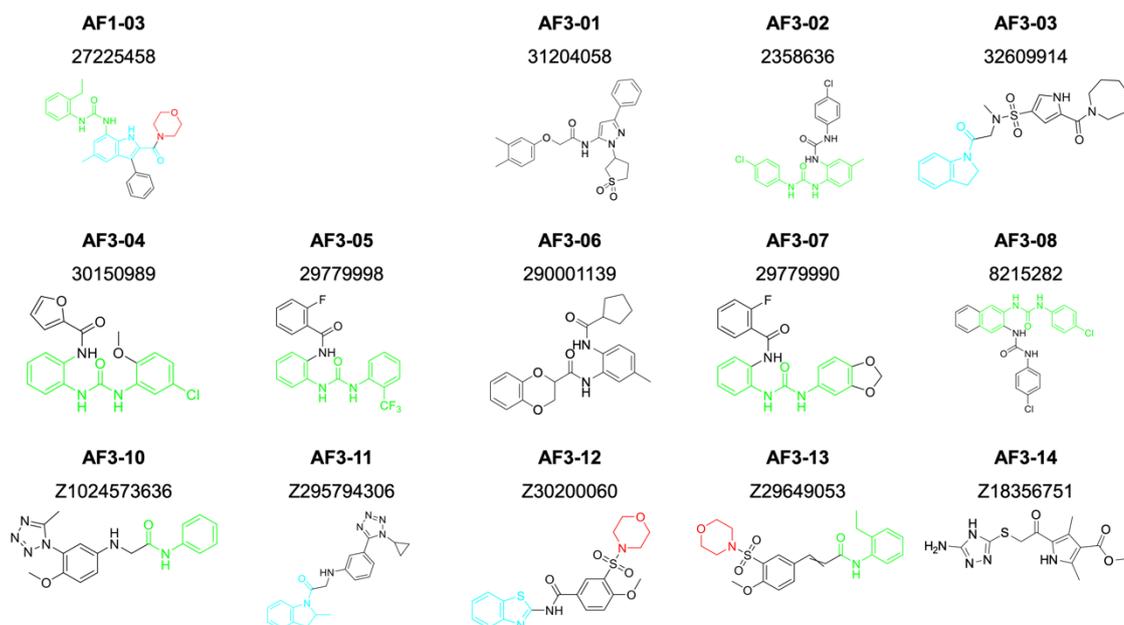
Since the top scoring compounds in the eMolecules library were not as similar to AF1-03 as was desired, the enamineHTS library<sup>247</sup> was screened using a similar workflow to potentially find better analogues (Figure 4.18C). However, the top scoring compounds also had the similar issues with some shared chemical motifs but striking differences. For example, compound Z29649053 also contained an ethylbenzene group and a similar amide backbone, but it had no benzene group or puckered ring like AF1-03. Finally, compound Z295794306 had a similar shape to AF1-93 but contained a completely new double ring.



**Figure 4.18 Analogues of Affimer-based compound AF1-03 were developed using virtual screens.** A) Workflow used to screen the eMolecules library for compounds similar to AF1-03. B) The most similar compounds from the eMolecules library overlaid with AF1-03 (green). C) Workflow used to screen the enamine HTS library for compounds similar to AF1-03. D) The most similar compounds from the enamine HTS library overlaid with AF1-03 (green).

Overall, neither the eMolecules or enamineHTS libraries contained the chemical space required to purchase true analogues of AF1-03. Even the most similar compounds had key differences that might remove their activity. However, these differences may also help highlight key interactions, so compounds were purchased from each library to tester whether they were more active.

In total, eight compounds were purchased from the eMolecules library and termed the AF3-0 compounds and five were purchased from the enamineHTS library and termed the AF3-1 compounds (Figure 4.19).



**Figure 4.19 Analogues of the active small molecule AF1-03 were purchased and termed the AF3 compounds.** Compound structures have been coloured to highlight the portions of the AF3 compounds which correspond to portions of AF1-03.

#### 4.3.2.2 AF1-03 analogues were inactive against CCHFV NP

The AF3 compounds were tested using an optimised version of the screening workflow used for the original compounds. For example, to identify binding of the AF1 and AF2 compounds to CCHFV NP, SPR, DSF then nanoDSF were employed, but nanoDSF was only method able to identify AF1-03 binding to NP due to compound interference in the prior assays. So, for the AF3 compounds, nanoDSF-based thermal shift assays were employed first, since analogues of AF1-03 were likely to interfere in the other binding assays too.

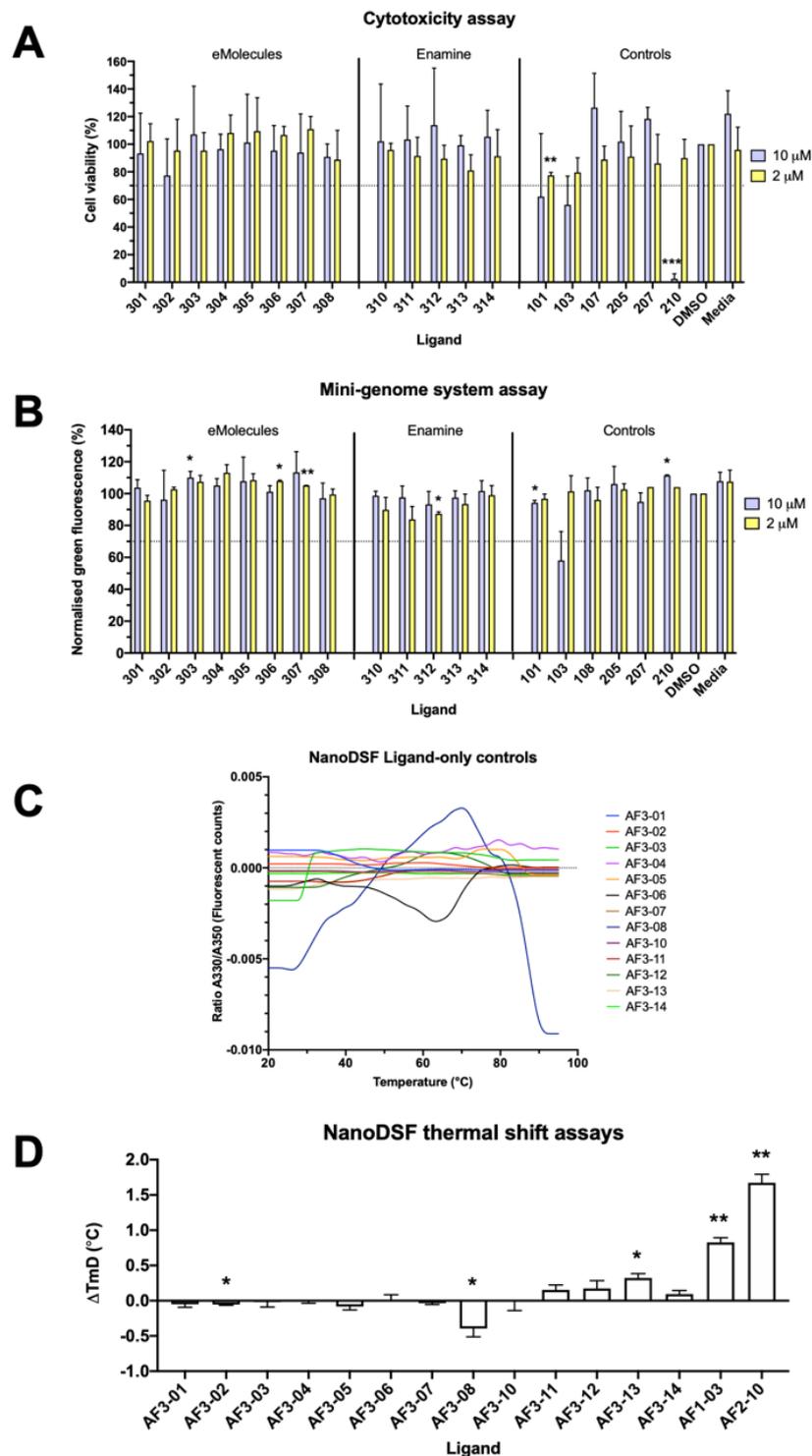
As with the AF1 and AF2 compounds, the first screening assays were cellular. In order to screen for inhibition of the CCHFV MGS in cells, compounds were first screened for cytotoxicity. Many of the original compounds were toxic at 10  $\mu\text{M}$ , so to have a better chance of finding a non-cytotoxic but active compound, analogues were screened at both 10  $\mu\text{M}$  and 2  $\mu\text{M}$ . Screening for cytotoxicity identified that no compounds led to BSR-T7 cell viability within the 70% threshold at 10  $\mu\text{M}$  or 2  $\mu\text{M}$  (Figure 4.20A).

Several original compounds were used as controls and their cytotoxicity was consistent with the original compound screens. However, cell viability was 12.4% lower in the presence of 10  $\mu\text{M}$  AF1-03 in this screen (56.1%) compared to in the initial screen (68.5%). This effect was clearly below the 70% threshold and calls in to question some of the activity of AF1-03 against the CCHFV MGS. It also highlights the irreproducibility of cell-based assays, since cells from the same cell line can behave differently if they are from different stocks, were started up at different times or are at different stages in

their life cycle. Whereas, protein-based assays show less variability by stripping away many of the cellular features.

When non-cytotoxic AF3 compounds were screened against the CCHFV MGS, there was no reduction in GFP fluorescence in the presence of any compounds (Figure 4.20B). AF1-03 was used as a control and showed a reduction in GFP fluorescence, indicative of inhibition, but it had also shown cytotoxicity far above the threshold. Interestingly, AF2-05 did not show an effect against the CCHFV MGS in this screen, but it showed inhibition of the MGS above the threshold in the initial screen. Again, the discrepancies between this MGS assay and the previous one are likely down to cell differences. Overall, the screen indicated that not one of the compounds had retained the activity of AF1-03. This was a clear sign that the compounds were not good analogues.

To verify if any of the analogues bound to NP, all of the AF3 compounds were screened for binding by nanoDSF-based thermal shift assays. Firstly, the compounds were screened without protein to check for interference (Figure 4.20C). None of the compounds showed large changes in the ratio ( $F_{330}/F_{350}$ ) as they were heated, outside of the noise of the experiment (note the low increments on the y axis). When the compounds were screened against NP, only AF3-13 generated a statistically significant positive shift in the  $T_m$  of NP, indicating AF3-13 retained some of AF1-03's activity (Figure 4.20D). AF3-08 caused a strong negative shift, indicative of it binding to CCHFV NP but destabilising it. This destabilisation might have been caused by AF3-13 binding to a less populated, partially unfolded state of CCHFV NP, which could be a potential avenue for developing CCHFV NP inhibitors<sup>248</sup>.



**Figure 4.20 Screening analogues of AF1-03 in cell-based and biophysical assays showed they were inactive.** A) Cytotoxicity assay of the AF3 compounds at 10  $\mu\text{M}$  and 2  $\mu\text{M}$  using some AF1 and AF2 compounds as controls (N=3). Results normalised to the DMSO control. Results marked with their statistical significance according to a paired t-test. Results are not significant unless otherwise marked with the following significance levels: \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ . This scheme is used throughout the figure. B) CCHFV mini-genome system assay on non-cytotoxic AF3 compounds (N=3). Results normalised to the DMSO control. C) Ligand-only controls for the nanoDSF thermal shift assay. D) NanoDSF thermal shifts of NP in the presence of AF3 compounds. Shifts are relative to the DMSO control.

Overall, the analogue screens were conclusive. None of the AF3 compounds retained AF1-03's biological activity and only AF3-13 partially retained its ability to bind NP,

likely because they were poor analogues of AF1-03. This highlights a limitation of commercially available libraries. Although they might contain the chemical space to identify initial hits, they do not necessarily contain the analogues required to generate lead molecules and understand structure-activity relationship requirements. Instead, going from hits to lead molecules is such a bespoke process that it requires medicinal chemistry input to generate customised compounds.

### **4.3.3 Affimer-derived small molecule OA3 inhibits influenza virus and binds to HA**

Eight small molecules termed the OA compounds were developed using the Affimer A5 as a template which was isolated against HA from IAV. In this section, I will describe how these compounds were tested to assess their activity against IAV using similar assays performed for CCHFV NP.

#### **4.3.3.1 OA3 inhibits live influenza virus in a TCID<sub>50</sub> assay**

Prior to screening compounds for IAV inhibition in a TCID<sub>50</sub> assay, compounds were first screened for cytotoxicity in Madin-Darby canine kidney (MDCK) cells, the same cell line used for the TCID<sub>50</sub> assay. Only eight OA compounds were purchased compared to the twenty one AF1 and AF2 compounds against CCHFV, so all the OA compounds were screened at a range of concentrations on one plate for the cell-based assays. Whereas, the AF1 and AF2 compounds were only screened at a single concentration to increase throughput. Screening for cytotoxicity showed that all compounds were toxic to some degree, with OA6 being the most toxic and OA7 and OA8 being the least toxic (Figure 4.21A). In TCID<sub>50</sub> assays, cell death is measured as a proxy for viral infection, meaning compound cytotoxicity would present a false result. The low tolerance for compound cytotoxicity meant that the highest concentration that led to >95% cell viability was selected for each compound as the maximum concentration for TCID<sub>50</sub> assay screening (Figure 4.21B).

When the OA compounds were screened for IAV inhibition, the only small molecule to show an effect was OA3, which protected cells against IAV-induced cell death at concentrations as low as 15  $\mu$ M. (Figure 4.21B and Figure 4.21C). Affimer A5 was used as a positive control, and as was shown in the initial Affimer screens, it protected cells against IAV infection at low nM concentrations. Therefore, Affimer A5 is four orders of magnitude more potent than OA3.

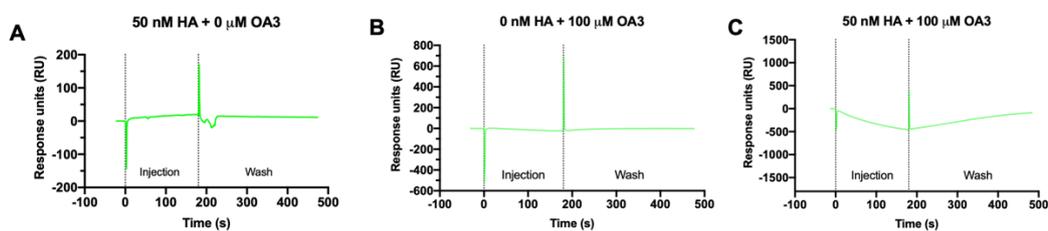
OA3's lower potency is not surprising given that OA3 is not an exact mimic of A5 and A5 is much larger, with A5 having a mass of 11,980 Da and OA3 having a mass of 486 Da. Rather than directly comparing potency via IC<sub>50</sub> measurements, it was useful to consider ligand efficiency; how effective a ligand is with its atoms. The binding efficiency index (BEI) is defined as  $\text{pIC}_{50}/(\text{Molecular mass (kDa)})^{249}$ , meaning A5 has



during the Affimer characterisation. The same setup was employed as that used by Dr Oliver Debski-Antoniak, whereby Affimers were chemically biotinylated, immobilised onto a streptavidin-conjugated gold SPR chip and then HA was flowed over the Affimer-coated flow cell. The hypothesis behind the assay was that since A5 was a template for OA3, OA3 should bind to the same location on HA as A5 and therefore HA's binding to A5 should be reduced in the presence of OA3.

This setup contrasts the traditional set up in which the target is immobilised and the ligand is flowed over the top. The SPR was performed this way because chemical biotinylation of a protein requires free lysine or cysteine residues. The lysine residues could not be biotinylated because the HA sample purchased from Sino Biologicala was supplied in a tris buffer. Tris molecules contain amine groups, meaning tris in the buffer would greatly outcompete the lysine residues for the chemical biotinylation. The cysteines could not be biotinylated either because all accessible cysteines on HA are involved in disulfide bonds. Additionally, previous attempts by Dr Debski-Antoniak to biotinylate HA were unsuccessful (Data not shown).

The arrangement was tested by flowing HA over the A5 flow-cell, which gave a positive binding signal of 19.5 response units (RUs) (Figure 4.22A). The positive signal indicated HA bound to A5 and could potentially be reduced by OA3 blocking the HA-A5 interactions. When OA3 was flowed over the A5 flow-cell as a negative control, it led to a negative binding signal (Figure 4.22B). A negative binding signal indicated that OA3 was binding to the reference flow cell more than the A5 flow cell likely due to a non-specific 'sticking' to the streptavidin or dextran layer. This is common for small molecules and was also measured for various AF1 and AF2 compounds. The negative binding signal also occurred when HA was present with OA3, meaning the effect of OA3 could not be measured by SPR (Figure 4.22C). Altering the concentrations of DMSO or Tween-20 detergent in the running buffer did not substantially improve the negative binding signal (data not shown).



**Figure 4.22 Surface plasmon resonance (SPR) spectroscopy could not identify biologically active compound OA3 binding to HA due to its non-specific sticking to the chip.** A) SPR trace of 50 nM HA injected over an A5 flow cell. B) SPR trace of 100 μM OA3 injected over an A5 flow cell. C) SPR trace of 50 nM HA and 100 μM OA3 injected over an A5 flow cell.

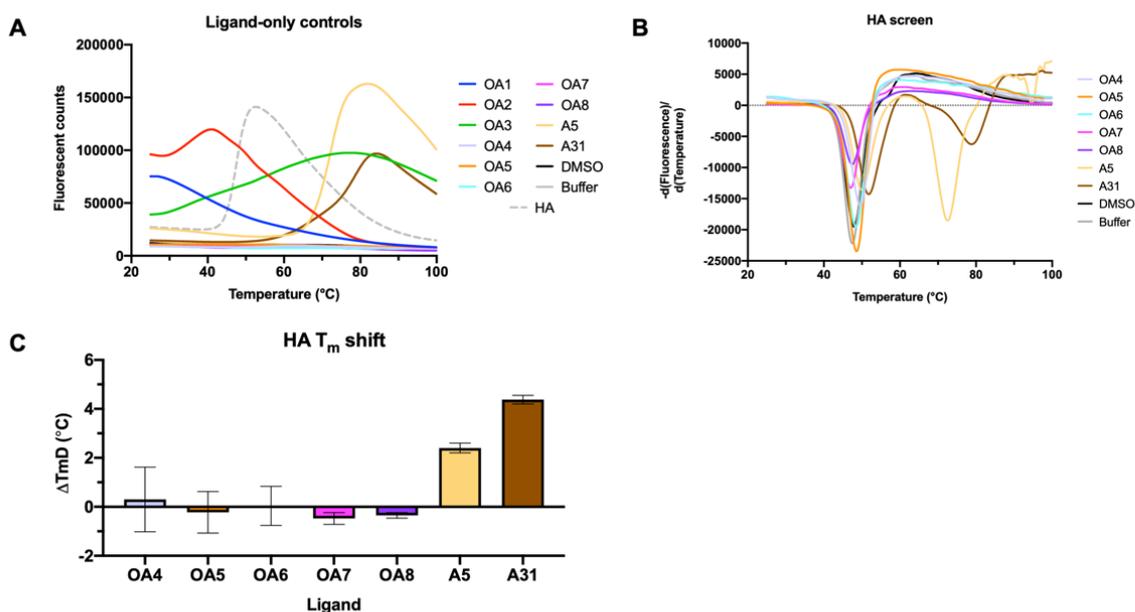
#### 4.3.3.3 DSF-based thermal shift assays did not identify compounds binding to HA

Since SPR was unsuccessful at detecting HA-OA3 binding due to OA3's non-specific binding to the SPR chip, solution-based thermal shift assays by DSF were employed instead. First, the ligands were screened without HA to check for compound interference with the dye. Most compounds generated a flat fluorescence readout, indicating no interference, but OA1, OA2 and OA3 showed a sloped fluorescence, meaning they were excluded from further analysis by DSF (Figure 4.23A). OA3's interference meant that DSF was another technique that was unable to assess its binding to HA.

In section 4.3.1.6, NP-Affimer was used as a positive control in the DSF thermal shift assays because it generated a strong CCHFV NP shift of 3.4°C and despite it being a protein reagent, its  $T_m$  was >95°C, meaning it was folded and active at the  $T_m$  for NP. The ligand-only control melts showed that the  $T_m$  for A5 and A31 were ~70°C, indicating they were less stable than NP-Affimer, but they could still be used as positive controls because they had  $T_m$ s 30 °C higher than the  $T_m$  of HA.

Monomeric HA was used to screen the compounds by DSF rather than the active trimeric form, because only the monomeric form was available in the quantities required for DSF. When the non-interfering compounds and Affimers were screened against HA, the derivative plots had clear melting curves around 48°C (Figure 4.23B). The inflection point for each melting curve in the derivative plot was measured and the  $\Delta T_m$  was calculated relative to the DMSO for compounds or the buffer control for Affimers (Figure 4.23C). From this, we saw that none of the compounds generated positive  $T_m$  shifts, but the Affimers A5 and A31, whose binding was validated during Affimer isolation, generated strong shifts of 2.4°C and 4.4°C.

The stronger binding by A31 compared to A5 is consistent with the TCID<sub>50</sub> assays performed during Affimer isolation, where A31 showed a higher potency with a lower TCID<sub>50</sub> of 0.44 nM compared to 1.96 nM for A5<sup>157</sup>. However, the  $K_D$  calculated by SPR showed that A5 had a higher affinity than A31, with a lower  $K_D$  for A5 of 2.80 nM compared to 5.94 nM.

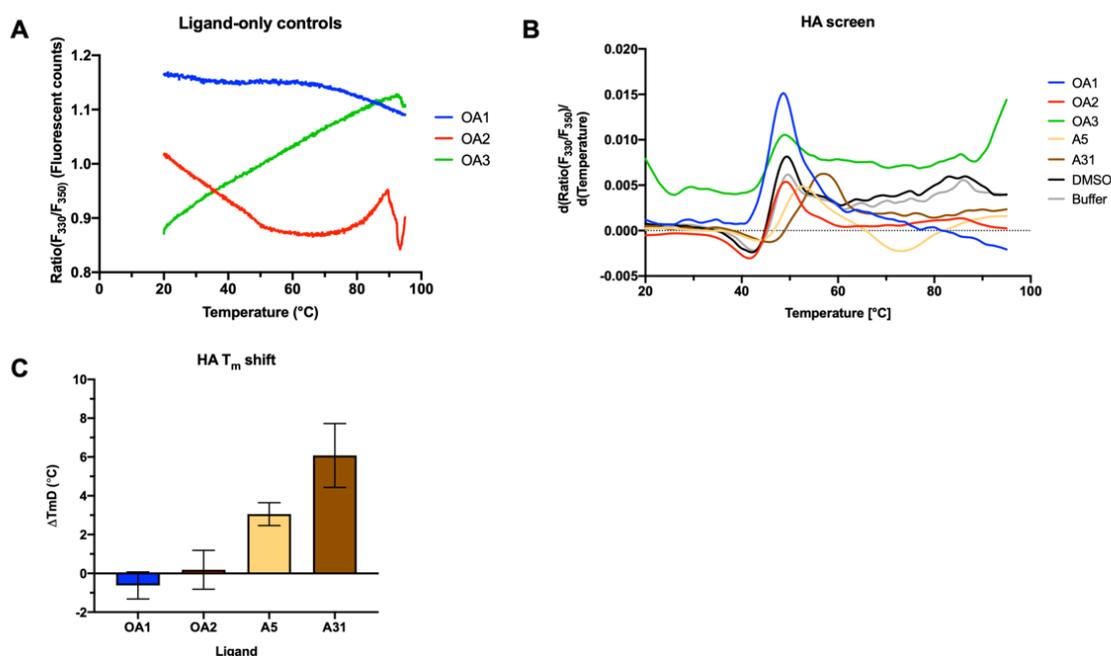


**Figure 4.23 Dye-based differential scanning fluorimetry (DSF) thermal shift assays did not identify any Affimer-based OA compounds binding to hemagglutinin (HA).** A) Ligand-only controls to identify interfering compounds. B) Screen of non-interfering compounds against HA. C) HA thermal shift in the presence of OA compounds and Affimers A5 and A31 (N=3). Compounds are relative to a DMSO control and Affimers are relative to a buffer control.

#### 4.3.3.4 NanoDSF thermal shift assays also did not identify compounds binding to HA

Because the binding of OA1, OA2 and OA3 to HA could not be measured by DSF due to compound interference, dye-free nanoDSF was used, which was previously successful at detecting the binding of AF1-03 to CCHFV NP. When the ligands were screened without HA present, all three generated gently sloped ratio( $F_{330}/F_{350}$ ) plots, indicating they all interfered with the fluorescence measurements at low levels (Figure 4.24A).

Since the interference was low, especially compared to the interference demonstrated by AF2-05 (Figure 4.14B), all three OA compounds were taken forward for screening against monomeric HA. The HA melting curves were noisier than those for CCHFV NP because the concentration of HA used in the assay was only 1  $\mu\text{M}$  compared to the 5  $\mu\text{M}$  NP used to screen compounds by DSF (Figure 4.24B). This discrepancy was due to the fact that NP was made in-house whereas HA was purchased commercially in smaller quantities. In OA3's melting curve, the melting peak was too close to the noise for accurate analysis, whereas the peaks in the other plots were distinguishable. The  $\Delta T_m$  was calculated for the ligands that could be accurately analysed (Figure 4.24C). From this we see that neither OA1 nor OA2 generated a positive shift that would indicate binding. However, A5 and A31 did generate strong positive binding shifts, again indicating that nanoDSF represents a useful tool for studying Affimer binding.

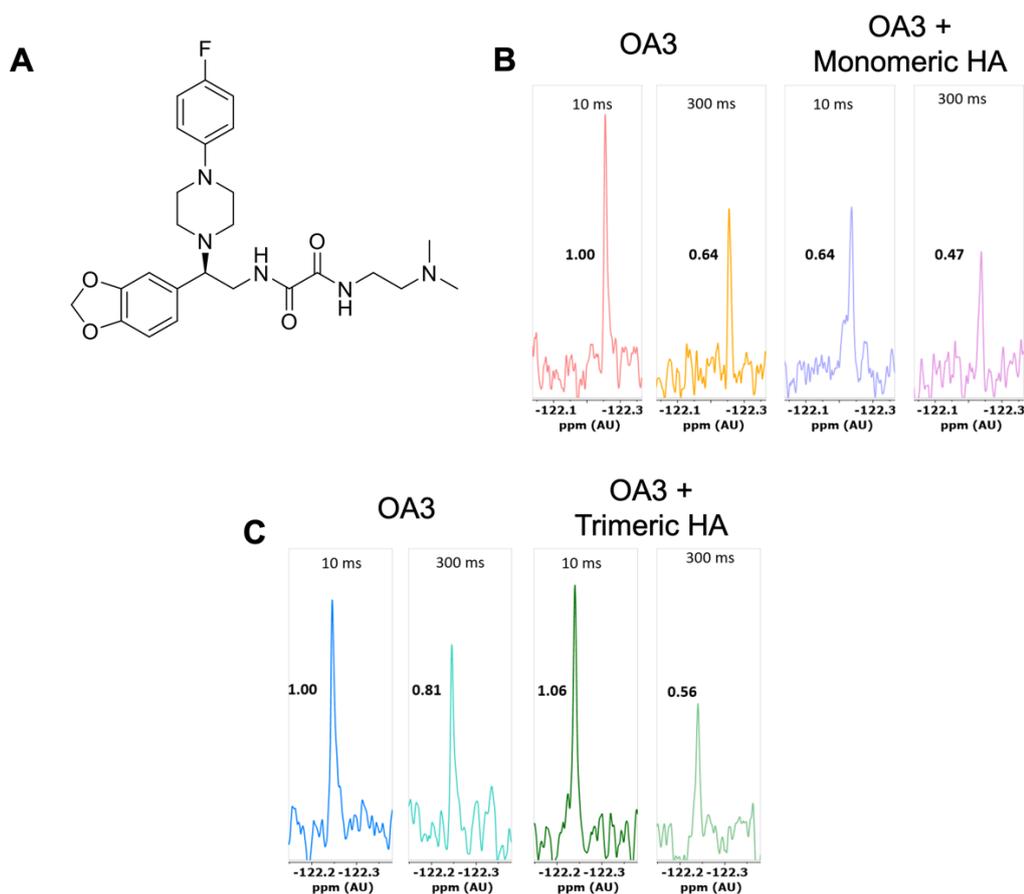


**Figure 4.24 NanoDSF thermal shift assays screened the remaining OA compounds but did not identify any binding to hemagglutinin (HA).** A) Ratio(fluorescence at 330nm/fluorescence at 350nm) for the ligand-only controls. B) Derivative of the fluorescence ratio for HA in the presence of ligands. C) Shift in HA melting temperature in the presence of various ligands. Compounds normalised to DMSO control and Affimers normalised to buffer control.

#### 4.3.3.5 $^{19}\text{F}$ Fluorine NMR experiments identified OA3 binding to HA

SPR, DSF and nanoDSF had all proved unsuitable for measuring binding of the biologically active compound OA3 to HA due to interference by various means. Ligand-based  $^{19}\text{F}$  NMR<sup>243</sup> has limited routes for interference because fluorine atoms are not found naturally in any biological material. Therefore, the peak from the fluorine in the small molecule is the only detectable fluorine peak in the spectrum. OA3 contains a single fluorine atom within its fluorobenzene ring, which should give a single peak in the  $^{19}\text{F}$  NMR trace (Figure 4.25A).

In F-NMR experiments, binding is indicated by the broadening and reduction of the  $^{19}\text{F}$  peak upon addition of the protein target, due to the changing of the chemical environment surrounding the fluorine atom. 75  $\mu\text{M}$  OA3 was screened against 0.757  $\mu\text{M}$  monomeric HA and 0.252  $\mu\text{M}$  trimeric HA (0.757 HA monomers), and for both forms of HA, there was a reduction in peak height at the 300 ms time points compared to OA3-only, indicating that OA3 had bound HA (Figure 4.25B-C). For monomeric HA, the peak of OA3 + HA at 300 ms was 47% of the height of OA3 at 10 ms, and for trimeric HA, it was 56% of the height of the OA3-only peak. Unfortunately, due to sample limitations, this experiment could only be completed once, so statistical significance could not be calculated via a paired t-test.



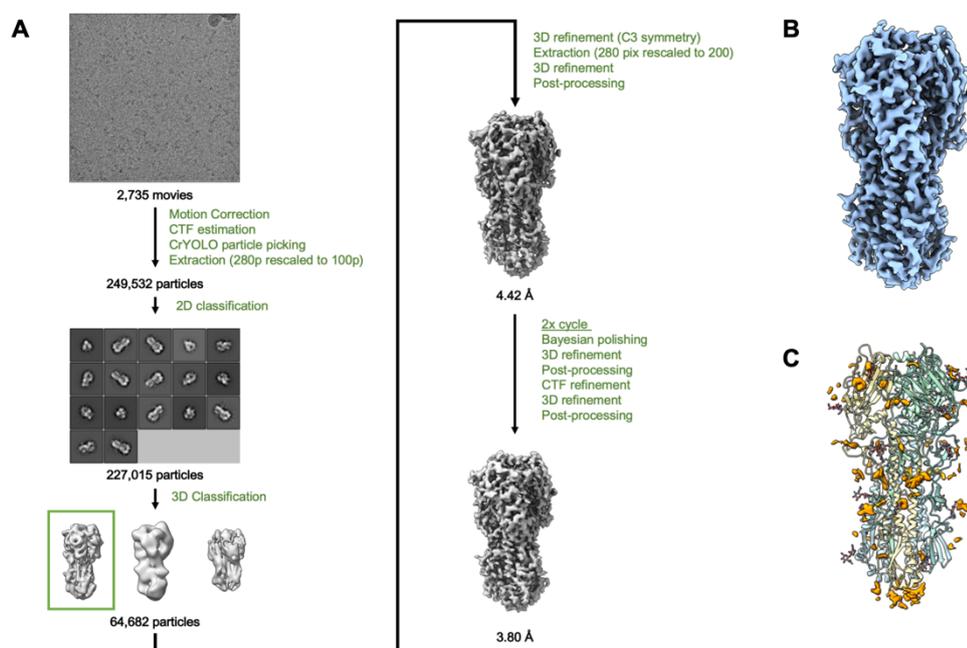
**Figure 4.25 19-Fluorine NMR experiments showing Affimer-based compound OA3 binding to hemagglutinin (HA).** A) Structure of OA3 showing its fluorine atom. B) 19-F CPMG of OA3-only and OA3 + monomeric HA at 10 ms and 300 ms. C) 19-F CPMG of OA3-only and OA3 + trimeric HA.

#### 4.3.3.6 Cryo-EM of HA with OA3 using a device for fast grid preparation generated an apo HA structure

Determining a high-resolution structure of OA3 bound to HA would have been useful both to assess whether OA3 bound at the same location on HA as its template Affimer and to guide future structural optimisation of OA3. Since the sprayer for fast cryo-EM grid preparation and time-resolved cryo-EM generated the first A5-bound HA structure, the same method was employed for the OA3-bound structure. OA3 and HA trimer (Provided by the Wilson group) were pre-mixed and sprayed onto a grid, imaged on a Krios microscope and processed in RELION by David Klebl and myself (Figure 4.26A). The 2D classes showed the highest variation in HA side views on a grid generated by the sprayer. HA particles were classified and refined into a final map with a global resolution of 3.8 Å (Figure 4.26B).

Docking in an atomic model of HA identified that there were glycosylated residues in the map at HA1 residues N165 and N285 and HA2 residue N154, so a crystal structure of the HA variant used to make grids that contains the same glycosylated residues was docked in. The waters were deleted as the resolution was not high enough to visualise

water molecules, then the remaining glycoprotein model was refined into the map. In order to locate OA3, regions of unmodeled density were identified in the map (Figure 4.26C). However, there were no clear unambiguous areas of density. Therefore, no clear conclusions can be drawn at this resolution, which is potentially too low for resolving small molecules.



**Figure 4.26 Using a sprayer for rapid cryo-EM grid preparation to generate a map of HA with OA3 bound. The resulting map was ambiguous for the position of the ligand.** A) The image processing workflow using RELION to generate a final map of 3.8 Å global resolution. B) Final map with 3.80 Å global resolution. C) Glycoprotein model of HA from influenza virus (A/Hong Kong/1/1968 (H3N2)) with areas of unmodeled cryo-EM density (orange surface).

## 4.4 Discussion

The aim of this chapter was to test the small molecules developed using VS methods in chapter three. This included the AF1 series which was identified through a novel LBVS workflow whereby compounds were identified which had similar shapes to the binding region of NP-Affimer, an inhibitory Affimer of CCHFV NP. CCHFV is a tick-borne virus with a fatality rate estimated at 5-30% and NP is involved in packaging and protecting its RNA genome. In addition, the AF2 compound series was developed using a traditional SBVS workflow around the CCHFV NP RNA-binding site, which was predicted to be a druggable site by two *in silico* methods.

The first screening assays for the AF1 and AF2 compounds were cellular. Although AF1-03 showed cytotoxicity just above the threshold, it was still taken forward as it was close. The issues with AF1-03 cytotoxicity could be circumvented in the future by

testing lower AF1-03 concentrations, in order to identify one where AF1-03 is non-cytotoxic but still biologically active.

Non-cytotoxic compounds were screened against the CCHFV MGS, which uses GFP expression as a proxy for CCHFV replication. Three compounds including AF1-03 caused a significant reduction in CCHFV replication above the threshold for inhibition. Compounds were tested using a different set up than that used during NP-Affimer characterisation by Alvarez-Rodriguez *et al.*<sup>86</sup>, whereby cells were co-transfected in a 12-well plate with a NP-Affimer expression plasmid and the MGS plasmids. However, this method was not high-throughput enough to convert to compound screening and did not miniaturise to a 96-well plate format. So, cells were first transfected with the MGS in a 6-well plate, allowing for high transfection efficiency, and after 24 hours they were re-seeded into a 96-well plate with the compounds, increasing the throughput and reducing the inter-well variation in fluorescence. The throughput could even be increased further by re-seeding cells into 384-well plates, which are compatible with the Incucyte live cell imager used for the assay. However, a disadvantage of the new set-up was that NP-Affimer did not function as a positive control, likely because expression of NP-Affimer when the MGS was already established was too late to inhibit CCHFV replication. To circumvent this, the MGS could be tested in a stable cell line which only expresses NP-Affimer when induced to do so by the addition of a molecule like Doxycycline.

Compounds were screened for binding to CCHFV NP using a range of assays. SPR and DSF-based thermal shift assays identified some of the SBVS-based AF2 compounds as binders of NP including AF2-02 and AF2-07. However, both assays had problems with compound interference, so dye-free nanoDSF-based thermal shift assays were employed instead. This identified several binders including AF1-03, making this the first example of a compound that was generated through the novel Affimer-guided LBVS method and that binds and inhibits the same target as its template Affimer. The  $K_D$  of AF1-03 was calculated using a new isothermal approach to be  $\sim 150 \mu\text{M}$ , which is low but could be improved by structural optimisation of AF1-03. To aim optimisation, we tried to generate a structure of AF1-03 bound to CCHFV NP using crystallography but this was unsuccessful.

Despite the lack of structural guidance, analogues of AF1-03 termed the AF3 series were developed, but these were not that similar to the template compound, indicating the compound libraires did not contain the chemical space for developing analogues of AF1-03. None of the AF3 series bound and inhibited CCHFV NP, highlighting a limitation of commercial libraries. Although they may contain initial hits, as evidenced by AF1-03 and OA3, the size limit of libraries means they have to balance containing a broad range of scaffolds as well as a good selection of analogues. These issues might be solved by screening the ultra-large Enamine REAL and REAL space libraries which

contain a wider selection of scaffolds and analogues, or through having bespoke compounds synthesised by a medicinal chemist.

The AF1 and AF2 compounds allowed for a comparison of compound development through the novel Affimer-guided LBVS workflow and a traditional SBVS workflow. Several of the AF2 compounds including AF2-02 and AF2-07 bound CCHFV NP but did not inhibit replication in the MGS assay and AF2-10 had a  $K_D$  of  $\sim 10 \mu\text{M}$  but was too cytotoxic to test against the MGS. These non-inhibitory binders highlight that SBVS programs only screen for binding and do not possess the intimate knowledge of protein function and inhibition required to translate binding into biological activity. The Affimer-guided method is advantageous in this respect because by using an inhibitory template, the resulting small molecules were more likely to be inhibitory themselves. Although, this pattern was for one system and may not be universal.

The OA compounds also were developed based on the Affimer A5 using the same LBVS workflow that generated the AF1 series but without guidance from a high-resolution target-Affimer structure. A5 inhibits the IAV spike protein HA, which mediates viral entry into the cell. OA3 was identified as the only inhibitor of IAV infection in the series using the cell-based TCID<sub>50</sub> assay with an IC<sub>50</sub> of 11  $\mu\text{M}$ . SPR, DSF and nanoDSF were employed to detect binding of OA3 to HA, but none were successful due to compound interference. It was only through using <sup>19</sup>F-NMR that we were able to identify OA3 binding to HA.

It is important that a small molecule Affimer mimic binds at the same site and has the same mode of action as its template Affimer. We tried to verify this for AF1-03 and OA3 using X-ray crystallography and cryo-EM, respectively, but in both cases, the resulting map did not contain the small molecule, so their exact binding modes remain unknown.

A range of binding assays were employed in this chapter, allowing the comparison of how effective each assay is. SPR is an established technique for screening small molecules, but it was not useful in this case due to non-specific binding of compounds to the chip. For most compounds, any binding signal was masked by non-specific sticking, and AF2-02 and AF2-07 were the only binders that were detected via SPR. DSF-based thermal shift assays identified some additional compounds on top of SPR, but still many compounds caused the dye to fluoresce in the absence of protein. NanoDSF proved to be a particularly useful screening technique, being the only method able to detect AF1-03 binding to CCHFV NP. NanoDSF is dye-free, removing one source of interference, and it can be used to calculate binding affinities using the isothermal method employed here. A main drawback is that the Prometheus NT.48 nanoDSF machine used here is not as high-throughput as the machines for SPR and DSF because the Prometheus only analyses 48 capillaries per run which require manual loading. However, there are more high-throughput nanoDSF systems available, such as the Prometheus Panta from NanoTemper, which can screen 1536 capillaries,

or the SUPR-DSF by Malvern Panalytical, which measures samples in a 384-well plate. However, these are not yet available to use near to Leeds.

The aim of this chapter was to identify hits from the compound series developed in chapter 3 through screening them in cell-based and biophysical assays. In summary, screening assays identified two hits from compound series generated using a novel Affimer-guided approach: AF1-03, which bound its target CCHFV NP and inhibited CCHFV replication; and OA3, which bound its target HA and inhibited IAV infection. Several of the AF2 series, which were generated through a traditional SBVS workflow, demonstrated binding to CCHFV NP but not inhibition of CCHFV replication, indicating a disadvantage of using SBVS compared to the novel Affimer-guided approach for this target.

#### **4.4.1 Recommended screening cascade**

Through completing these screening workflows, the advantages and disadvantages of each assay have been identified. Figure 4.27 shows the recommended workflow for screening Affimer-based small molecules in the future. Combined with Figure 3.20, this forms a full workflow for discovering small molecules based on Affimers. The workflow outlined here is for screening ~50 starting compounds with a quantity of protein sample that limits screening to only the necessary compounds.

The workflow starts by identifying inhibitors of the system in cells, because this is the fundamental function of a therapeutic small molecule. Firstly, compounds are screened in a cytotoxicity assay, so that cytotoxic compounds can be removed prior to screening for inhibition screening. To increase throughput, compounds should be screened at a single concentration such as 10  $\mu$ M. Although compounds may be missed that are not active at 10  $\mu$ M, that is the trade-off for higher throughput. Then, compounds can be screened for inhibition of the target system in a cell-based assay. Again, compounds should be screened at a single concentration, but any inhibitors can then be tested at a range of concentrations to determine their potencies.

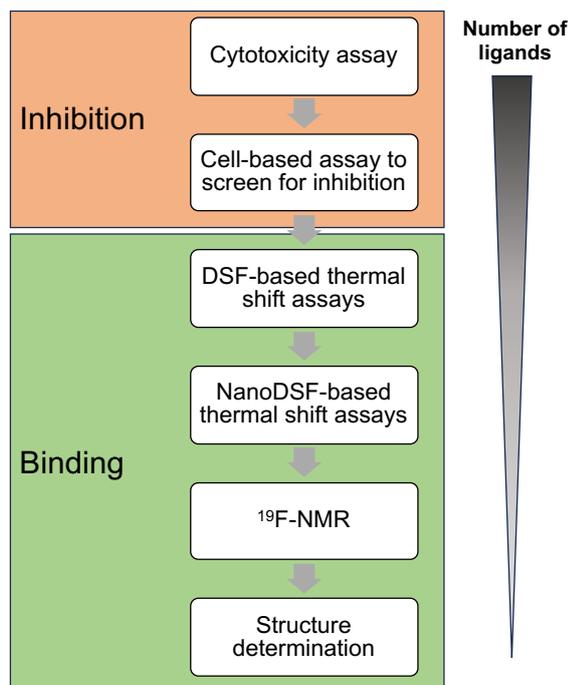
If there are fewer than 15 starting compounds and throughput is not an issue, they could be screened at a range of concentrations in the cytotoxicity and inhibition assays, as was performed for the OA compounds in section 4.3.3.1. Three replicate wells per compound is recommended for the cell-based assays, due to their high variability, but this can be reduced to two replicates to increase the throughput if there are >50 compounds to screen. Furthermore, 96-well plates are recommended for cell-based assays, but they could be miniaturised to 384-well plates to increase throughput.

Once inhibitory compounds are identified, those that bind the target are identified. A DSF-based thermal shift assays is the best leading binding assay for this as it is in-solution and plate-based. The first step is to identify the lowest concentration of target protein which generates a melt with strong signal to noise, which is used to calculate

the ligand screening concentration as this should be 100x the protein concentration. The Affimer template should be used as positive control since Affimers are typically thermostable up to 100 °C. The compounds must be screened without protein to identify those that interact with the dye and will mask the protein signal. Any non-interfering compounds are then screened against the protein to identify those that generate a significant positive  $T_m$  shift. Two replicates is sufficient for DSF as it has low variability. Binders can be tested at a concentration range to calculate affinities using the isothermal approach.

Any compounds that caused fluorescence in the DSF-based thermal shift assays can then be tested in nanoDSF-based thermal shift assays using the Prometheus NT.48 machine at The University of York. As for DSF, the optimum target protein concentration should be identified and used to calculate ligand screening concentration at 100x molar excess. Compounds that emit fluorescence at 330 nm or 350 nm should be identified by performing a ligand-only control, then non-interfering compounds should be screened against target protein. Running samples in duplicate capillaries, means 24 ligands can be screened per run on the Prometheus NT.48. The affinity can then be calculated for any binders using the isothermal method. If the sample quantities allow, then all compounds could be tested by DSF and nanoDSF thermal shift assays to identify non-inhibitory binders which may have uses as tool compounds, and to verify the results from one binding assay with another.

For any compounds which interfere with DSF and nanoDSF and contain a fluorine atom, then  $^{19}\text{F}$ - NMR represents a useful tool to measure binding. This is more sample intensive and low throughput than the previous techniques but it is less likely for a compound to interfere with it. After the screening cascade, any inhibitory binders should be taken forward for structure determination to verify that they bind at the same site as their template Affimer, if this is known.



**Figure 4.27** Template workflow for screening compounds developed using an Affimer-guided virtual screening workflow

# Chapter 5 : Discussion

## 5.1 Discussion

The aim of this PhD project was to investigate how to accelerate the small molecule discovery pipeline using protein structures.

Chapter 2 addressed protein structure generation and how it has been enhanced by advances in cryo-EM. The structures of Ax qNOR and bovine Cyt bc1 had been determined previously to ~4 Å resolution by cryo-EM<sup>107,108</sup>. By re-collecting cryo-EM data on better detector setups, new structures were generated at ~2 Å resolution, providing further information on electron donor binding and the roles of water molecules and lipids, which could all be used to aid identification of small molecule modulators of the target. It then discussed how advances in technology have facilitated this increase in resolution.

Chapter 3 investigated how protein structures can be used to guide the generation of small molecule inhibitors beyond traditional SBVS workflows. In particular, how Affimers can be used to identify small molecule binding to a protein target, since the applications of Affimers can be limited due to their inability to reach some locations e.g. they are non-membrane permeable. A LBVS workflow was developed to generate small molecule mimics of Affimer binding regions using 3D shape and electrostatic similarity algorithms. The workflow was applied to NP-Affimer, an inhibitory Affimer of CCHFV NP, and A5, an inhibitory Affimer of the HA spike protein of IAV. For NP-Affimer, the two NP-Affimer loops were narrowed down to the five key binding residues. In addition, a traditional SBVS workflow was performed at the CCHFV NP RNA-binding site, generating the AF2 series to compare the two approaches.

In chapter 4, the compounds generated by virtual screening were tested for their ability to bind and inhibit their target protein. A panel of screening assays identified hits in each Affimer-based series, namely AF1-03 and OA3, indicating that the Affimer-based LBVS workflow was successful and that an Affimer-target structure is not essential for the method. Comparatively, the SBVS approach generated several binders that did not inhibit CCHFV replication, suggesting that the RNA-binding site is not an effective inhibitory site, and suggesting some advantages of LBVS techniques guided by Affimers.

The improvement in ~2 Å resolution of the Ax qNOR and bovine Cyt bc1 cryo-EM maps compared to those generated ~5 years prior matches the trend that resolutions of cryo-EM maps are increasing year-on-year. These improvements came about primarily from advances in cryo-EM detectors which collect better quality data more rapidly. For example, only 5.6% of the cytochrome bc1 newly-acquired particles were required to generate a map with the same resolution as the one generated previously<sup>107</sup>. These improvements have led to a shortening in the lengths of cryo-EM data collections from the previous 24-72 hours to 4-12 hours per dataset, depending on the

sample<sup>70,141</sup>. Multiple datasets of different ligands bound to the same target can now be collected within a single day using new tools like multigrid EPU to automate this<sup>70,135</sup>.

This increase in the speed of cryo-EM structure determination changes the questions in drug discovery that cryo-EM can answer. Rather than visualising the best inhibitor bound to a target, a panel of inhibitors could be visualised to screen for those with favourable binding modes. To further accelerate this screening, the overall binding sites could be quickly identified for all ligands through 3-4 Å structures, then structures with the promising compounds could be determined to 2-3 Å by collecting more data. Screening a range of inhibitors by cryo-EM, rather than the single most promising, would identify compounds with more diverse modes of inhibition, broadening and accelerating the ligand discovery pipeline.

Assuming the speed of cryo-EM structure determination continues to rise, it is useful to identify the bottlenecks in the pipeline and how they might be removed. Processing multiple different datasets requires good organisation and automation of the full image processing workflow. Although more automation is becoming available, like automated 2D class selection in Relion 4.0<sup>105</sup>, setting up a fully automated processing pipeline is still challenging in Relion. Furthermore, the Relion GUI is still set-up for individual projects with bespoke processing pipelines which makes high-throughput structure-determination challenging. Conversely, the cryoSPARC GUI better organises processing, dividing different datasets into projects and workspaces that can be easily inspected, and it better facilitates the queuing of automated processing pipelines<sup>104</sup>. However, since cryoSPARC requires a for-profit license that starts at ~£100,000, smaller, younger companies engaging in cryo-EM are forced into using the cheaper alternative software Relion. Thankfully, automation in Relion is likely to improve soon with the release of Doppio, the new CCPEM GUI, which will facilitate automated Relion processing pipelines and allow for better organisation of different projects<sup>250</sup>. Furthermore, the software will even include automated model building using the neural network-based ModelAngelo, meaning the pipeline will be able to convert cryo-EM movies into an atomic structure with little user input<sup>251</sup>. Although, a potential issue with such automated, unsupervised processing pipelines is that they could lead to incorrectly calculated structures.

Another bottleneck to high throughput cryo-EM beyond image processing automation is the slow speed of changing grids. Krios autoloader cassettes contain slots for 12 grids and it takes ~4 minutes to change grids within the column and ~10 minutes to atlas each grid. To screen more than 12 grids, the autoloader cassette must then be removed, the grids changed, then the cassette reinserted. The total time for these processes is a minimum of 40 minutes, which is microscope downtime during which data cannot be collected. There are three advances that would reduce this bottleneck:

increasing the autoloader speeds; adding more slots to autoloader cassettes, and the ability for a grid to contain multiple samples, perhaps with different ligands. The first two are problems that must be addressed by microscope manufacturers, but the latter is one that users might be able to solve. A grid containing the same protein mixed with different ligands depending on the location around the grid would require the automated dispensing of protein and compound onto distinct areas of the grid. There are two systems that might allow for this. Firstly, the Chameleon (SPT Labtech) is a machine for automated cryo grid preparation which dispenses sample in a stripe on an EM grid that is travelling into liquid ethane, similar to the nanosprayer used here to visualise HA<sup>65,252</sup>. A Chameleon with multiple nozzles could potentially dispense a target mixed with different ligands in adjacent stripes on the same grid. The presence of dry squares between the stripes would validate that there was no mixing of different ligands. With each stripe being ~3 grid squares wide, on a typical 300-mesh grid with 30 squares in each direction, ~10 different protein-ligand mixtures could be dispensed<sup>129</sup>. In addition, the Chameleon can dispense and freeze grids in ~54 ms, allowing the structure determination of proteins with preferred orientation like HA<sup>252</sup>. The second system is the Vitrojet (CryoSol), which prints sample onto an EM grid before vitrifying it<sup>253</sup>. If a VitroJet had multiple nozzles, as was suggested in the original VitroJet paper by Ravelli *et al.*, a grid could potentially be produced with regions containing different ligands<sup>253</sup>. The VitroJet can dispense sample in smaller regions of 3x3 squares, potentially allowing for up to 100 ligands to be dispensed per grid. The ligand at each grid location could be correlated using labelled cryo grids like Finder grids (Agar Scientific). With 100 ligands/per grid and 12 grids/autoloader, >1000 ligands could be screened per session, moving cryo-EM even further upstream in the screening pipeline, being used to screen for binding of untested ligands to a target, rather than for information on the binding modes of validated targets. Although it is worth noting we will need further increases in throughput to achieve these screening rates.

With such increases in cryo-EM throughput, a cryo-EM-based fragment screening pipeline could be provided at national cryo-EM facilities, similar to XChem at the Diamond Light Source<sup>52</sup>. At X-Chem, protein crystals are soaked in a library of compound fragments, then X-ray crystal structures are solved of the protein in the presence of each ligand. This screens for binding of the fragment to the target based on its presence or absence within the electron density, and any target-bound structure generated can then be used to rationally develop the fragment into a lead molecule. The same principles could be applied to cryo-EM, given a high enough throughput that is likely to occur in the future.

In a world where such high throughput is possible for cryo-EM single particle analysis (SPA), and high resolution structures can be generated automatically without significant user input, the focus would likely to move to challenging targets and

visualising proteins in a truly native environment. The disadvantage of SPA is that proteins are visualised away from their native cellular environment potentially without potential binding partners or stabilising molecules. For example, membrane proteins, which make up the majority of drug targets, are extracted from their environment of the lipid bilayer and analysed by cryo-EM in membrane mimetic systems like detergents, amphipols or styrene-maleic acid lipid particles (SMALPs)<sup>150,254,255</sup>. Although amphipols and SMALPs are closer to the native environment, none are exact mimics of the bilayer. In particular, ion channels, which make up a large proportion of drug targets, change state depending on changes in their environment including membrane polarisation or mechanical stress<sup>53</sup>. It is very challenging to visualise extracted ion channels under such changes in their membrane environment by cryo-EM SPA.

With cryo electron tomography (cryo-ET), targets can be visualised within the cell, reducing concerns about how native the resulting structure is<sup>256</sup>. The resolution limit caused by the low electron dose applied during tomography can be improved by sub-tomogram averaging (STA), the averaging of repeated regions within tomograms. The resolution of maps generated by STA have been improving greatly thanks to new software like Warp<sup>257</sup>, M<sup>55</sup> and Relion-4.0<sup>258</sup>, leading to impressive structures generated through STA including the membrane coat of trafficking vesicles determined to 12 Å resolution, with the inner layer resolved to 4.6 Å<sup>259</sup>, the bacterial surface layer resolved to 3.7 Å<sup>260</sup> and the SARS-CoV-2 spike in whole viruses resolved to 4.9 Å<sup>261</sup>. Its large size of ~2 MDa in bacteria and 3.2 MDa in eukaryotes means the ribosome has been studied extensively in the cellular environment by STA, uncovering the different conformational states it undergoes during translation, and even visualising antibiotic binding at 3.5 Å resolution<sup>55,262</sup>. So far, the application of cryo-ET and STA to membrane proteins has been limited by the challenges in identifying specific proteins within tomograms of the crowded lipid bilayer. Tagging the protein of interest can be done with quantum dots that have high electron scattering potential<sup>263</sup> or DNA origami signposts which are large and physically distinguishable<sup>264</sup>, but so far we are yet to see the breakthrough required for this challenge. In summary, enabling small molecule discovery using cryo-ET is promising but likely requires several scientific developments and years to be fully realised. To give it an approximate timeframe, there were eight years between the publishing of the first <4 Å eukaryotic ribosome structure by cryo-EM SPA (2014)<sup>265</sup> and cryo-ET (2022)<sup>262</sup>.

Once protein structures have been determined, they enable small molecule discovery by identifying predicted binders through SBVS, turning hits in lead molecules through SBDD and turning fragments into lead molecules through FBDD. Here, I identified an alternative use of protein structures whereby Affimer structures were used as templates to guide small molecule development, either using an experimentally derived structure like the CCHFV NP:NP-Affimer structure or a predicted structure like the A5 homology

model<sup>86</sup>. This method used a LBVS workflow based on the shape similarity program ROCS, which was employed in a similar fashion by Celis *et al.* to develop small molecule inhibitors of protein-protein interactions (PPIs)<sup>34,213</sup>. Although, there were some key differences in the approaches. Celis *et al.* generated inhibitors of hDM2-p53 interactions by using a section of p53 as a template for ROCS, so the resulting inhibitors blocked native interactions formed by their template<sup>34</sup>. Whereas, the Affimer-guided small molecules described here would not block native interactions by their template because Affimers are not natively expressed. Although, their large size means Affimers do tend to disrupt PPIs, for example A5 blocks HA-receptor binding and NP-Affimer blocks NP oligomerisation, so Affimer-guided small molecules will also likely mimic these properties.

Since beginning the project developing the Affimer-guided approach, the size of commercially available libraries has grown extraordinarily. For example, the eMolecules library contained 12.6 million unique molecules when it was screened here and now contains 20.3 million molecules<sup>228</sup>. Furthermore, there are now make-on demand compound libraries which did not exist when I began the project, such as the Enamine REAL SPACE library of 36 billion<sup>266</sup>. In order to increase the rate of hit identification, future Affimer-guided small molecule screens should make use of these larger libraries with vaster chemical space. Compared to docking algorithms used in SBVS, shape similarity algorithms for LBVS screen compounds faster because they search through a considerably smaller space, meaning the Affimer-guided approach is better suited to screening these larger libraries. The local installation of CPU-based ROCS used here has now been improved upon by a GPU-based FastROCS algorithm which runs fastest when run in the cloud<sup>267</sup>. OpenEye report that whereas CPU-based ROCS searched through one billion compounds in 11 CPU years (2 days on 2000 CPU cluster), FastROCS screened one billion compounds for shape similarity in 5 GPU days (<1 day on a 4 GPU cluster) when run locally and 30 minutes when using GPU cores in the cloud via Amazon Web Services (AWS)<sup>267</sup>. Ten billion compounds were screened using FastROCS in one hour using >200 GPU cores rented from AWS<sup>268,269</sup>. For comparison, docking one billion compounds using the VirtualFlow program would take ~15 hours using a considerable 160,000 CPUs available through cloud computing services<sup>39</sup>.

As the speeds of docking and shape similarity algorithms increase, it allows for the screening of proteins in different conformations to overcome the problems of proteins being highly flexible and dynamic. This idea is complementary with cryo-EM, which can generate structures of a protein in different conformation more easily than X-ray crystallography. For example, multiple protein conformations can be determined from micrographs of the same cryo-EM grid<sup>64</sup>.

Following the established method for Affimer isolation with the new method for Affimer-guided small molecule development could lead to a fast workflow for generating new inhibitors, in particular for emerging virus threats. Affimer isolation takes ~12 days and is highly automated<sup>83</sup>. The best Affimers would then require further testing for binding and inhibition, which could take two weeks using robust, automated assays.

Generating an Affimer-bound target structure was helpful but not essential for small molecule development. The amount of time required to generate a new structure by cryo-EM is highly variable, but it took only two months from the emergence of SARS-CoV in December 2019 to the publication of the cryo-EM structure of its spike protein in February 2020<sup>270</sup>. The LBVS took <1 day and compounds took three weeks to arrive, although this could likely be sped up in emergencies. With an established target, robust assays and automated data processing, testing could potentially be completed in ~2 weeks if completed by a team. In total, Affimer-guided small molecule hit discovery could be completed in 3-4 months.

As well as Affimers, there are many additional antibody-like proteins that are isolated against targets and could serve as templates in the LBVS workflow. For example, nanobodies are the antigen-specific portions of single chain antibodies found in camelids<sup>271</sup>. Like Affimers, they can be isolated using phage display and expressed in *E. coli*<sup>271,272</sup>. They are 12-14 kDa in size and bind to targets via three variable loops denoted CDR1, CDR2 and CDR3, with CDR3 being the main contributor for antigen recognition. Although CDR3 is on average 18 residues long, making it longer than a 9-residue Affimer loop, it could be focused to a core interaction site using a target-nanobody structure, as was performed here for NP-Affimer loop two. On the other hand, DARPinS are 14-21 kDa proteins that consist of adjacent, parallel helices which form a concave binding surface<sup>273</sup>. The extended shape of this binding surface makes it much larger than the binding loops in Affimers or Nanobodies and too extensive to accurately mimic using a small molecule. Therefore, as well as Affimers, the LBVS workflow described here could likely be applied to other antibody-like proteins with small binding regions.

## 5.2 Final thoughts

In summary, protein structure determination by cryo-EM continues to improve in terms of data quality, speed of collection, the resolution of 3D reconstructions and the level of automation. These factors are changing the questions we can answer in drug discovery. Cryo-EM can now be used to screen a panel of tens of compounds for their binding modes, and in the near future, we will likely even use it to screen compound fragments for binding. Furthermore, the information gained from the higher resolution structures determined nowadays is rich. At resolutions close to 2 Å, we can visualise

water molecules, lipids bound to membrane proteins, more unambiguous side chain positions and ligand binding at high details.

Once these protein structures are generated, there are various points in which they can feed into the small molecule development pipeline, even beyond traditional SBVS. For example, structures of Affimers bound to their target can feed into a new Affimer-guided method of small molecule discovery based on LBVS programs described here. This novel method was validated by the identification of two Affimer-guides small molecules as inhibitors of CCHFV and IAV, which can be used for future optimisation into lead candidates for early stage drug discovery projects.

## Appendix

Table 5.1 Information on the cryo-EM data collection and model building for a structure of Ax qNOR

<b>Data Accession</b>	
PDB	8BGW
EMDB	16041
<b>Data Collection</b>	
Microscope	FEI Titan Krios
Voltage (kV)	300
Detector	Falcon 4
Energy filter slit width (eV)	10
Nominal magnification	130k
Pixel size (Å/pixel)	0.91
Defocus range (μm)	-0.9 - -2.7
Exposure time	6.11
Frames	44
Exposure rate (e <sup>-</sup> pixel <sup>-1</sup> s <sup>-1</sup> )	4.61
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	34.90
Dose per frame (e <sup>-</sup> /Å <sup>2</sup> )	0.8
Micrographs collected	5466
<b>Reconstruction</b>	
Software	RELION 3.1
Particles used in refinement	404,950
Symmetry	C2
Overall resolution when FSC=0.143 (masked) (Å)	2.2
Map sharpening B-factor (Å <sup>2</sup> )	-47.68
<b>Model Refinement</b>	
Software	REFMAC5
Non-hydrogen atoms	12845

Protein residues	11649
Ligands	746
Average B factors (Å <sup>2</sup> )	61.128
Protein	63.29
Ligands and water	91.95
<b>R.M.S. deviations</b>	
Bond length (Å)	0.012
Bond angle (Å)	1.649
<b>Ramachandran statistics (%)</b>	
Outliers	0.27
Allowed	5.47
Favoured	94.26
MolProbity score	2.36
ClashScore	15.19
Poor rotamers (%)	2.41
Model vs. Map FSC	0.854

**Table 5.2 Information on cryo-EM data collection for a structure of bovine cytochrome bc1**

<b>Data Accession</b>	
PDB	In progress
EMDB	In progress
<b>Data Collection</b>	
Microscope	FEI Titan Krios
Voltage (kV)	300
Detector	Falcon 4i
Energy filter slit width (eV)	10
Nominal magnification	165k
Pixel size (Å/pixel)	0.74
Defocus range (µm)	-0.9 - -2.7

Exposure time	3.22
Frames	51
Exposure rate ( $e^- \text{ pixel}^{-1} \text{ s}^{-1}$ )	6.9
Electron exposure ( $e^-/\text{\AA}^2$ )	40.57
Dose per frame ( $e^-/\text{\AA}^2$ )	0.8
Micrographs collected	8000
<b>Reconstruction</b>	
Software	RELION 4.0
Particles used in refinement	288,568
Symmetry	C2
Overall resolution when FSC=0.143 (masked) ( $\text{\AA}$ )	2.1
Map sharpening B-factor ( $\text{\AA}^2$ )	-37.56
<b>Model Refinement</b>	
Software	REFMAC5
Non-hydrogen atoms	36179
Protein residues	
Ligands	
Waters	2435
Average B factors ( $\text{\AA}^2$ )	62.296
Protein	
Ligands and water	
<b>R.M.S. deviations</b>	
Bond length ( $\text{\AA}$ )	0.009
Bond angle ( $\text{\AA}$ )	1.602
<b>Ramachandran statistics (%)</b>	
Outliers	0.88
Allowed	2.95
Favoured	96.17

MolProbity score	1.80
ClashScore	3.85
Poor rotamers (%)	3.08
Model vs. Map FSC	0.856

**Table 5.3 Information on cryo-EM data collection for a structure of bovine cytochrome bc1 with small molecule SG-114**

<b>Data Accession</b>	
PDB	In progress
EMDB	In progress
<b>Data Collection</b>	
Microscope	FEI Titan Krios
Voltage (kV)	300
Detector	Falcon 4i
Energy filter slit width (eV)	10
Nominal magnification	165k
Pixel size (Å/pixel)	0.74
Defocus range (µm)	-0.9 - -2.7
Exposure time	2.49
Frames	44
Exposure rate (e <sup>-</sup> pixel <sup>-1</sup> s <sup>-1</sup> )	7.81
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	35.51
Dose per frame (e <sup>-</sup> /Å <sup>2</sup> )	0.8
Micrographs collected	2,500
<b>Reconstruction</b>	
Software	RELION 4.0
Particles used in refinement	121,966
Symmetry	C2
Overall resolution when FSC=0.143 (masked) (Å)	2.0

Map sharpening B-factor (Å <sup>2</sup> )	-28.13
<b>Model Refinement</b>	
Software	REFMAC5
Non-hydrogen atoms	36120
Waters	2587
Average B factors (Å <sup>2</sup> )	58.344
<b>R.M.S. deviations</b>	
Bond length (Å)	0.008
Bond angle (Å)	1.606
<b>Ramachandran statistics (%)</b>	
Outliers	0.64
Allowed	2.96
Favoured	96.40
MolProbity score	2.06
ClashScore	7.5
Poor rotamers (%)	3.48
Model vs. Map FSC	0.850

**Table 5.4 Information on cryo-EM data collection for a structure of bovine cytochrome bc1 with small molecule ligand CK-267**

	Collection 1	Collection 2
<b>Data Accession</b>		
PDB	In progress	In progress
EMDB	In progress	In progress
<b>Data Collection</b>		
Microscope	FEI Titan Krios	FEI Titan Krios
Voltage (kV)	300	300
Detector	Falcon 4i	Falcon 4i
Energy filter slit width (eV)	10	10
Nominal magnification	165k	165k

Pixel size (Å/pixel)	0.74	0.74
Defocus range (µm)	-0.9 - -2.7	-0.9 - -2.7
Exposure time	2.49	2.78
Frames	44	44
Exposure rate (e <sup>-</sup> pixel <sup>-1</sup> s <sup>-1</sup> )	7.81	6.99
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	35.51	35.49
Dose per frame (e <sup>-</sup> /Å <sup>2</sup> )	0.8	0.8
Micrographs collected	1936	5,312
<b>Reconstruction</b>		
Software	RELION 4.0	RELION 4.0
Particles used in refinement	44,627	192,016
Symmetry	C2	
Overall resolution when FSC=0.143 (masked) (Å)	2.0	
Map sharpening B-factor (Å <sup>2</sup> )	-35.77	
<b>Model Refinement</b>		
Software	REFMAC5	
Non-hydrogen atoms	34852	
Waters	872	
Average B factors (Å <sup>2</sup> )	48.776	
<b>R.M.S. deviations</b>		
Bond length (Å)	0.009	
Bond angle (Å)	1.690	
<b>Ramachandran statistics (%)</b>		
Outliers	0.96	
Allowed	2.96	
Favoured	96.08	
MolProbity score	1.71	
ClashScore	2.39	

Poor rotamers (%)	3.61
Model vs. Map FSC	0.798

## Bibliography

1. Lei, S., Chen, X., Wu, J., Duan, X. & Men, K. Small molecules in the treatment of COVID-19. *Signal Transduct. Target. Ther.* **7**, 1–39 (2022).
2. Beall, R. F., Hwang, T. J. & Kesselheim, A. S. Pre-market development times for biologic versus small-molecule drugs. *Nat. Biotechnol.* **37**, 708–711 (2019).
3. Schlander, M., Hernandez-Villafuerte, K., Cheng, C. Y., Mestre-Ferrandiz, J. & Baumann, M. How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment. *Pharmacoeconomics* **39**, 1243–1269 (2021).
4. Batta, A., Kalra, B. S. & Khirasaria, R. Trends in FDA drug approvals over last 2 decades: An observational study. *J. Fam. Med. Prim. Care* **9**, 105–114 (2020).
5. Plenge, R. M. Disciplined approach to drug discovery and early development. *Sci. Transl. Med.* **8**, 1–5 (2016).
6. Childs-Disney, J. L. *et al.* Targeting RNA structures with small molecules. *Nat. Rev. Drug Discov.* **21**, 736–762 (2022).
7. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discov.* **21**, 181–200 (2022).
8. New Drug Therapy Approvals 2022 | FDA. <https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/new-drug-therapy-approvals-2022>.
9. Foley, T. L. *et al.* Selecting Approaches for Hit Identification and Increasing Options by Building the Efficient Discovery of Actionable Chemical Matter from DNA-Encoded Libraries. *SLAS Discov.* **26**, 263–280 (2021).
10. Entzeroth, M., Flotow, H. & Condron, P. Overview of High-Throughput Screening. *Curr. Protoc. Pharmacol.* **44**, 9.4.1-9.4.27 (2009).
11. Dorr, P. *et al.* Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob. Agents Chemother.* **49**, 4721–4732 (2005).
12. Duffy, K. J. *et al.* Identification of a pharmacophore for thrombopoietic activity of small, non-peptidyl molecules. 1. Discovery and optimization of salicylaldehyde thiosemicarbazone thrombopoietin mimics. *J. Med. Chem.* **45**, 3573–3575 (2002).
13. Duffy, K. J. *et al.* Hydrazinonaphthalene and azonaphthalene thrombopoietin mimics are nonpeptidyl promoters of megakaryocytopoiesis. *J. Med. Chem.* **44**, 3730–3745 (2001).
14. Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M. & Taranto, A. G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* **8**, 481382 (2020).
15. Wong, P. C., Pinto, D. J. P. & Zhang, D. Preclinical discovery of apixaban, a direct and orally bioavailable factor Xa inhibitor. *J. Thromb. Thrombolysis* **31**, 478 (2011).
16. Butler, D. Translational research: Crossing the valley of death. *Nature* **453**, 840–842 (2008).
17. UK, C. R. Phases of clinical trials. <https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/what-clinical-trials-are/phases-of-clinical-trials>.

18. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).
19. Fan, H. *et al.* Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature* **573**, 287–290 (2019).
20. Keown, J. R. *et al.* Mapping inhibitory sites on the RNA polymerase of the 1918 pandemic influenza virus using nanobodies. *Nat. Commun.* **13**, 1–11 (2022).
21. Amaro, R. E. Will the Real Cryptic Pocket Please Stand Out? *Biophys. J.* **116**, 753–754 (2019).
22. Nussinov, R. & Tsai, C. J. Allostery in Disease and in Drug Discovery. *Cell* **153**, 293–305 (2013).
23. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 1–11 (2009).
24. Halgren, T. New Method for Fast and Accurate Binding-site Identification and Analysis. *Chem. Biol. Drug Des.* **69**, 146–148 (2007).
25. Ghersi, D. & Sanchez, R. EasyMIFs and SiteHound: a toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics* **25**, 3185–3186 (2009).
26. Brylinski, M. & Feinstein, W. P. EFindSite: Improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands. *J. Comput. Aided. Mol. Des.* **27**, 551–567 (2013).
27. Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A. S. & De Fabritiis, G. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* **33**, 3036–3042 (2017).
28. Meller, A. *et al.* Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat. Commun.* **2023** **14**, 1–15 (2023).
29. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* vol. 432 862–865 (2004).
30. Houston, D. R. & Walkinshaw, M. D. Consensus docking: Improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.* **53**, 384–390 (2013).
31. Zoete, V., Daina, A., Bovigny, C. & Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. *J. Chem. Inf. Model.* **56**, 1399–1404 (2016).
32. Rush, T. S., Grant, J. A., Mosyak, L. & Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **48**, 1489–1495 (2005).
33. Wang, L. *et al.* Discovery of new selective human aldose reductase inhibitors through virtual screening multiple binding pocket conformations. *J. Chem. Inf. Model.* **53**, 2409–2422 (2013).
34. Celis, S. *et al.* Query-guided protein-protein interaction inhibitor discovery. *Chem. Sci.* **12**, 4753–4762 (2021).
35. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
36. REAL Database - Enamine. <https://enamine.net/compound-collections/real-compounds/real-database>.
37. Tingle, B. I. *et al.* ZINC-22—A Free Multi-Billion-Scale Database of Tangible

- Compounds for Ligand Discovery. *J. Chem. Inf. Model.* **63**, 1166–1176 (2023).
38. Gentile, F. *et al.* Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **17**, 672–697 (2022).
  39. Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663 (2020).
  40. Jaskolski, M., Dauter, Z. & Wlodawer, A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *FEBS J.* **281**, 3985 (2014).
  41. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **2016** *159* **15**, 605–619 (2016).
  42. Saur, M. *et al.* Fragment-based drug discovery using cryo-EM. *Drug Discov. Today* **25**, 485–490 (2020).
  43. Hartshorn, M. J. *et al.* Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **48**, 403–413 (2005).
  44. Murray, C. W., Newell, D. R. & Angibaud, P. A successful collaboration between academia, biotech and pharma led to discovery of erdafitinib, a selective FGFR inhibitor recently approved by the FDA. *Medchemcomm* **10**, 1509–1511 (2019).
  45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
  46. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science (80-. )*. **373**, 871–876 (2021).
  47. Mullard, A. What does AlphaFold mean for drug discovery? *Nat. Rev. Drug Discov.* **20**, 725–727 (2021).
  48. Dessau, M. A. & Modis, Y. Protein Crystallization for X-ray Crystallography. *J. Vis. Exp.* (2011) doi:10.3791/2285.
  49. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153 (2008).
  50. Maveyraud, L. & Mourey, L. Protein X-ray Crystallography and Drug Discovery. *Molecules* **25**, (2020).
  51. Goodwill, K. E., Tennant, M. G. & Stevens, R. C. High-throughput x-ray crystallography for structure-based drug design. *Drug Discov. Today* **6**, 113–118 (2001).
  52. Douangamath, A. *et al.* Achieving Efficient Fragment Screening at XChem Facility at Diamond Light Source. *JoVE (Journal Vis. Exp.* e62414 (2021) doi:10.3791/62414.
  53. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2016).
  54. White, J. B. R. *et al.* Single Particle Cryo-Electron Microscopy: From Sample to Structure. *J. Vis. Exp.* **2021**, (2021).
  55. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat. Methods* **18**, 186–193 (2021).
  56. Kühlbrandt, W. The Resolution Revolution. *Science (80-. )*. **343**, 1443–1444 (2014).
  57. García-Nafria, J. & Tate, C. G. Structure determination of GPCRs: cryo-EM compared with X-ray crystallography. *Biochem. Soc. Trans.* **49**, 2345 (2021).

58. Choy, B. C., Cater, R. J., Mancina, F. & Pryor, E. E. A 10-Year Meta-Analysis Of Membrane Protein Structural Biology: Detergents, Membrane Mimetics, And Structure Determination Techniques. *Biochim. Biophys. acta. Biomembr.* **1863**, 183533 (2021).
59. Wu, X. & Rapoport, T. A. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2115001118 (2021).
60. Han, Y. *et al.* High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1009–1014 (2020).
61. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10**, 1–7 (2019).
62. Macé, K. *et al.* Cryo-EM structure of a type IV secretion system. *Nature* **607**, 191–196 (2022).
63. Ma, J. *et al.* Structural basis of energy transfer in *Porphyridium purpureum* phycobilisome. *Nature* **579**, 146–151 (2020).
64. Kwon, D. H., Zhang, F., Fedor, J. G., Suo, Y. & Lee, S. Y. Vanilloid-dependent TRPV1 opening trajectory from cryoEM ensemble analysis. *Nat. Commun.* **13**, 1–12 (2022).
65. Kontziampasis, D. *et al.* A cryo-EM grid preparation device for time-resolved structural studies. *IUCrJ* **6**, 1024–1031 (2019).
66. Klebl, D. P., White, H. D., Sobott, F. & Muench, S. P. On-grid and in-flow mixing for time-resolved cryo-EM. *Acta Crystallogr. Sect. D, Struct. Biol.* **77**, 1233–1240 (2021).
67. Benton, D. J., Gamblin, S. J., Rosenthal, P. B. & Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* 1–4 (2020) doi:10.1038/s41586-020-2333-6.
68. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
69. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
70. Drulyte, I. *et al.* High-throughput cryo-EM epitope mapping of SARS-CoV-2 spike protein antibodies using EPU Multigrid. (2022).
71. Ceska, T., Chung, C. W., Cooke, R., Phillips, C. & Williams, P. A. Cryo-EM in drug discovery. *Biochemical Society Transactions* vol. 47 281–293 (2019).
72. Pinto, D. J. P. *et al.* Discovery of 1-(4-methoxyphenyl)-7-oxo-6-(4-(2-oxopiperidin-1-yl)phenyl)-4, 5,6,7-tetrahydro-1H-pyrazolo[3,4-c]pyridine-3-carboxamide (Apixaban, BMS-562247), a highly potent, selective, efficacious, and orally bioavailable inhibitor of blood coagulation factor Xa. *J. Med. Chem.* **50**, 5339–5356 (2007).
73. Boike, L., Henning, N. J. & Nomura, D. K. Advances in covalent drug discovery. *Nat. Rev. Drug Discov.* **21**, 881–898 (2022).
74. Davids, M. S. & Brown, J. R. Ibrutinib: a first in class covalent inhibitor of Bruton's tyrosine kinase. *Future Oncol.* **10**, 957 (2014).
75. Halford, B. Arvinas unveils PROTAC structures. *Chem. Eng. News* **99**, 5–5 (2021).
76. Senior, M. Fresh from the biotech pipeline: fewer approvals, but biologics gain share. *Nat. Biotechnol.* **41**, 174–182 (2023).
77. 50 best-selling pharmaceuticals of 2022: Market leaders and trends.

<https://www.drugdiscoverytrends.com/50-of-2022s-best-selling-pharmaceuticals/>.

78. Rau, R. Adalimumab (a fully human anti-tumour necrosis factor  $\alpha$  monoclonal antibody) in the treatment of active rheumatoid arthritis: the initial results of five trials. *Ann. Rheum. Dis.* **61**, ii70–ii73 (2002).
79. Mullard, A. FDA approves 100th monoclonal antibody product. *Nat. Rev. Drug Discov.* **20**, 491–495 (2021).
80. Lu, R. M. *et al.* Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* **27**, 1–30 (2020).
81. Baker, M. Upping the ante on antibodies. *Nat. Biotechnol.* **23**, 1065–1072 (2005).
82. Gray, A. *et al.* Animal-free alternatives and the antibody iceberg. *Nat. Biotechnol.* **38**, 1234–1239 (2020).
83. Tiede, C. *et al.* Affimer proteins are versatile and renewable affinity reagents. *Elife* **6**, 24903 (2017).
84. Tiede, C. *et al.* Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng. Des. Sel.* **27**, 145–55 (2014).
85. Haza, K. Z. *et al.* RAS-inhibiting biologics identify and probe druggable pockets including an SII- $\alpha$ 3 allosteric site. *Nat. Commun.* **12**, 1–15 (2021).
86. Álvarez-Rodríguez, B. *et al.* Characterization and applications of a crimean-congo hemorrhagic fever virus nucleoprotein-specific affimer: Inhibitory effects in viral replication and development of colorimetric diagnostic tests. *PLoS Negl. Trop. Dis.* **14**, 1–20 (2020).
87. Hesketh, E. L. *et al.* Affimer reagents as tools in diagnosing plant virus diseases. *Sci. Rep.* **9**, (2019).
88. Lopata, A. *et al.* Affimer proteins for F-actin: novel affinity reagents that label F-actin in live and fixed cells. *Sci. Rep.* **8**, 6572 (2018).
89. Cordell, P. *et al.* Affimers and nanobodies as molecular probes and their applications in imaging. *J. Cell Sci.* **135**, (2022).
90. Gil-Garcia, A. I. *et al.* Comparative diagnostic performance of the new chromatographic Affimer®-based rapid antigen detection against SARS-CoV-2 and other standard antigen tests for COVID-19 in a clinical setting. *medRxiv* 2022.01.18.22269401 (2022) doi:10.1101/2022.01.18.22269401.
91. Pipeline | Avacta Life Sciences Limited. <https://avacta.com/therapeutics/pipeline/>.
92. Lawson, A. D. G. Antibody-enabled small-molecule drug discovery. *Nat. Rev. Drug Discov.* **11**, 519–525 (2012).
93. Mullard, A. Antibody clamps pry open small-molecule drug discovery opportunities. *Nat. Rev. Drug Discov.* **21**, 247–248 (2022).
94. Saragovi, H. U. *et al.* Design and Synthesis of a Mimetic from an Antibody Complementarity-Determining Region. *Science (80- )*. **253**, 792–796 (1991).
95. Van Dongen, M. J. P. *et al.* An orally active small molecule fusion inhibitor of influenza virus. *Science (80- )*. **363**, (2019).
96. *Single-particle Cryo-EM of Biological Macromolecules*. (IOP Publishing, 2021). doi:10.1088/978-0-7503-3039-8.
97. McMullan, G., Faruqi, A. R., Clare, D. & Henderson, R. Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156–163 (2014).

98. GATAN. *K3 Direct Detection Cameras*. (2021).
99. Thermo Scientific. *Falcon 3EC Direct Electron Detector Datasheet*. <https://assets.thermofisher.com/TFS-Assets/MSD/Technical-Notes/Falcon-3EC-Datasheet.pdf> (2018).
100. Thermo Fisher Scientific. *Falcon 4i Direct Electron Detector Datasheet*. <https://assets.thermofisher.com/TFS-Assets/MSD/Datasheets/falcon-detector-datasheet-ds0371.pdf> (2022).
101. Thermo Fisher Scientific. Selectris Cryo Electron Microscopy Imaging Filter. <https://www.thermofisher.com/uk/en/home/electron-microscopy/products/accessories-em/selectris.html>.
102. Scientific, T. F. *Falcon 4 Detector Datasheet*. (2020).
103. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
104. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
105. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169 (2021).
106. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
107. Gopalasingam, C. C. *et al.* Dimeric structures of quinol-dependent nitric oxide reductases (qNORs) revealed by cryo–electron microscopy. *Sci. Adv.* **5**, eaax1803 (2019).
108. Ampornanai, K. *et al.* X-ray and cryo-EM structures of inhibitor-bound cytochrome bc<sub>1</sub> complexes for structure-based drug discovery. *IUCrJ* **5**, 200–210 (2018).
109. Gonska, N. *et al.* Characterization of the quinol-dependent nitric oxide reductase from the pathogen *Neisseria meningitidis*, an electrogenic enzyme. *Sci. Rep.* **8**, 1–13 (2018).
110. Householder, T. C., Fozo, E. M., Cardinale, J. A. & Clark, V. L. Gonococcal Nitric Oxide Reductase Is Encoded by a Single Gene, *norB*, Which Is Required for Anaerobic Growth and Is Induced by Nitric Oxide. *Infect. Immun.* **68**, 5241 (2000).
111. Bogdan, C. Nitric oxide and the immune response. *Nat. Immunol.* **2**, 907–916 (2001).
112. Gómez-Cerezo, J. *et al.* *Achromobacter xylosoxidans* bacteremia: a 10-year analysis of 54 cases. *Eur. J. Clin. Microbiol. Infect. Dis.* **22**, 360–363 (2003).
113. Aisenberg, G., Rolston, K. V. & Safdar, A. Bacteremia caused by *Achromobacter* and *Alcaligenes* species in 46 patients with cancer (1989–2003). *Cancer* **101**, 2134–2140 (2004).
114. Ramsey, B. W. Management of pulmonary disease in patients with cystic fibrosis. *N. Engl. J. Med.* **335**, 179–188 (1996).
115. Saiman, L. *et al.* Identification and Antimicrobial Susceptibility of *Alcaligenes xylosoxidans* Isolated from Patients with Cystic Fibrosis. *J. Clin. Microbiol.* **39**, 3942 (2001).
116. Burns, J. L. *et al.* Microbiology of Sputum from Patients at Cystic Fibrosis Centers in the United States. *Clin. Infect. Dis.* **27**, 158–163 (1998).

117. Hino, T. *et al.* Structural basis of biological N<sub>2</sub>O generation by bacterial nitric oxide reductase. *Science (80-. )*. **330**, 1666–1670 (2010).
118. Matsumoto, Y. *et al.* Crystal structure of quinol-dependent nitric oxide reductase from *Geobacillus stearothermophilus*. *Nat. Struct. Mol. Biol.* **2011 192** **19**, 238–245 (2012).
119. Jamali, M. A. M. *et al.* The active form of quinol-dependent nitric oxide reductase from *Neisseria meningitidis* is a dimer. *IUCrJ* **7**, 404–415 (2020).
120. Fisher, N., Meunier, B. & Biagini, G. A. The cytochrome bc<sub>1</sub> complex as an antipathogenic target. *FEBS Lett.* **594**, 2935–2952 (2020).
121. Zhang, Z. *et al.* Electron transfer by domain movement in cytochrome bc<sub>1</sub>. *Nature* **392**, 677–684 (1998).
122. McKeage, K. *et al.* Atovaquone/proguanil: A review of its use for the prophylaxis of *Plasmodium falciparum* malaria. *Drugs* **63**, 597–623 (2003).
123. Health Organization, W. *World malaria report 2022*. [https://cdn.who.int/media/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2022.pdf?sfvrsn=40bfc53a\\_4](https://cdn.who.int/media/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2022.pdf?sfvrsn=40bfc53a_4) (2022).
124. Xia, D. *et al.* Crystal structure of the cytochrome bc<sub>1</sub> complex from bovine heart mitochondria. *Science (80-. )*. **277**, 60–66 (1997).
125. Iwata, S. *et al.* Complete structure of the 11-subunit bovine mitochondrial cytochrome bc<sub>1</sub> complex. *Science (80-. )*. **281**, 64–71 (1998).
126. Guo, R., Zong, S., Wu, M., Gu, J. & Yang, M. Architecture of Human Mitochondrial Respiratory Megacomplex I2III2IV2. *Cell* **170**, 1247–1257.e12 (2017).
127. Swainsbury, D. J. K. *et al.* Cryo-EM structure of the four-subunit *Rhodobacter sphaeroides* cytochrome bc<sub>1</sub> complex in styrene maleic acid nanodiscs. *Proc. Natl. Acad. Sci. U. S. A.* **120**, (2023).
128. Ampornnanai, K. *et al.* Targeting the Ubiquinol-Reduction (Qi) Site of the Mitochondrial Cytochrome bc<sub>1</sub> Complex for the Development of Next Generation Quinolone Antimalarials. *Biology (Basel)*. **11**, 1109 (2022).
129. Thompson, R. F., Walker, M., Siebert, C. A. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3–15 (2016).
130. Ripstein, Z. A. & Rubinstein, J. L. Processing of Cryo-EM Movie Data. *Methods Enzymol.* **579**, 103–124 (2016).
131. Dubochet, J. & McDowell, A. W. Vitrification of pure water for electron microscopy. *J. Microsc.* **124**, 3–4 (1981).
132. Zi Tan, Y. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **2017 148** **14**, 793–796 (2017).
133. UK, T. F. S. EPU EM Software. <https://www.thermofisher.com/uk/en/home/electron-microscopy/products/software-em-3d-vis/eput-software.html#documents>.
134. Karuppasamy, M., Karimi Nejadasl, F., Vulovic, M., Koster, A. J. & Ravelli, R. B. G. Radiation damage in single-particle cryo-electron microscopy: Effects of dose and dose rate. *J. Synchrotron Radiat.* **18**, 398–412 (2011).
135. Scientific, T. F. *Smart EPU Software: A platform for automated screening and data acquisition in single particle analysis using cryo-EM.* (2022).
136. (EMDB), E. M. D. B. EMD software usage per year. [https://www.ebi.ac.uk/emdb/statistics/emdb\\_software\\_year](https://www.ebi.ac.uk/emdb/statistics/emdb_software_year).

137. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
138. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
139. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 1–13 (2019).
140. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).
141. Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat. Protoc.* **2018** *141* **14**, 100–118 (2018).
142. Moriya, T. *et al.* Size matters: optimal mask diameter and box size for single-particle cryogenic electron microscopy. *bioRxiv* 2020.08.23.263707 (2020) doi:10.1101/2020.08.23.263707.
143. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 22–25 (2010).
144. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. Sect. D, Struct. Biol.* **73**, 469–477 (2017).
145. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D, Struct. Biol.* **67**, 355–367 (2011).
146. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
147. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D, Struct. Biol.* **66**, 486–501 (2010).
148. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 1282–1291 (2021).
149. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
150. Flynn, A. J., Antonyuk, S. V., Eady, R. R., Muench, S. P. & Hasnain, S. S. A 2.2 Å cryoEM structure of a quinol-dependent NO Reductase shows close similarity to respiratory oxidases. *Nat. Commun.* **14**, 1–12 (2023).
151. Electron Microscopy Data Bank (EMDB). <https://www.ebi.ac.uk/emdb/>.
152. Lawson, C. L. *et al.* EMDatabank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403 (2016).
153. Lin, J., Balabin, I. A. & Beratan, D. N. The nature of aqueous tunneling pathways between electron-transfer proteins. *Science (80-. )*. **310**, 1311–1313 (2005).
154. Antonyuk, S. V., Han, C., Eady, R. R. & Hasnain, S. S. Structures of protein-protein complexes involved in electron transfer. *Nature* **496**, 123–126 (2013).
155. Stagg, S. M., Noble, A. J., Spilman, M. & Chapman, M. S. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *J. Struct. Biol.* **185**, 418–426 (2014).
156. Wright, D. J. *et al.* Human TRPC5 structures reveal interaction of a xanthine-based TRPC1/4/5 inhibitor with a conserved lipid binding site. *Commun. Biol.* **3**, 1–11 (2020).

157. Debski-Antoniak, O. *et al.* Exploiting the Affimer platform against influenza A virus. *bioRxiv* 2023.08.22.554342 (2023) doi:10.1101/2023.08.22.554342.
158. Abudurexiti, A. *et al.* Taxonomy of the order Bunyvirales: update 2019. *Arch. Virol.* **164**, 1949–1965 (2019).
159. Kuhn, J. H. *et al.* 2020 taxonomic update for phylum Negarnaviricota (Riboviria: Orthornavirae), including the large orders Bunyvirales and Mononegavirales. *Arch. Virol.* **165**, 3023–3072 (2020).
160. Aytekin, F. Y. *et al.* Factors related to fatalities and clinical progression of Crimean-Congo hemorrhagic fever patients and the effects of IL 28-B gene polymorphism. *Arch. Virol.* **164**, 547–557 (2019).
161. Swanepoel, R. *et al.* The Clinical Pathology of Crimean-Congo Hemorrhagic Fever. *Rev. Infect. Dis.* **11**, 794–800 (1989).
162. Lozach, P. Y. *et al.* Entry of bunyaviruses into mammalian cells. *Cell Host Microbe* **7**, 488–499 (2010).
163. Koch, J., Xin, Q., Tischler, N. D. & Lozach, P. Y. Entry of Phenuiviruses into Mammalian Host Cells. *Viruses* vol. 13 299 (2021).
164. Hulswit, R. J. G., Paesen, G. C., Bowden, T. A. & Shi, X. Recent Advances in Bunyavirus Glycoprotein Research: Precursor Processing, Receptor Binding and Structure. *Viruses* vol. 13 353 (2021).
165. Bente, D. A. *et al.* Crimean-Congo hemorrhagic fever: History, epidemiology, pathogenesis, clinical syndrome and genetic diversity. *Antiviral Research* vol. 100 159–189 (2013).
166. Mild, M., Simon, M., Albert, J. & Mirazimi, A. Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. *J. Gen. Virol.* **91**, 199–207 (2010).
167. Crimean-Congo haemorrhagic fever case identified in England, following travel to Central Asia - GOV.UK. <https://www.gov.uk/government/news/crimean-congo-haemorrhagic-fever-case-identified-in-england-following-travel-to-central-asia>.
168. Crimean-Congo haemorrhagic fever: origins, reservoirs, transmission and guidelines - GOV.UK. <https://www.gov.uk/guidance/crimean-congo-haemorrhagic-fever-origins-reservoirs-transmission-and-guidelines#cases-imported-into-the-uk>.
169. Jameson, L. J., Morgan, P. J., Medlock, J. M., Watola, G. & Vaux, A. G. C. Importation of *Hyalomma marginatum*, vector of Crimean-Congo haemorrhagic fever virus, into the United Kingdom by migratory birds. *Ticks Tick. Borne. Dis.* **3**, 95–99 (2012).
170. Public Health England. *Qualitative assessment of the risk that Crimean-Congo haemorrhagic fever (CCHF) virus presents to the UK human population*. <https://www.gov.uk/government/publications/hairs-risk-assessment-crimean-congo-haemorrhagic-fever> (2021).
171. Bonnet, S. I. *et al.* The control of *Hyalomma* ticks, vectors of the Crimean–Congo hemorrhagic fever virus: Where are we now and where are we going? *PLoS Negl. Trop. Dis.* **16**, e0010846 (2022).
172. Hawman, D. W. & Feldmann, H. Crimean–Congo haemorrhagic fever virus. *Nat. Rev. Microbiol.* **21**, 463–477 (2023).
173. Suda, Y. *et al.* Analysis of the entry mechanism of Crimean-Congo hemorrhagic fever virus, using a vesicular stomatitis virus pseudotyping system. *Arch. Virol.* **161**, 1447 (2016).
174. Xiao, X., Feng, Y., Zhu, Z. & Dimitrov, D. S. Identification of a putative Crimean-

- Congo hemorrhagic fever virus entry factor. *Biochem. Biophys. Res. Commun.* **411**, 253–258 (2011).
175. Garrison, A. R. *et al.* Crimean–Congo hemorrhagic fever virus utilizes a clathrin- and early endosome-dependent entry pathway. *Virology* **444**, 45–54 (2013).
  176. Hover, S. *et al.* Bunyavirus requirement for endosomal K<sup>+</sup> reveals new roles of cellular ion channels during infection. *PLoS Pathog.* **14**, (2018).
  177. Punch, E. K. *et al.* Potassium is a trigger for conformational change in the fusion spike of an enveloped RNA virus. *J. Biol. Chem.* **293**, 9937–9944 (2018).
  178. Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6895 (2001).
  179. Aurobindo Pharma. SmPC - Ribavirin 200 mg film-coated tablets. (2021).
  180. Keshtkar-Jahromi, M. *et al.* Crimean-Congo hemorrhagic fever: Current and future prospects of vaccines and therapies. *Antiviral Research* vol. 90 85–92 (2011).
  181. Johnson, S. *et al.* Ribavirin for treating Crimean Congo haemorrhagic fever. *Cochrane Database of Systematic Reviews* vol. 2018 (2018).
  182. Oestereich, L. *et al.* Evaluation of antiviral efficacy of ribavirin, arbidol, and T-705 (favipiravir) in a mouse model for Crimean-Congo hemorrhagic fever. *PLoS Negl. Trop. Dis.* **8**, (2014).
  183. Bertolotti-Ciarlet, A. *et al.* Cellular Localization and Antigenic Characterization of Crimean-Congo Hemorrhagic Fever Virus Glycoproteins. *J. Virol.* **79**, 6152 (2005).
  184. Zivcec, M. *et al.* Identification of broadly neutralizing monoclonal antibodies against Crimean-Congo hemorrhagic fever virus. *Antiviral Res.* **146**, 112 (2017).
  185. World Health Organisation. Prioritizing diseases for research and development in emergency contexts. <https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts> (2021).
  186. Tang, A. A. S., Tiede, C., Hughes, D. J., McPherson, M. & Tomlinson, D. C. Isolation of isoform-specific binding proteins (Affimers) by phage display using negative selection. *Sci. Signal.* **10**, 14 (2017).
  187. Miller, N. L., Clark, T., Raman, R. & Sasisekharan, R. Learned features of antibody-antigen binding affinity. *Front. Mol. Biosci.* **10**, (2023).
  188. Krammer, F. *et al.* Influenza. *Nat. Rev. Dis. Prim.* **4**, 1–21 (2018).
  189. Paget, J. *et al.* Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *J. Glob. Health* **9**, (2019).
  190. Spreeuwenberg, P., Kroneman, M. & Paget, J. Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic. *Am. J. Epidemiol.* **187**, 2561–2567 (2018).
  191. Rossman, J. S. & Lamb, R. A. Influenza Virus Assembly and Budding. *Virology* **411**, 229 (2011).
  192. Lampejo, T. Influenza and antiviral resistance: an overview. *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 1201 (2020).
  193. Treanor, J. J. *et al.* Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group. *JAMA* **283**, 1016–1024 (2000).
  194. Public Health England. *Surveillance and Laboratory Testing of Influenza Neuraminidase Inhibitor Resistance*. [www.facebook.com/PublicHealthEngland](http://www.facebook.com/PublicHealthEngland)

- (2018).
195. Uehara, T. *et al.* Treatment-Emergent Influenza Variant Viruses With Reduced Baloxavir Susceptibility: Impact on Clinical and Virologic Outcomes in Uncomplicated Influenza. *J. Infect. Dis.* **221**, 346–355 (2020).
  196. Van Dongen, M. J. P. *et al.* A small-molecule fusion inhibitor of influenza virus is orally active in mice. *Science (80-. )*. **363**, (2019).
  197. Zhu, L. *et al.* Inhibition of Influenza A Virus (H1N1) Fusion by Benzenesulfonamide Derivatives Targeting Viral Hemagglutinin. *PLoS One* **6**, (2011).
  198. Schrödinger LLC. Drug Discovery Platform. <https://www.schrodinger.com/platform/drug-discovery> (2023).
  199. Scientific, O. OpenEye Solutions. <https://www.eyesopen.com/solutions> (2023).
  200. Schrödinger, L. Schrödinger Release 2023-1: Maestro. (2023).
  201. Schrödinger LLC. Schrödinger Release 2023-1: Protein Preparation Wizard; Epik, Prime. (2023).
  202. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* **27**, 221–234 (2013).
  203. LLC, S. Schrödinger Release 2023-1: SiteMap. (2023).
  204. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).
  205. Schrödinger, L. Schrödinger Release 2023-1: LigPrep. (2023).
  206. OpenEye Scientific Software. OMEGA. (2013).
  207. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **50**, 572–584 (2010).
  208. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
  209. Schrödinger LLC. Schrödinger Release 2023-1: Glide. (2023).
  210. Friesner, R. A. *et al.* Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).
  211. How long does it take to screen 10,000 compounds with Glide? | Schrödinger. <https://www.schrodinger.com/kb/1012>.
  212. Hawkins, P. C. D., Skillman, A. G. & Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **50**, 74–82 (2007).
  213. OpenEye Cadence Molecular Sciences. ROCS 3.2.1.4. (2015).
  214. OpenEye Scientific Software. EON 2.2.0.5. (2013).
  215. Baell, J. & Walters, M. A. Chemical con artists foil drug discovery. *Nature* **513**, 481–483 (2014).
  216. Schrodinger LLC. The PyMOL Molecular Graphics System Version 2.3.2.
  217. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
  218. Rogers, D. J. & Tanimoto, T. T. A computer program for classifying plants.

- Science (80-. ).* **132**, 1115–1118 (1960).
219. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
  220. Kidmose, R. T. *et al.* Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526–531 (2019).
  221. Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
  222. Cianfrocco, M. A., Wong-Barnum, M., Youn, C., Wagner, R. & Leschziner, A. COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination. in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact* (Association for Computing Machinery, 2017). doi:10.1145/3093338.3093390.
  223. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174 (2015).
  224. Miles, J. A. *et al.* Selective Affimers Recognise the BCL-2 Family Proteins BCL-xL and MCL-1 through Noncanonical Structural Motifs. *Chembiochem* **22**, 232 (2021).
  225. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
  226. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**, 2403–2408 (1990).
  227. Sheu, S. Y., Yang, D. Y., Selzle, H. L. & Schlag, E. W. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12683–12687 (2003).
  228. eMolecules. No Title. <https://www.emolecules.com/> (2023).
  229. Carter, S. D. *et al.* Structure, Function, and Evolution of the Crimean-Congo Hemorrhagic Fever Virus Nucleocapsid Protein. *J. Virol.* **86**, 10914 (2012).
  230. Glaeser, R. M. & Han, B.-G. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys. Reports* **3**, 1 (2017).
  231. Noble, A. J. *et al.* Reducing effects of particle adsorption to the air-water interface in cryoEM. *Nat. Methods* **15**, 793 (2018).
  232. Wei, H. *et al.* Optimizing “Self-Wicking” Nanowire Grids. *J. Struct. Biol.* **202**, 170 (2018).
  233. ESPript. ESPript. <https://espript.ibcp.fr/ESPript/ESPript/index.php> (2023).
  234. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
  235. Klebl, D. P. *et al.* Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. *Structure* **28**, 1238 (2020).
  236. Garcia-Moro, E. *et al.* Reversible structural changes in the influenza hemagglutinin precursor at membrane fusion pH. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
  237. Morciano, G. *et al.* Use of luciferase probes to measure ATP in living cells and animals. *Nat. Protoc.* **12**, 1542–1562 (2017).
  238. Klimov, A. *et al.* Influenza virus titration, antigenic characterization, and serological Mmethods for antibody detection. *Methods Mol. Biol.* **865**, 25–51 (2012).

239. *Handbook of Surface Plasmon Resonance*. (Royal Society of Chemistry, 2017). doi:<https://doi.org/10.1039/9781788010283>.
240. Renaud, J. P. *et al.* Biophysics in drug discovery: impact, challenges and opportunities. *Nat. Rev. Drug Discov.* **15**, 679–698 (2016).
241. Cull, M. G. & Schatz, P. J. [26] Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol.* **326**, 430–440 (2000).
242. Gao, K., Oerlemans, R. & Groves, M. R. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85 (2020).
243. Dalvit, C. & Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **62**, 2218–2244 (2019).
244. Ekiert, D. C. *et al.* A Highly Conserved Neutralizing Epitope on Group 2 Influenza A Viruses. *Science (80- )*. **333**, 843 (2011).
245. Niebling, S. *et al.* FoldAffinity: binding affinities from nDSF experiments. *Sci. Rep.* **11**, 1–17 (2021).
246. Hu, Y., Stumpfe, D. & Bajorath, J. Recent Advances in Scaffold Hopping. *J. Med. Chem.* **60**, 1238–1246 (2017).
247. Enamine. HTS Collection. <https://enamine.net/compound-collections/screening-collection/hts-collection>.
248. Askin, S. *et al.* Selective protein unfolding: a universal mechanism of action for the development of irreversible inhibitors. *Chem. Commun.* **54**, 1738–1741 (2018).
249. Abad-Zapatero, C. & Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **10**, 464–469 (2005).
250. CCPEM. Doppio User Guide. [https://www.ccpem.ac.uk/docs/doppio/user\\_guide.html](https://www.ccpem.ac.uk/docs/doppio/user_guide.html).
251. Jamali, K., Kimanius, D. & Scheres, S. H. W. A Graph Neural Network Approach to Automated Model Building in Cryo-EM Maps. *arXiv* (2022).
252. Darrow, M. C., Moore, J. P., Walker, R. J., Doering, K. & King, R. S. Chameleon: Next Generation Sample Preparation for CryoEM based on Spotiton. *Microsc. Microanal.* **25**, 994–995 (2019).
253. Ravelli, R. B. G. *et al.* Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nat. Commun.* **2020 111** **11**, 1–9 (2020).
254. Higgins, A. J. *et al.* Cycloalkane-modified amphiphilic polymers provide direct extraction of membrane proteins for CryoEM analysis. *Commun. Biol.* **2021 41** **4**, 1–9 (2021).
255. Parmar, M. *et al.* Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim. Biophys. Acta - Biomembr.* **1860**, 378–383 (2018).
256. Dunstone, M. A. & de Marco, A. Cryo-electron tomography: an ideal method to study membrane-associated proteins. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).
257. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **2019 1611** **16**, 1146–1152 (2019).
258. Zivanov, J. *et al.* A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11**, (2022).
259. Hutchings, J. *et al.* Structure of the complete, membrane-assembled COPII coat reveals a complex interaction network. *Nat. Commun.* **12**, (2021).

260. von Kügelgen, A. *et al.* In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell* **180**, 348-358.e15 (2020).
261. Turoňová, B. *et al.* In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203 (2020).
262. Xue, L. *et al.* Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205 (2022).
263. Gold, V. A. M. *et al.* Visualizing active membrane protein complexes by electron cryotomography. *Nat. Commun.* **2014 51** **5**, 1–9 (2014).
264. Silvester, E. *et al.* DNA origami signposts for identifying proteins on cell membranes by electron cryotomography. *Cell* **184**, 1110-1121.e16 (2021).
265. Amunts, A. *et al.* Structure of the Yeast Mitochondrial Large Ribosomal Subunit. *Science (80-. )*. **343**, 1485–1489 (2014).
266. REAL Space - Enamine. <https://enamine.net/compound-collections/real-compounds/real-space-navigator>.
267. FastROCS Shape based virtual screening software. <https://www.eyesopen.com/fastrocs>.
268. Grebner, C. *et al.* Virtual Screening in the Cloud: How Big Is Big Enough? *J. Chem. Inf. Model.* (2019) doi:10.1021/acs.jcim.9b00779.
269. Petrović, D. *et al.* Virtual Screening in the Cloud Identifies Potent and Selective ROS1 Kinase Inhibitors. *J. Chem. Inf. Model.* **62**, 3832–3843 (2022).
270. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-. )*. **367**, 1260–1263 (2020).
271. Jovčevska, I. & Muyldermans, S. The Therapeutic Potential of Nanobodies. *Biodrugs* **34**, 11 (2020).
272. Huo, J. *et al.* Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat. Struct. Mol. Biol.* **2020 279** **27**, 846–854 (2020).
273. Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPins: A new generation of protein therapeutics. *Drug Discov. Today* **13**, 695–701 (2008).