Structural and Biophysical Studies of HIV Rev and HBV e-Antigen

Michael A. DiMattia

Nuffield Department of Clinical Medicine
Magdalen College
University of Oxford

Laboratory of Structural Biology Research
National Institute of Arthritis and Musculoskeletal and Skin Diseases

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2012

Structural and Biophysical Studies of HIV Rev and HBV e-Antigen

Michael A. DiMattia, Magdalen College, University of Oxford
D.Phil, Nuffield Department of Clinical Medicine
Michaelmas 2012

Human immunodeficiency virus (HIV) Rev and Hepatitis B virus (HBV) e-antigen are both viral proteins that have key functions in their respective viral replication cycles. Both have evaded crystallisation for decades due to their tendency to aggregate and/or form higher-order species. In this thesis the structure determination of HIV Rev and HBV e-antigen is presented—achieved via complexing with monoclonal antibody Fab fragments—and their structures are analysed.

HIV Rev is a small regulatory protein that mediates the nuclear export of viral mRNAs, an essential step in the HIV replication cycle. In this process, Rev cooperatively oligomerises onto a highly structured RNA motif, the Rev response element. The structure of Rev (complexed with Fab), determined to 2.3 Å resolution, reveals a molecular dimer where the ordered portion of each subunit (N-terminal domain; NTD; residues 9-65) contains two coplanar α-helices arranged in hairpin fashion. Rev subunits dimerise via interaction of identical hydrophobic patches that overlap to form a V-shaped assembly. Mating of hydrophobic patches on the outer surface of the dimer promotes higher order interactions. Cryo-electron microscopy and helical image reconstruction of *in vitro* assembled Rev filaments were performed to better understand higher-order Rev oligomerisation. Reconstructions of Rev filaments were determined to ~13 Å resolution, permitting docking of the Rev NTD structure. Conformational variability of the Rev dimer subunits and use of a third oligomerisation interface engender filaments that can expand and contract. Both characteristics were also observed in the crystal structures of Rev. Surface features of the Rev filaments are altered in different expansion states, which may have implications for the assembled forms that Rev adopts during nuclear export of RNA and subsequent re-import into the nucleus. Various models for Rev oligomerisation onto the viral RNA are proposed.

Chronic Hepatitis B virus (HBV) infection afflicts millions worldwide with cirrhosis and liver cancer. HBV e-antigen (HBeAg), a clinical marker for disease severity, is a soluble variant of the protein (core antigen, HBcAg) that forms the building-blocks of capsids. HBeAg is not required for virion production, but is implicated in establishing immune tolerance and chronic infection. The crystal structure of HBeAg clarifies how the short N-terminal propeptide of HBeAg induces a radically altered mode of dimerisation relative to HBcAg (~140° rotation), which is locked into place through formation of intramolecular disulfide bridges. This structural switch precludes capsid assembly and engenders a distinct antigenic repertoire, explaining why the two antigens are cross-reactive at the T-cell level (through sequence identity) but not at the B-cell level (through conformation). The structure offers insight into how HBeAg may establish immune tolerance for HBcAg while evading its robust immunogenicity.

I would like to dedicate this thesis to my parents
Michael and Angela DiMattia

Acknowledgements

First I would like to thank Dr. Jonathan Grimes, Prof. David Stuart, and Dr. Alasdair Steven for the opportunity to carry out this exciting research under their supervision. Thank you for your dedicated guidance and support: scientifically, professionally, and personally. Thank you to Dr. Paul Wingfield for a great collaboration. I thank specifically Dr. Norman Watts for all his input and insight discussions. A most humble thank you is extended to Dr. Naoko Mizuno for all of her fast-paced training in cryo-EM analysis before leaving for her new post in Munich.

Many people at the Division of Structural Biology (Oxford) and the Laboratory of Structural Biology (NIH) have contributed to this work. Thank you to all my friends at STRUBI for creating such a wonderful place to do research. Special thanks to Dan Whalen, Dr. Robert Gilbert, Dr. Erika Mancini, Jenny Hayward, Eli Fulcini, Dr. Doryen Bubeck, Margaret Jones, Dr. Dennis Winkler, Dr. Paul Keller, Dr. Juan Fontana, Dr. Rick Huang, Phil West, and Dr. Bernard Heymann. Your friendship and guidance has made all the difference. A specific thank you to Dr. Stephen Graham, Dr. Karl Harlos, Dr. Tom Walter and Jun Dong for their help with structure refinement, crystal handling, and computing. I graciously thank Dr. Cedric Leyrat for help with complementary biophysical experiments. Thank you to Nea Dam for administrative support. I also thank my viva examiners, Dr. Kay Grünewald and Prof. Felix Rey for their rigorous examination and thoughtful discussion regarding my D.Phil research.

I would like to say a heartfelt thank you to my dear friends Ellene Mashalidis, Joe Murphy, Leigh Shaprio, Trish Ryan, Ian Peahota, Nick Otto, Kovia Gratzon-Erskine, and Sarah Clarke for their wonderful friendship and support, even through the hard times. An unparalleled thank you to Hania Pavlou, Megan Watts, Elizabeth Humberstone, Erin Garrigan, my parents, sisters, and grandpa for their unwavering love and support. I would not have been able to get through this without you.

Declaration of Work

The work described herein is my own except as described below. This work is part of a collaborative project with Dr. Paul Wingfield of the Protein Expression Laboratory (PEL) and Dr. Alasdair Steven of the Laboratory of Structural Biology Research (LSBR) (NIAMS, NIH).

Chapter 2: The isolation of anti-Rev Fab fragments by phage display was performed by Christoph Rader (NCI, NIAMS) and Paul Wingfield (PEL). The preparation and characterisation of the Rev-Fab and Rev-scFv complexes were carried out by Stephen Stahl, Norman Watts, Ira Palmer, and Josh Kaufman as part the same collaborative project with PEL and Alasdair Steven LSBR. All negative-stain electron micrographs of Rev, RevL64A, and HBeAg were recorded by Norman Watts. The Fab-HBeAg complex was jointly prepared by myself and Norman Watts.

Chapter 3: AUC experiments on Rev L64A were carried out by me under the guidance of Robert Gilbert (STRUBI). Negative stain EM on Rev L64A was performed by Norman Watts (PEL).

Chapter 4: Cryo-electron microscopy data collection on Rev filaments was performed by Naiqian Cheng (LSBR).

Chapter 5: Analytical ultracentrifugation (AUC) on HBeAg samples was carried out by Paul Wingfield (PEL).

Table of Contents

List of Figures

List of Tables

Abbreviations

E.coli	*Escherichia coli*
AIDS	Acquired Immunodeficiency Syndrome
ARM	arginine-rich RNA binding motif
ASU	asymmetric unit
AUC	analytical ultracentrifugation
BSA	buried surface area
CA	HIV capsid protein
CCD	charge-coupled device
Crm-1	exportin-1
CTD	C-terminal domain
CTF	contrast transfer function
DAP	diammonium phosphate
DLS	dynamic light scattering
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	etyhlenediaminetetraacetic acid
eIF-5A	eukaryotic initiation factor 5A
EM	electron microscopy
ESRF	European Synchrotron Radiation Facility
Fab	antibody fragment, antigen binding
FSC	Fourier shell correlation
GDP	guanosine diphosphate
GTP	guanosine triphosphate
HAART	Highly-active anti-retroviral therapy
HBcAg	HBV core antigen
HBeAg	HBV e antigen
HBV	Hepatitis B virus
HIV	Human Immunodeficiency Virus
HIV	Human Immunodeficiency Virus
hRIP	human Rev-interacting protein
Ig	immunoglobulin
IHRSR	iterative helical real space reconstruction
IN	HIV integrase
LTR	Long Terminal Repeat
MA	HIV matrix protein
MD	molecular dynamics

MHC Major Histocompatibility Complex
miRNA micro RNA
MR molecular replacement
mRNA messenger RNA
NC HIV nucleocapsid protein
NCS non-crystallographic symmetry
NES nuclear export sequence
NLS nuclear localisation signal
NMR Nuclear magnetic resonance
NPC Nuclear Pore Complex
NTD N-terminal domain
Nup nucleoporin
OLIGO oligomerisation motif
PCA principal component analysis
PDB Protein Data Bank
PIC HIV pre-integration complex
PR HIV protease
PTHrP parathyroid hormone-related protein
RanGAP Ran-GTPase-activating protein
RanGEF Ran-GDP-exchange factor
RMSD root mean square deviation
RNA ribonucleic acid
RNP ribonucleoprotein
RRE Rev Response Element
rRNA ribosomal RNA
RT HIV reverse transcriptase
SAM68 Src-associated protein (68 kDa)
SAXS small angle X-ray scattering
scFv antibody fragment, single-chain variable domain
SDS PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLIIB Stem Loop IIB
snRNA small nuclear RNA
SPR surface plasmon resonance
SREBP-2 sterol regulatory element-binding protein 2
ssRNA single-stranded RNA
SV spatial variance

TLS translation-libration-screw motion
tRNA transfer RNA
WT wild-type

Chapter 1

Introduction

1.1 Human immunodeficiency virus (HIV)

1.1.1 HIV replication and regulatory protein Rev

Human immunodeficiency virus (HIV) is a retrovirus that can infect humans and cause acquired immunodeficiency syndrome (AIDS). AIDS is a disease resulting in the progressive destruction of the immune system, primarily via the depletion of $CD4^+$ helper T cells (Miedema et al., 1990). Loss of cell-mediated immunity occurs once the count of $CD4^+$ T cells falls below a critical threshold, leaving the body increasingly susceptible to life-threatening, opportunistic infections (Hazenberg et al., 2003). When HIV was first isolated in 1983, the average time from diagnosis to death was approximately four years. In the three decades since, the development of a number of antiviral drugs has led to a dramatic increase in longevity for HIV-positive individuals with access to adequate health care (Carpenter et al., 1996; Gilks et al., 2006). The predominant treatment for HIV infection is highly-active anti-retroviral therapy (HAART), which is the administration of several antiviral drugs in a temporally-staggered manner to inhibit HIV replication for as long as possible (Grabar et al., 2000). Development of new antiviral drugs may expand our clinical repertoire for treatment of HIV-positive individuals (Padian et al., 2011; Shattock et al., 2011)

The majority of the drugs developed against HIV are nucleoside and non-nucleoside inhibitors of

the viral enzymes (reverse transcriptase, protease, and integrase) (Pomerantz and Horn, 2003). In recent years, fusion inhibitors have also been developed to block cellular entry of the virion. Arresting the HIV replication cycle at any step is, in theory, a potential avenue for reducing the pathogenicity of AIDS. However, great potential for improvement remains with regard to the efficacy of current anti-retroviral therapy. Obstacles that impede complete suppression of the virus in infected individuals include the development of resistance to antivirals, latency of the virus in tissue-specific reservoirs, and metabolic complications and drug toxicity effects (Pomerantz and Horn, 2003).

The first part of this dissertation is focused on the structural study of HIV Rev, which is a regulatory viral protein that regulates the nuclear export of HIV mRNA transcripts, an essential step of the HIV replication cycle. The HIV regulatory and accessory proteins, including Rev, have not yet been targeted for drug design, and development of antivirals against these proteins has great potential for additional avenues to inhibition of viral replication.

1.1.1.1 Virion genome and structure

The HIV genome comprises 9 genes that code for 15 proteins. From the 5'- to 3'-end of the genome, the three largest genes are *gag*, *pol*, and *env* (**Fig 1.1A**) (Votteler and Schubert, 2008). The *gag* gene encodes the main structural proteins: matrix (MA), capsid (CA), nucleocapsid (NC), and p6. These four proteins are initially expressed as one polyprotein precursor that is ultimately cleaved by the viral protease. The *pol* gene is expressed as a Gag-Pol polyprotein precursor that results from a ribosomal frameshifting event. Once processed, the *pol*-encoded proteins comprise the three viral enzymes: protease (PR), reverse transcriptase (RT), and integrase (IN) (Votteler and Schubert, 2008). The *env* gene produces the glycoprotein spikes that stud the lipid bilayer: gp120 (surface) and gp41 (transmembrane).

HIV has developed a number of strategies to express all of its viral proteins in a limited amount

of genomic space (Feinberg and Greene, 1992). As depicted in **Fig 1.1A**, genes are coded for in all three reading frames. Furthermore, expression of *pol* results from a rare frameshifting event during translation of the Gag polyprotein (Freed, 2001). The genome also contains many splice sites, which result in more than 30 differentially spliced mRNAs (this will be explored in greater detail in the next section).

In addition to *gag*, *pol*, and *env*, HIV-1 also codes for a number of regulatory and accessory proteins. Tat (expressed from *tat* gene) is essential for transcription of the viral genome from the HIV-1 long terminal repeat (LTR) (**Fig 1.1A**), and Rev (expressed from *rev*), also essential for HIV replication, mediates the transport of viral mRNAs from the nucleus to the cytoplasm (Votteler and Schubert, 2008). There are also a number of smaller genes that encode the auxiliary proteins Vpu, Vif, Vpr and Nef. These proteins have important functions for increasing the efficacy of HIV infectivity and pathogenesis; however, they are not essential for viral replication (for full details on their known functions, see (Freed, 2001) or Li et al. (2005)).

The structural proteins form the fundamental framework of the HIV virion (shown in shades of blue and red in **Fig 1.1B,C**), which is roughly spherical with a diameter of ~120 nm (range: 110-128 nm (Gentile et al., 1994). Within the mature virion is the viral genome, which is composed of two copies of positive-sense single-stranded RNA that are each ~9 kb in length (Freed, 2001). Copies of NC are tightly bound to the ssRNA genome. Surrounding the genome is a conical protein shell composed of CA. Reverse transcriptase and IN are also localised within the capsid (green), tightly bound to the viral genome (**Fig 1.1C**). Exterior to the capsid shell is a matrix protein layer that organises underneath the host-derived lipid bilayer. The gp120 and gp41 viral glycoproteins are used for host cell recognition and fusion (orange and yellow, respectively) (**Fig 1.1C**). The matrix layer helps to anchor the gp41/gp120 glycoprotein spikes in the lipid bilayer by interacting with gp41.

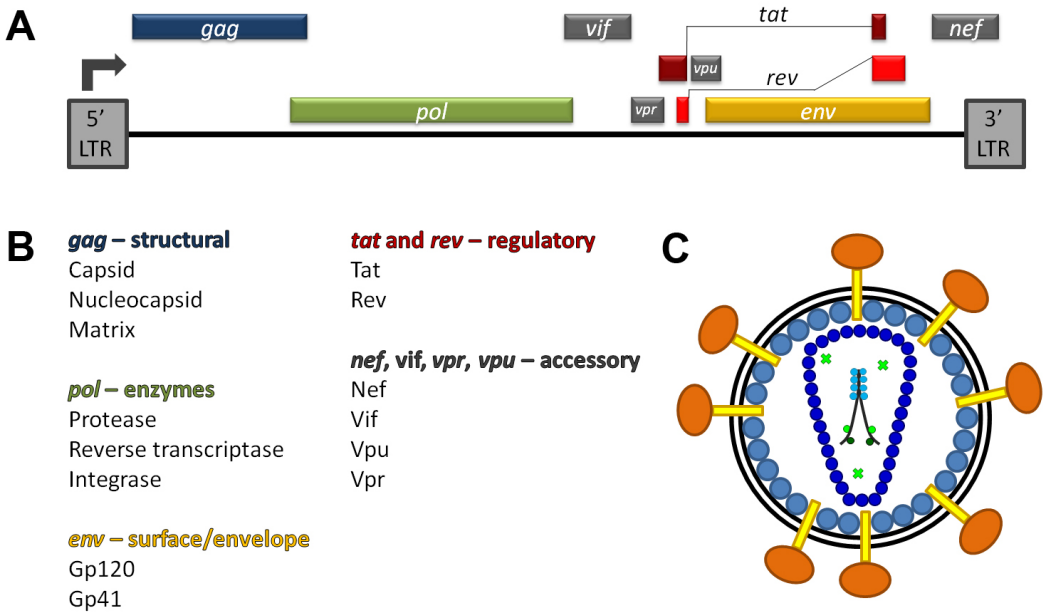


Figure 1.1: HIV-1 Genome Organisation and Virion Structure. **(A)** The HIV-1 genome. From left to right, the genome is depicted from 5'-long terminal repeat (LTR) to 3'-LTR. The three main viral genes (*gag*, *pol*, and *env* are shown in blue, green, and yellow, respectively. These genes code for the viral structural, enzymatic and surface proteins, respectively. *tat* and *rev* encode the regulatory proteins Tat and Rev (shades of red). Accessory genes (*vif*, *nef*, *vpr*, *vpu*) are shown in shades of gray. **(B)** The mature HIV-1 virion. The mature HIV-1 virion is roughly spherical, with an outer shell composed of a host-derived lipid bilayer (double black) that is supported by a matrix protein layer (large blue). Anchored within these layers are the surface glycoproteins, gp41 (yellow; transmembrane) and gp120 (orange; surface). The virion core comprises a conical shell of capsid protein (small, dark blue), which contains two copies of the viral ssRNA (black) bound to nucleocapsid (NC; small, light blue) and many of the other viral proteins, particularly the enzymes (green; protease, integrase, and reverse transcriptase).

1.1.1.2 HIV replication cycle

HIV replication has been studied in depth for decades and a full review of the molecular details is beyond the scope of this introduction. However, a general view of the events that take place during HIV replication is presented in **Fig 1.2** and described below.

HIV infection begins with attachment of the virion to the host cell surface. The viral glycoprotein gp120 binds to CD4, the primary cell surface receptor present on macrophages, some T cell lineages, and other types of immune cells. Interaction of gp120 and CD4 induces conformational changes in gp120 that increase its affinity for HIV's host cell co-receptor: either CCR5 or CXCR4 (Klasse, 2012). Co-receptor engagement induces conformational changes in gp120 that prime the associated gp41 to catalyse membrane fusion, which takes place between the lipid bilayer of the viral envelope and the host cell plasma membrane (Klasse, 2012). The fusion event allows the viral core—defined as the structure that remains after the lipid bilayer and most of MA is removed—to gain access to the cytoplasm.

Subsequent to fusion, the viral core is converted to the preintegration complex (PIC), wherein the genome container is opened and the viral RT catalyses the conversion of the RNA genome into double-stranded DNA (Piller et al., 2003). NC, some MA, the accessory protein Vpr, and viral enzymes RT and IN remain associated as the PIC translocates to and into the nucleus. In the nucleus, IN catalyses the integration of the HIV DNA into the host cell genome (Krishnan and Engelman, 2012). Following integration into the host chromosome, the integrated provirus becomes the template from which all of the viral proteins are transcribed. Tat is a transactivator of HIV transcription, acting on an RNA element in the 5'-LTR to upregulate RNA synthesis (by more than two logs) by recruiting various host cell transcription factors to the integrated viral template (Freed, 2001).

The HIV genome contains a number of splice sites that, upon transcription, give rise to more than

30 viral mRNAs, many of which contain introns. Intronic RNA is normally not permitted to leave the nucleus. HIV has overcome this problem through the evolution of a novel viral protein, Rev ("regulator of expression of viral proteins") and a *cis*-acting RNA element, the Rev response element (RRE) (Pollard and Malim, 1998). Because of its role in regulating the expression of viral proteins, Rev is considered to govern the switch between "early" and "late" viral replication (Kim et al., 1989). Synthesis of all of the viral proteins, including the structural proteins and enzymes, marks the beginning of late HIV replication. Nascent virion assembly takes place at the inner surface of the host cell plasma membrane. Further details of known Rev function, and mechanism(s) by which such functions occur, are considered in the subsequent sections.

The Gag polyprotein is targeted to the plasma membrane, where it interacts with the membrane itself and also multimerises (Bell and Lever, 2013). A *cis*-acting element in the RNA termed the packaging signal (Ψ) targets the genome to Gag, where two copies of the newly-synthesised genome are encapsidated (Lu et al., 2011). Gag interacts with Env glycoproteins in the plasma membrane (which arrived via the endoplasmic reticulum) and then—aided by the host cell ESCRT machinery—stimulates the necessary curvature and eventual budding from the host cell, wherein the host-derived lipid bilayer is acquired (Lee et al., 2012). During or just subsequent to virion release from the plasma membrane, the viral PR cleaves the Gag- and Gag-Pol polyproteins, inducing structural rearrangement of the viral proteins within the maturing virion. The steps of HIV replication for which antiviral drugs have been developed are marked with a red "X" in **Fig 1.2**.

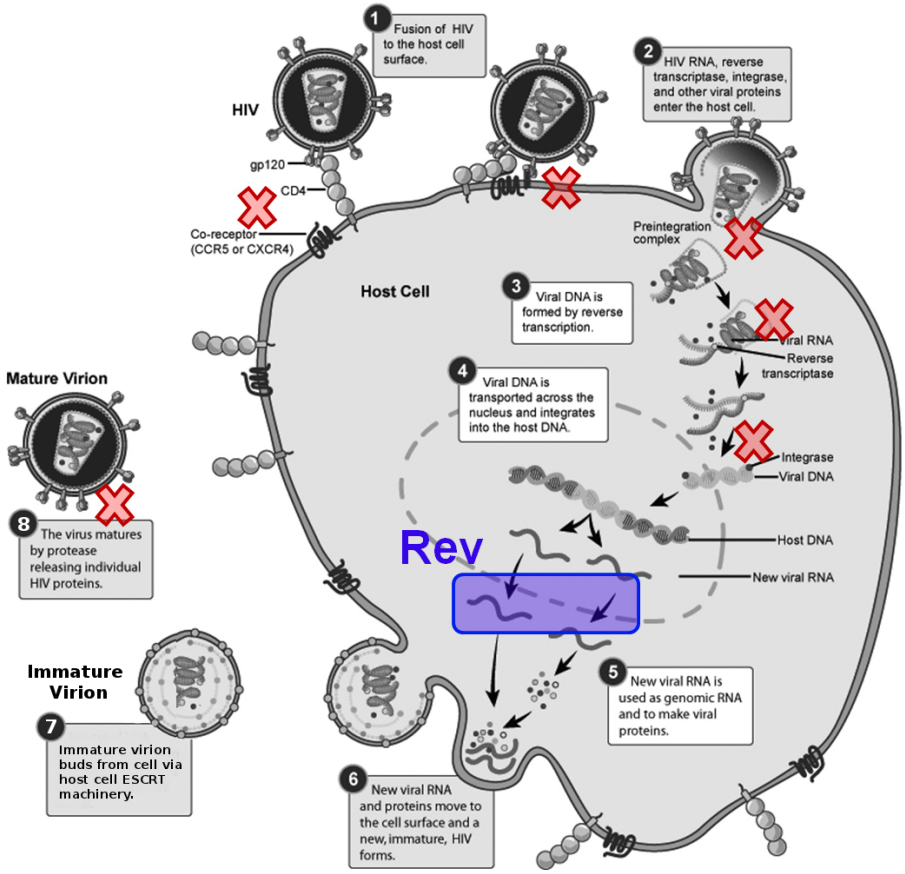


Figure 1.2: Replication cycle of HIV-1.
(1) Attachment of the HIV-1 virion to cell via interaction of gp120 with host cell proteins CD4 (receptor) and either CCR5 or CXCR4 (co-receptor).
(1) Fusion of viral lipid bilayer with host cell plasma membrane via HIV gp41 (and gp120).
(2) Uncoating of virion. The viral core, leaving behind the lipid bilayer, translocates to nucleus.
(3) Reverse transcription. Viral RT catalyses reverse transcription of viral RNA genome into dsDNA
(4) Integration. Viral DNA is transported across the nucleus and IN integrates into the host chromosome.
(4) Tat upregulates transcription of viral genome.
(blue box) Rev regulates the export of viral mRNAs.
(5) Exported viral mRNAs are used as genomes for nascent virions and for expression of viral proteins.
(6) Assembly of immature virions at inner surface of plasma membrane.
(7) Viral PR cleaves Gag- and Gag-Pol polyproteins, inducing virus maturation.
The steps of HIV replication for which antiviral drugs have been developed are marked with a red 'X'. Figure modified from National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD USA

1.1.1.3 HIV's nuclear export problem and Rev

HIV, like most retroviruses, has a small RNA genome that once reverse transcribed into DNA, is integrated into the host genome as a single provirus. Reducing integration to a single event comes at a cost for a virus that has 9 genes that encode 15 proteins (**Fig 1.3A**). One strategy HIV has evolved to deal with this issue is a high degree of post-transcriptional splicing. While transcription initiates from the 5'-LTR, multiple splice sites give rise to a complex pattern of downstream RNA processing whereby more than 30 different mRNA transcripts are alternatively spliced (**Fig 1.3B**). Broadly, these transcripts fall into three size classes (Sodroski et al., 1986; Emerman et al., 1989; Malim et al., 1989b) (**Fig 1.3C**):

1. 9-kb unspliced (full length) transcripts, used as genomic RNA for packaging into assembling virions and for translation of many of the viral proteins;
2. 4-kb partially spliced (single splicing) transcripts; used for translation of the viral structural proteins and enzymes, along with full-length transcripts;
3. and 2-kb fully spliced transcripts; translation of regulatory proteins Rev, Tat and Nef.

However, a problem arises with this post-transcriptional scheme: the unspliced and partially spliced transcripts contain introns, and therefore, the transcripts are typically retained in the nucleus until they can be fully spliced by the host cell's splicing machinery. This default host process is incompatible with HIV replication due to the requirement of the longer transcripts for packaging into nascent virions (as RNA genomes) and for expression of the major viral proteins (structural and enzymatic) (Tiley et al., 1992). HIV circumvents this issue with Rev, a regulatory viral protein that shuttles the longer transcripts (9-kb and 4-kb) from the nucleus to the cytoplasm, prior to complete splicing (Malim et al., 1989b; Pollard and Malim, 1998). Because Rev function is correlated with the expression of the structural viral proteins required in late stages of the replication cycle, it is thought to regulate the transition from

early to late phase HIV infection (Pollard and Malim, 1998).

1.1.1.4 Domain organisation of Rev

With a molecular mass of only 13 kDa, Rev is a relatively small protein with a high affinity for RNA (Heaphy et al., 1991; Wingfield et al., 1991; Zemmel et al., 1996). Despite its diminutive size, Rev contains multiple motifs that confer it with the capacity to shuttle RNA out of the nucleus (Malim et al., 1989a). The Rev N-terminal domain (NTD) contains an arginine-rich RNA binding motif (ARM) that is used for targeting viral mRNAs (**Fig 1.4**) (Cochrane et al., 1990; Kjems et al., 1992). The ARM is flanked by two oligomerisation (OLIGO) motifs that are utilised for Rev multimerisation (Zapp et al., 1991; Hope et al., 1992), an essential process in Rev-mediated nuclear export of RNA. Its C-terminal domain (CTD) contains a nuclear export sequence (NES) that engages host cell protein Crm-1 (exportin-1) to effect nuclear export of the entire ribonucleoprotein (RNP) complex through the nuclear pore complex (NPC)(Malim et al., 1989a; Fischer et al., 1995). Once the RNA is released into the cytoplasm, Rev can be imported back into the nucleus through the direct interaction of its nuclear localisation signal (NLS) with importin-β (Szebeni et al., 1997; Henderson and Percipalle, 1997). Interestingly, the ARM and NLS motifs utilise the same stretch of residues, hence their coincident labeling in **Fig 1.4**. Subsequent to nuclear import, Rev can engage new viral RNA for another round of nuclear export.

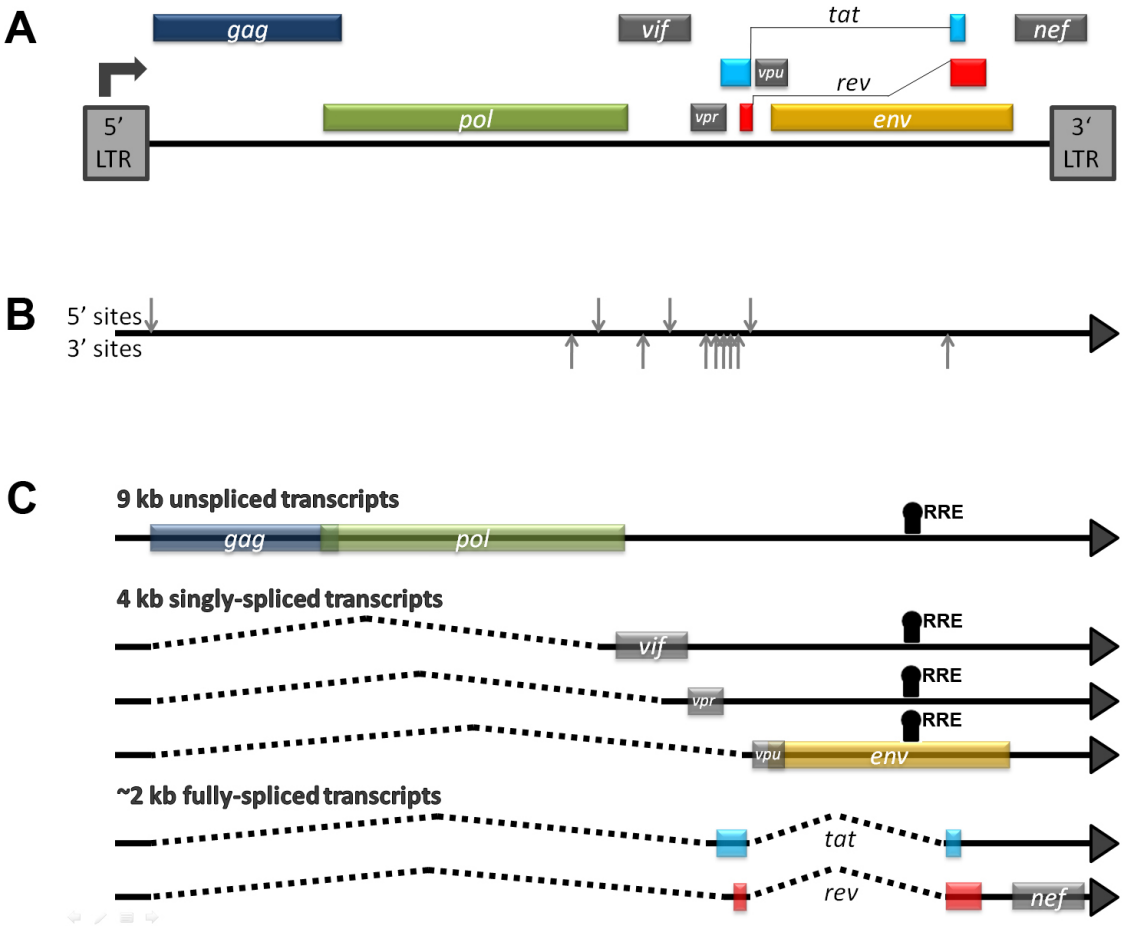


Figure 1.3: The RNA Nuclear Export Problem. Depicted at top is the HIV-1 genome, comprising nine genes in a single ORF. Many 5' and 3' splice sites (grey arrows) result in the production of 30+ different mRNA viral transcripts, clustered into three classes based on degree of splicing (and for ease of communication, also on size: 9-kb, 4-kb, and ~2-kb). The nuclear export of the 9-kb and 4-kb transcripts, which code for the majority of viral proteins, is regulated by Rev (in red).

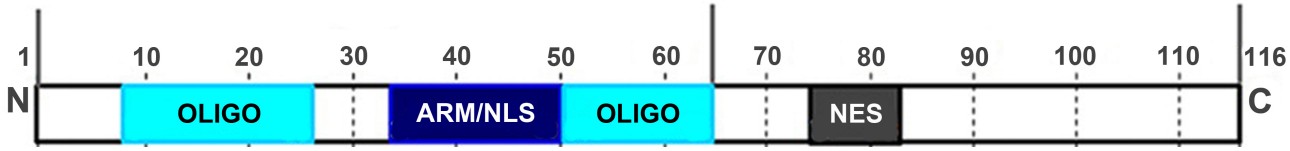


Figure 1.4: Domain organisation of HIV-1 Rev. The Rev polypeptide is 116 amino acids in length and can be divided into two functional domains, the N-terminal domain (NTD) and the C-terminal domain (CTD). The NTD comprises two stretches of amino acids (residues 9-26 and 51-65) that are involved in Rev self-association. These oligomerisation motifs (in cyan) flank an arginine-rich RNA-binding motif (ARM; in navy blue) that is responsible for interaction with HIV viral RNAs. Coincident with the ARM is a nuclear localisation sequence (NLS). The CTD contains a nuclear export sequence (NES; in gray).

1.1.2 Nucleocytoplasmic Transport in Eukaryotes

HIV Rev modulates the nuclear export of viral RNAs by adapting itself for use in host cell nuclear export and import pathways. This section briefly reviews how eukaryotic cells transport macromolecules to and from the nucleus, which will provide context for better understanding how Rev appropriates access to these host cell processes for the benefit of HIV replication.

The central dogma of molecular biology—DNA codes for RNA which codes for proteins—is upheld in all forms of cellular life. In prokaryotes, the genome is transcribed and translated in the same physical space. The appearance of a nuclear compartment in eukaryotes, separate from the cytoplasm, marked a major evolutionary transition. In it, the flow of genetic information encounters a physical barrier in the form of the nuclear envelope: a two-layered membrane that prevents free diffusion of macromolecules between the nucleus and cytoplasm (Hoelz et al., 2011). The spatial segregation of transcription (in the nucleus) and translation (in the cytoplasm) necessitated the development of molecular machineries, known as the nuclear pore complexes (NPC), which allow the transport of many different molecules (RNA, proteins, ribonucloprotein (RNP) complexes) between the nucleus and cytoplasm (Nakielny and Dreyfuss, 1999). Recent research has elucidated strategies that certain virus families (in particular parvoviruses and polyomaviruses) which can bypass the NPC and pass directly through the nuclear membrane (Kobiler et al., 2012).

1.1.2.1 Nuclear Pore Complex

Central to the eukaryotic transport machinery are the nuclear pore complexes (NPCs), which are large (120 nm diameter; 9 nm 'opening' diameter) hollow, highly-specialised cylinders embedded in the nuclear envelope that selectively allow passage of macromolecules carrying specific import and export

signals. The NPC comprises more than 30 different proteins, termed nucleoporins (Nups), that assemble with eight-fold symmetry into several major domains (**Fig 1.5**): the selective central channel, core scaffold, transmembrane regions, nuclear basket, and cytoplasmic filaments (Davis and Blobel, 1986; Rout et al., 2000; Cronshaw et al., 2002; DeGrasse et al., 2009). Within the central channel are Nups that form largely unstructured filaments rich in phenylalanine and glycine, termed FG Nups. The core scafford and transmembrane regions support the central channel and stabilise the NPC, respectively. The nuclear basket and cytoplasmic filaments reach ~100 nm into the nucleus and cytoplasm, respectively, and are thought to be involved in guiding macromolecules to the central channel for transport.

1.1.2.2 Selectivity of Transport

Passive transport through the NPC is size-restricted; macromolecules below a mass of ~60 kDa (as well as water and metabolites) can diffuse across the central channel (Mattaj and Englmeier, 1998; Pemberton and Paschal, 2005). However, the majority of macromolecules that are imported into the nucleus or exported to the cytoplasm contain either a nuclear localisation sequence (NLS) or nuclear export sequence (NES), respectively. These transport signals are recognised by transport receptors. The karyopherin families, comprising the various importin and exportin proteins, are often the transport receptors. These proteins directly engage the NLSs or NESs of their cognate cargoes, and to the FG Nups of the central channel, serving as adaptor proteins that mediate selective transport across the NPC. Of the many NPC domains, the central channel and the FG Nups are sufficient to mediate selective receptor-mediated transport (Strawn et al., 2004; Jovanovic-Talisman et al., 2009).

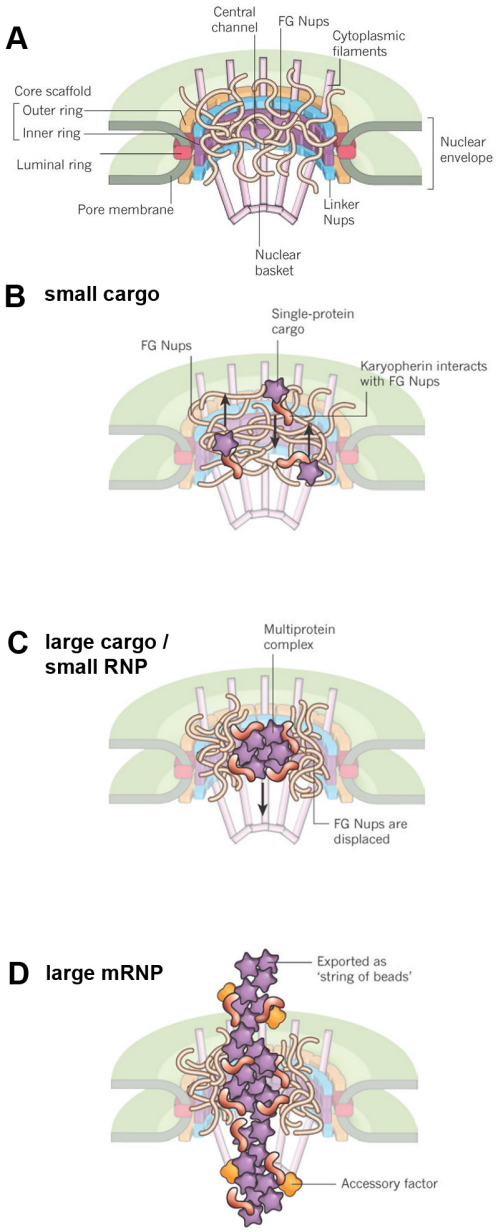


Figure 1.5: The Nuclear Pore Complex (A) Schematic representation of the nuclear pore complex (NPC). The major structure features are identified. The central channel is the space through which cargoes are transported. An outer and inner ring of nucleoporins (Nups) comprise the fundamental structure of the NPC, situating it within the nuclear envelope. Unstructured Fg Nups localize within the central channel, which prevent access to macromolecules larger than 60 kDa, but mediate transport of cargoes bound to karyopherins (exportins and importins). These karyopherins bind to FG Nups directly. **(B-D)** Grossly different sized cargoes are transported through the NPC. Small cargoes containing a protein and its respective karyopherin are transported without displacement of the FG Nups. Large cargoes and RNPs contain multiple transport factors, which must displace the FG Nups. Large mRNPs are translocated as a 'string of beads' in which each 'bead' behaves as a large cargo. It remains unclear which method is used for transport of Rev-RRE complexes, but panels (**C**) and (**D**) are most plausible. Figure modified from (Grunwald et al., 2011).

1.1.2.3 Directionality of Transport

Macromolecules undergoing transport through the NPC move by diffusion and therefore change their direction stochastically. Consequently, an external system that imposes directionality on transport is required. The directionality of karyopherin-mediated transport is modulated by a small GTPase called Ran, which exists in different states on either side of the nuclear envelope (Moore and Blobel, 1993). Interaction of Ran with karyopherins permits release from the NPC on only one side of the nuclear envelope, depending on the GDP/GTP state of Ran itself.

In the nucleus, Ran is GTP-bound and in the cytoplasm, Ran exists in a GDP-bound state. This RanGTP–RanGDP gradient is maintained across the nuclear membrane with the aid of two regulatory proteins: RanGEF (Ran-GDP-exchange factor) in the nucleus and RanGAP (Ran-GTPase-activating protein) in the cytoplasm. In general, import receptors bind cargo in the cytoplasm and release it in the nucleus upon RanGTP binding (Moroianu et al., 1996). Export receptors bind cargo in the nucleus in complex with RanGTP and dissociation of cargo is triggered upon GTP hydrolysis. Not all transport receptors require Ran, nor belong to the karyopherin family. However, notably, all transport receptors can interact directly with FG Nups (22).

1.1.2.4 Nuclear export of RNA

The NPC is used for export of many different macromolecules; however, it is probable that the majority of export cargoes are RNAs, usually complexed with proteins. The most abundant such complexes are messenger RNPs: ribosomal subunits and mRNAs. Such mRNP cargoes can be as large as 100 MDa (Mehlin et al., 1992). These large cargoes are transported through the NPC alongside relatively small cargoes, comprising just one protein and its corresponding karyopherin transporters (**Fig 1.5B-D**). How

these large mRNP complexes, which can be up to 50 nm in diameter, are transported through the NPC remains unclear, but it is known that the quaternary structure of very large RNPs must be remodelled (i.e. unraveled at the nucleoplasmic face of the NPC and threaded through as a thin strand) (Grunwald et al., 2011). Despite the need for remodelling of some of the largest complexes, the diameter of the NPC central channel is variable and can readily pass molecules up to ~30 nm in diameter (Pante and Kann, 2002; Fahrenkrog and Aebi, 2003). No Rev polymers have been observed *in vitro*, either alone or in complex with RNA, that are larger than ~15 nm in diameter; therefore it is theoretically possible that no remodelling would be needed if such filamentous complexes are made *in vivo* during HIV replication.

Different classes of RNA are exported from the nucleus using different nuclear export factors, which are depicted in **Fig 1.6**. Our understanding of the mechanisms and dynamics of nuclear export of RNAs other than mRNA (tranfer RNA (tRNA), microRNA (miRNA), small nuclear (snRNA), ribosomal (rRNA), and viral RNA) is murky. The export of small RNAs such as tRNAs and miRNAs is similar to that of proteins, involving similar karyopherin transport factors (Kohler and Hurt, 2007). Instead of a classical NES protein sequence, these RNAs appear to carry structural RNA elements that are recognised by exportins in a manner analogous to NES-karyopherin recognition (Kohler and Hurt, 2007). Interestingly, the pathway that host cells use for mRNA transport uses an entirely different set of transport factors than the one used by HIV Rev for transport of its viral mRNAs. Rev serves as an adaptor protein that engages both its cognate RNA (explained further below) and Crm-1 for transport, similar to snRNA and rRNA, which make use of endogenous NES-containing adaptor proteins (shown in blue in **Fig 1.6**) that bind both the snRNA and Crm-1 (Fornerod et al., 1997). The Rev-RNA-Crm1 route out of the nucleus is independent of the path used by cellular mRNAs; inhibition of one pathway does not affect the other one (Pasquinelli et al., 1997; Saavedra et al., 1997).

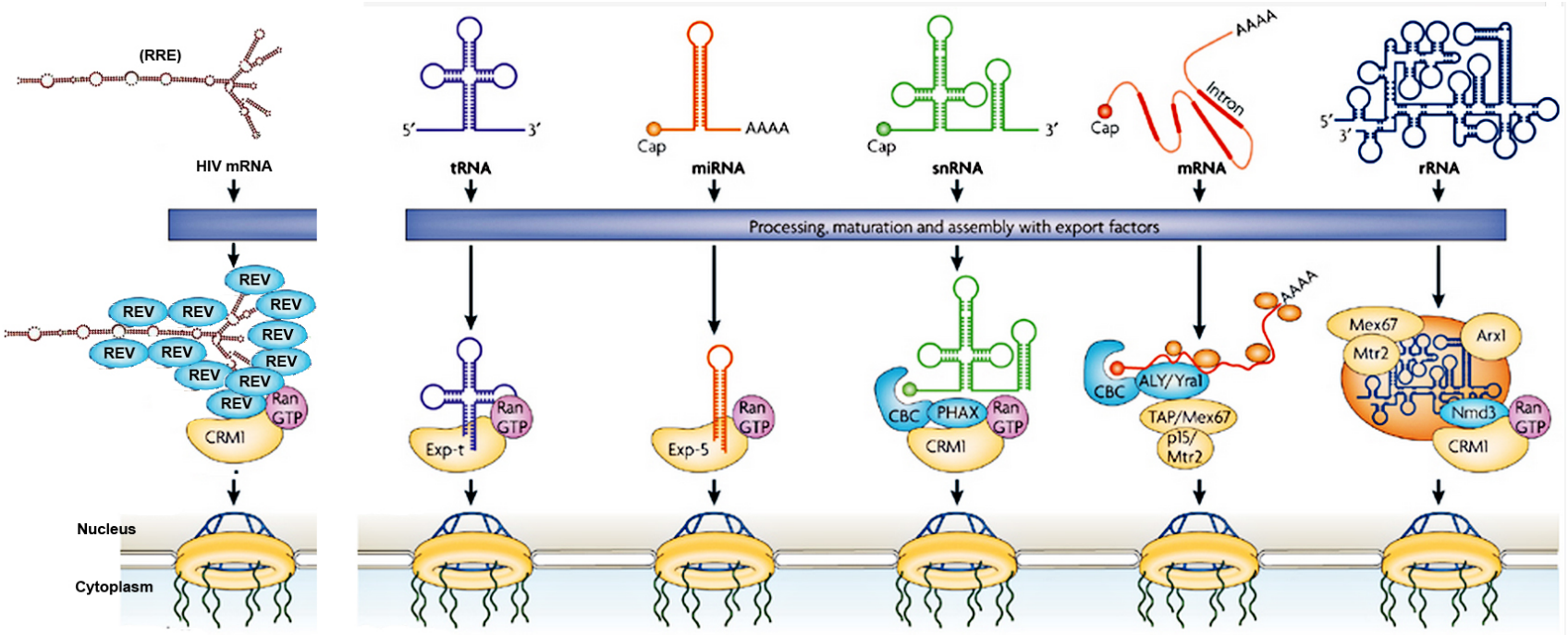


Figure 1.6: Types of RNA Nuclear Export The different known nuclear export pathways of RNA are depicted. At top are the different primary RNA transcripts: transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA), messenger RNA (mRNA), ribosomal RNA (rRNA). Below them are the transport-competent RNAs after undergoing processing, maturation, and assembly with export factors. All of the RNAs except for mRNA engage a host cell exportin (yellow) and RanGTP (purple) to mediate translocation through the NPC. tRNA and miRNA have structural motifs that directly interact with karyopherins, whereas snRNA, rRNA, and HIV RNA engage NES-containing adaptor proteins (in blue). Figure modified from (Kohler and Hurt, 2007).

1.1.3 Rev-mediated nuclear export of viral RNAs

Until very recently, structural data on Rev and on the mechanism underlying Rev-mediated nuclear export of RNA, have been relatively sparse. However, much of the molecular biology research on Rev has proceeded in the absence of a high-resolution structure.

1.1.3.1 The Rev Response Element (RRE)

Rev targets viral mRNAs by using its ARM motif to specifically recognise a *cis*-acting element within unspliced and partially-spliced HIV mRNAs known as the Rev Response Element (RRE) (Daly et al., 1989). The RRE is a ~350 nt tract that is located within an intron of the *env* gene (Malim et al., 1989b). The secondary structure of the RRE is shown in **Fig 1.7** in the context of the full HIV-1 RNA genome (Watts et al., 2009). Five stem-loop structures within the RRE have been characterised. Stem loop I (SLI) is a largely double-stranded region that comprises ~200 nt which connect the other RRE stem loops with the rest of the HIV genome. It also contains many irregularities (often termed bulges) where base-pairing is interrupted. Stem loop II-V localise to the distal end of SLI, each forming a relatively short loop relative to SLI.

1.1.3.2 Nucleation of the Rev-RRE complex

Functional characterisation of the RRE remains scarce. However, a high affinity interaction between Rev and the RRE has been localised at a purine-rich bulge within stem loop IIb of the RRE (Malim et al., 1989b) and this interaction appears to nucleate assembly of the Rev-RRE RNP complex (Malim et al., 1989a; Bogerd and Greene, 1993). We understand this interaction at the molecular level in large part due to an NMR structure that has been determined of stem loop IIb with a polypeptide stretch

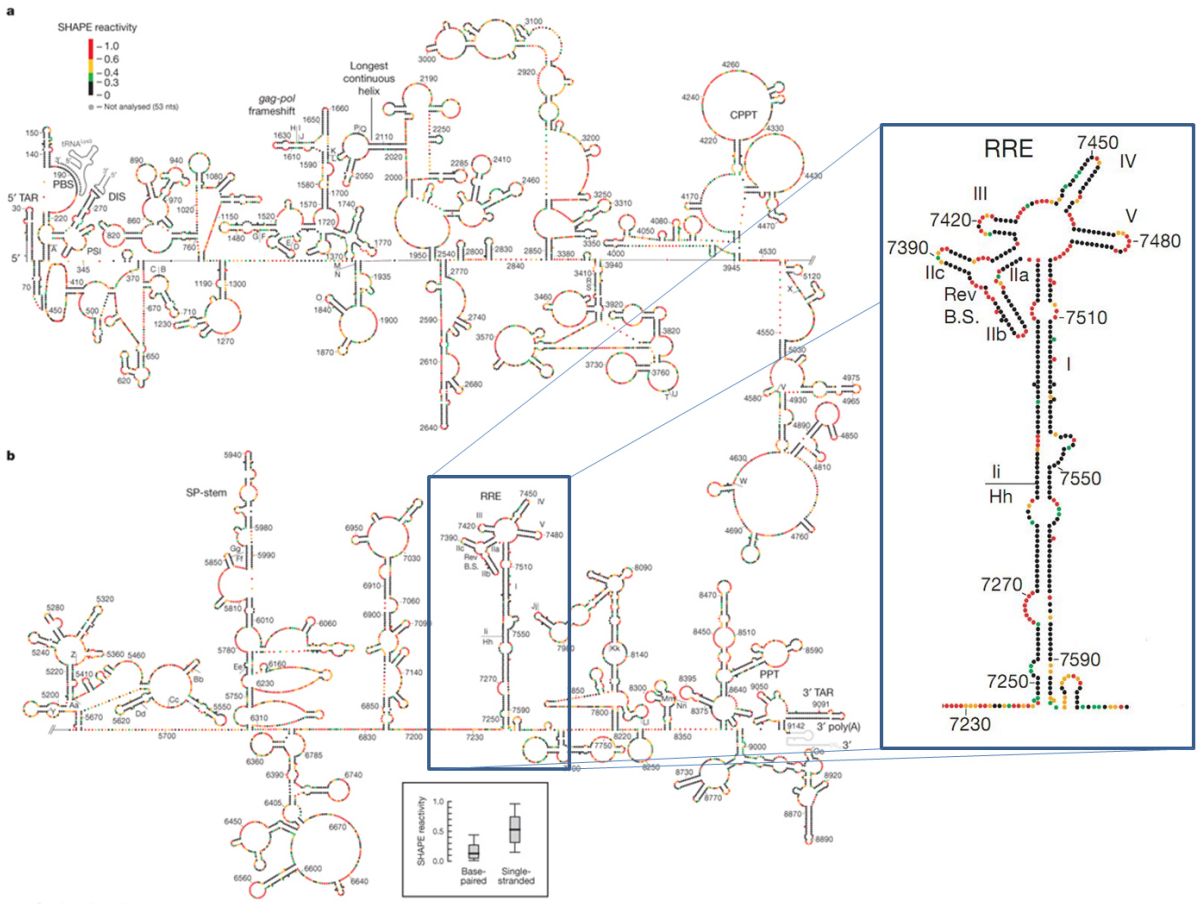


Figure 1.7: Rev Response Element (RRE) Secondary structure of HIV-1 Rev response element. The secondary structure of the entire HIV-1 genome is shown (split into two half images for space considerations, (**A**) and (**B**)), with the RRE inset. Rev first binds to the RRE via a high-affinity site within stem loop IIb, after which additional Rev molecules cooperatively bind both RRE-bound Rev and the RRE itself to form the full Rev-RRE RNP complex. The RRE contains a high degree of secondary structure relative to many other regions of the viral genome. Figure modified from Watts et al. (2009).

corresponding to the Rev ARM (Battiste et al., 1996).

1.1.3.3 Multimerisation of Rev on the RRE

Binding of one Rev dimer to the RRE is insufficient for nuclear export of RRE-containing RNAs to proceed (Malim and Cullen, 1991; Mann et al., 1994; Daugherty et al., 2008). Assembly of multiple Rev molecules onto the RRE—forming a large ribonucleoprotein (RNP) complex—is essential for Rev-mediated export (Malim et al., 1990; Iwai et al., 1992; Tiley et al., 1992; Jain and Belasco, 2001). *In vitro*, Rev molecules can multimerise onto the RRE via each subunit's ARM (Daly et al., 1993; Mann et al., 1994), effectively masking its NLS motifs (Fankhauser et al., 1991; Henderson and Percipalle, 1997). The ordering of Rev molecules on the RRE is a highly cooperative process mediated primarily through Rev-Rev interactions: the second Rev interacts with the RRE with at least ~130x greater affinity than the first (Cook et al., 1991; Mann et al., 1994; Daugherty et al., 2008; Pond et al., 2009). It remains unclear how Rev oligomerisation generates cooperativity and specificity for RRE binding.

1.1.3.4 Recruitment of Crm-1 and other host cell factors

Once Rev has multimerised onto the RRE, Crm-1 (exportin-1) engages the RNP via Rev's NES motifs (Stade et al., 1997; Fornerod et al., 1997; Cullen, 2003). Association of Rev-RRE with Crm1 mediates translocation of the RNP through the NPC and into the cytoplasm via direct interaction between Crm-1 and Nups (Fornerod et al., 1997; Neville et al., 1997; Fukuda et al., 1997). Subsequent to interaction with Crm-1, a cohort of other host cell proteins become involved to facilitate nuclear export of the HIV RNA (Strebel, 2003; Suhasini and Reddy, 2009). These factors include eukaryotic initiation factor (eIF-5A) (Elfgang et al., 1999), Src-associated protein (68 kDa) SAM68, DEAD box protein RNA helicases (DDX1 and DDX3), hRIP (human Rev-interacting protein), (Nekhai and Jeang, 2006). All

of these host cell factors have been shown to be essential for Rev-mediated nuclear export of HIV RNA (for reviews see Suhasini and Reddy (2009); Li et al. (2005)). For example, eIF-5A is essential for Rev-mediated nuclear export and mutations in eIF-5A inhibit HIV-1 replication in lymphocytes. DDX3 and DDX1 are RNA helicases which shuttle between the nucleus and cytoplasm, and DDX3 has been shown to directly interact with the Rev-RRE-Crm1 ternary complex (Dayton, 2004). It remains to be seen whether this interaction results in the formation of a quaternary complex, or rather involves binding of partners in a sequential, pass-along manner. DDX3 upregulates Rev-dependent RNA export, but not other RNA export pathways (Yedavalli et al., 2004; Fang et al., 2005).

The mechanisms by which these host cell proteins interact with the RNP and/or with the NPC to effect transport remain obscure. eIF-5A has been localised at the nuclear face of the NPC via immunogold electron microscopy, and it has been proposed to serve as an adaptor that targets Rev-RRE-Crm1 to the nucleoplasmic face of the NPC (Hofmann et al., 2001). DDX3 may be involved in remodelling of the RRE-containing RNA at some stage of transport or in facilitating its release from the NPC on the cytoplasmic face. A cartoon summarising the known host cell factors involved in export of Rev-RRE complexes and import of Rev (see next section) is shown in **Fig 1.8**. Once the Rev RNP translocates to and through the nuclear pore (Gorlich and Mattaj, 1996; Fritz and Green, 1996; Gerace, 1995), Rev may then escort mRNAs further into the cytoplasm to promote polysome association. Rev co-factors Sam68, eiF5A, hRIP, and DDX3 have all been implicated to function in translation of HIV RNA, further suggesting that Rev may play a role in translation of the viral proteins once export of the viral mRNAs into the cytoplasm has been achieved (Liu et al., 2011).

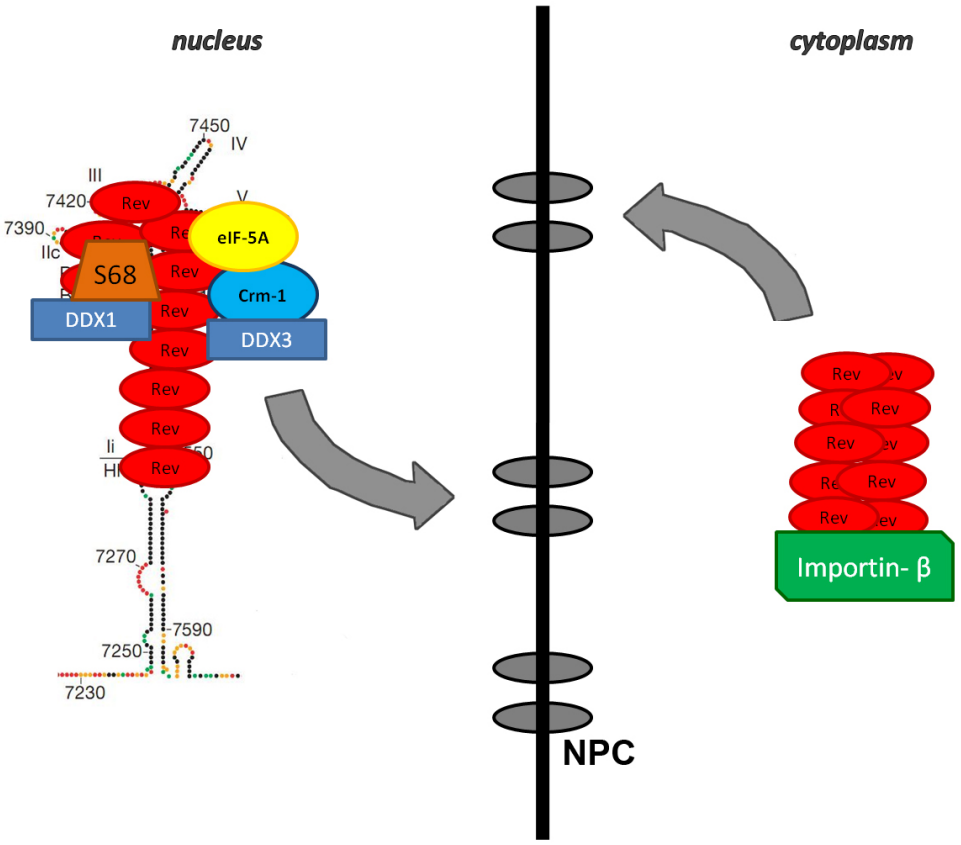


Figure 1.8: Rev-mediated RNA nuclear export Multimerisation of Rev subunits on the RRE is essential for Rev-mediated export of RRE-containing RNAs. Prior to nuclear export, a number of host cell factors are either recruited to the RNP or engaged at the nuclear pore complex (NPC) to aid in transport. RNA helicases DDX3 and DDX1 interact with Crm-1. Translation factor eIF-5A, Sam68, and many other host cell proteins (not shown) are involved, but their mechanistic roles remain unclear. Once through the NPC, the RNA is released and Rev returns to the nucleus (probably in a multimerised form) via the engagement of importin-β.

1.1.4 Nuclear import of Rev

1.1.4.1 Rev directly interacts with importin-β

Rev is first synthesised in the cytoplasm, after which it travels to the nucleus via direct engagement of its NLS with importin-β (Truant and Cullen, 1999) and a nucleolar phosphoprotein known as B23 (Szebeni et al., 1997). Importin-β interacts with nucleoporins on the NPC to permit passage across nuclear envelope into the nucleus (Pollard and Malim, 1998). Once in the nucleus, Rev self-assembly on the RRE is thought to displace B23 from Rev and to mask the NLS (Henderson and Percipalle, 1997; Fankhauser et al., 1991).

The majority of cargoes that are imported via karyopherins do so by engagement of their NLS with importin-α (Chook and Suel, 2011). Importin-α then interacts with importin-β to form a ternary complex (cargo, importin-α, importin-β) that assembles in the cytoplasm and translocates through the NPC. An alternative, less common pathway for nuclear import involves direct binding of importin-β to its cargo. A small number of protein cargoes have been found to directly engage importin-β in this manner. A non-exhaustive list includes T-cell protein tyrosine phosphatase, the parathyroid hormone-related protein, cyclin B1, transcription factors c-Jun and c-Fos, and most relevant to the topic at hand, HIV-1 Rev (Arnold et al., 2006).

1.1.4.2 Flexibility of the importin-β solenoid

The structure of importin-β helps to understand how it can directly engage so many different cargoes. Like all members of the β-karyopherin superfamily, importin-β is a solenoid protein composed of tandem HEAT repeats. Each HEAT repeat is a pair of anti-parallel α-helices that arrange in a superhelical fashion free from constraints by any intramolecular contacts between distal regions of the polypep-

tide sequence (Forwood et al., 2010). Consequently, they are often more flexible than other structured proteins and have been considered to comprise an emerging structural class between globular and intrinsically unstructured proteins (Wright and Dyson, 1999). This flexibility can be observed in crystal structures of importin-β with various cargoes (**Fig 1.9A**). A structural analysis of the conformational variability of importin-β by mutagenesis, SAXS modelling , and molecular dynamics (MD) analysis showed that the solenoid is highly flexible and in solution can adopt a fully extended conformation (Forwood et al., 2010) (**Fig 1.9B**). Cargos, via the NLS, bind to the concave side of importin-β, and the convex side interacts with the FG Nups of the NPC to allow passage (**Fig 1.9C**).

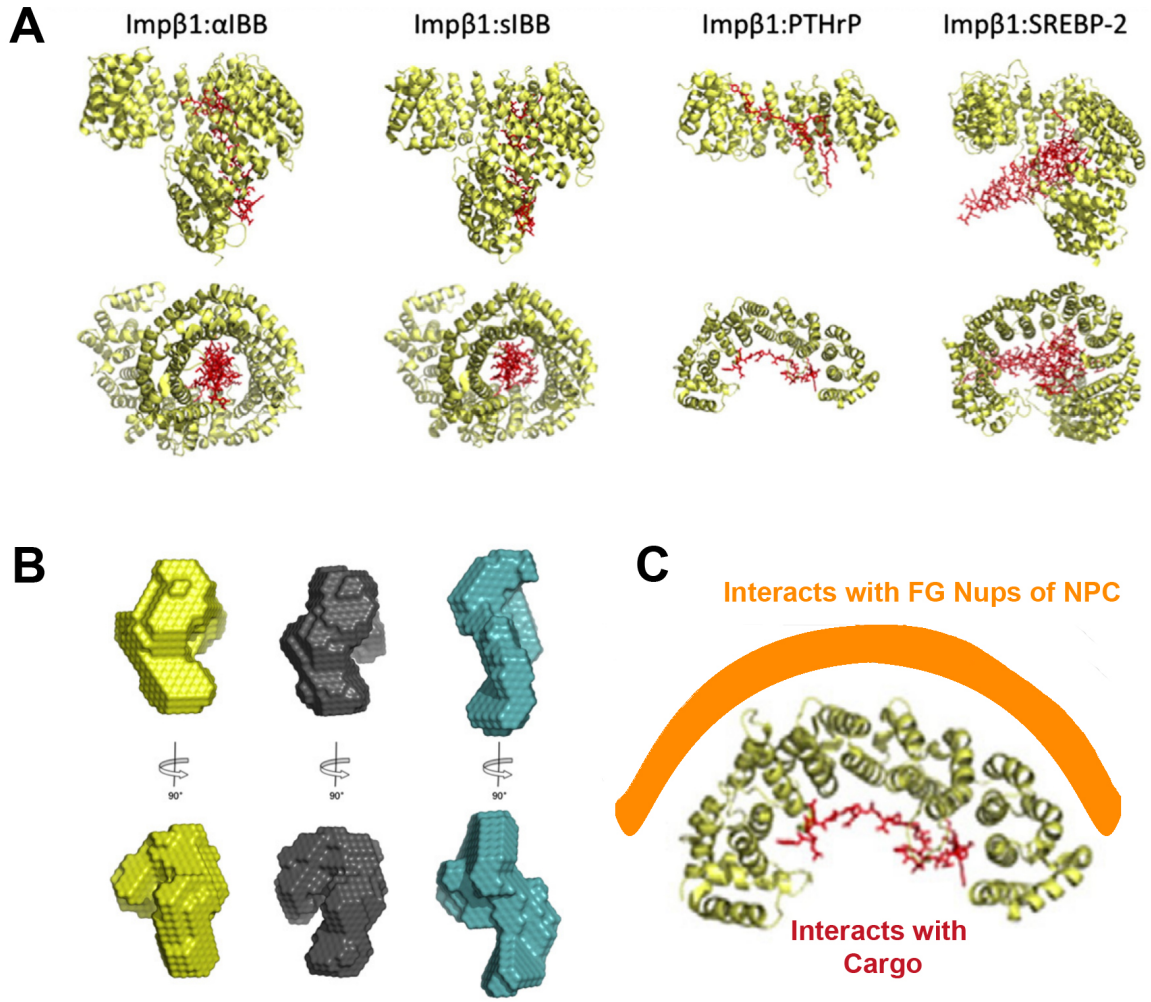


Figure 1.9: Importin-β flexibility and specificity for NPC (A) Crystal structures of importin-β complexed with various cargoes. From left, cargoes are IBB domain of importin-α (1QGK), IBB domain of snuportin-1 (2P8Q), parathyroid hormone-related protein (PTHrP) (1M5N), and sterol regulatory element-binding protein 2 (SREBP-2) (1UKL). The superhelical solenoid structure of importin-β is highly flexible to accommodate binding to different cargoes. **(B)** Low-resolution SAXS models of importin-β in various conformational states. **(C)** Cargos, via the nuclear localisation sequence, bind to the concave side of importin-β, and the convex side interacts with the FG Nups of the NPC to allow transport. Figure modified from Marfori et al. (2011) and Forwood et al. (2010).

1.1.5 *In vitro*-assembled Rev polymers

1.1.5.1 Rev and Rev-RRE complexes are qualitatively different by EM

The interactions whereby Rev recognises RRE-containing RNAs have been intensively studied, as has Rev's propensity to oligomerise, both in the absence of RNA and in the context of the RRE (Bogerd and Greene, 1993; Malim et al., 1989a; Watts et al., 1998; Wingfield et al., 1991). *In vitro* Rev readily forms polymers with two distinct morphologies: one in the presence of RNA and the other without any nucleic acid (**Fig 1.10**). Rev alone forms regular, hollow filaments with a diameter of ~15 nm. With RNA, Rev forms poorly ordered filaments ~7-9 nm in diameter with length roughly proportional to the size of the RNA.

1.1.5.2 Rev-RRE RNP: discrete complex or long filament?

We know that Rev functions as an adaptor protein that can bind both cognate HIV mRNAs and host cell exportin, Crm-1. This direct coupling of the viral RNA with Crm-1 circumvents the typical processing checkpoints that cellular mRNA must undergo prior to export (see **Fig 1.6**). However, in order to effect this Crm-1 mediated transport, Rev must first multimerise onto the RRE. This complex is thought to comprise the ~350 nt of the RRE and 8-12 Rev monomers. In this model, the Rev-RRE complex forms a discrete architecture that mediates nuclear export through the NPC, with the remainder of the Rev-unbound HIV mRNA towing behind. There has been some effort to isolate and structurally characterise such a discrete Rev-RRE complex (Daugherty et al., 2010a). However, the majority of reported structural investigations into Rev-RRE interactions actually support the formation of RNA-protein filaments, nucleated by interaction of Rev with stem loop IIb of the RRE:

- In 1991, Wingfield et al. observed Rev-RRE filaments by negative-stain EM that were 8 nm in diameter and more flexible in appearance than Rev filaments. The Rev-RNA filaments formed at lower Rev concentrations than Rev alone polymerises, suggesting that the RNA clearly stimulated the RNP formation and that the interaction is thermodynamically favorable.

- Also in 1991, Heaphy et al. observed Rev-RRE filaments by negative-stain EM. Complexing of Rev with the RRE (a 238-nt transcript) gave rise to short, rod-like RNP complexes with a length of ~60 nm. However, complexing of Rev with a 2.4 kb RNA (corresponding to *env* gene) resulted in long filaments 500-700 nm in length.

- In 2007, Havlin et al. performed negative-stain EM, AFM, and 2D solid state NMR on Rev and Rev-RRE assemblies. They found that the morphologies of Rev filaments and Rev-RNA co-assemblies are qualitatively different, but that Rev has the same conformation in both supramolecular assemblies. The Rev-RRE assemblies were estimated to be 7 ± 2nm in diameter. The authors suggested that Rev-RRE assemblies have a negatively-charged exterior surface, due to the absence of detectable adsorption of Rev-RRE to mica at pH 7. Lastly, for both Rev and Rev-RRE, the 2D spectra suggested the presence of additional α-helical content than what is present in the NTD.

These results altogether suggest that Rev and RNA can form filaments with a length proportional to the size of RNA. This suggests that the RNA is bound to Rev throughout its entire length, subsequent to the initial nucleation of Rev with the RRE. Rev-RNA filaments are often described as more poorly ordered (on the 5-10 nm length scale) than Rev filaments. Indeed, they are definitely more flexible; however, solid state NMR spectra for Rev- and Rev-RNA assemblies were nearly indistinguishable, suggesting that Rev has the same secondary structure in both assemblies (Havlin et al., 2007).

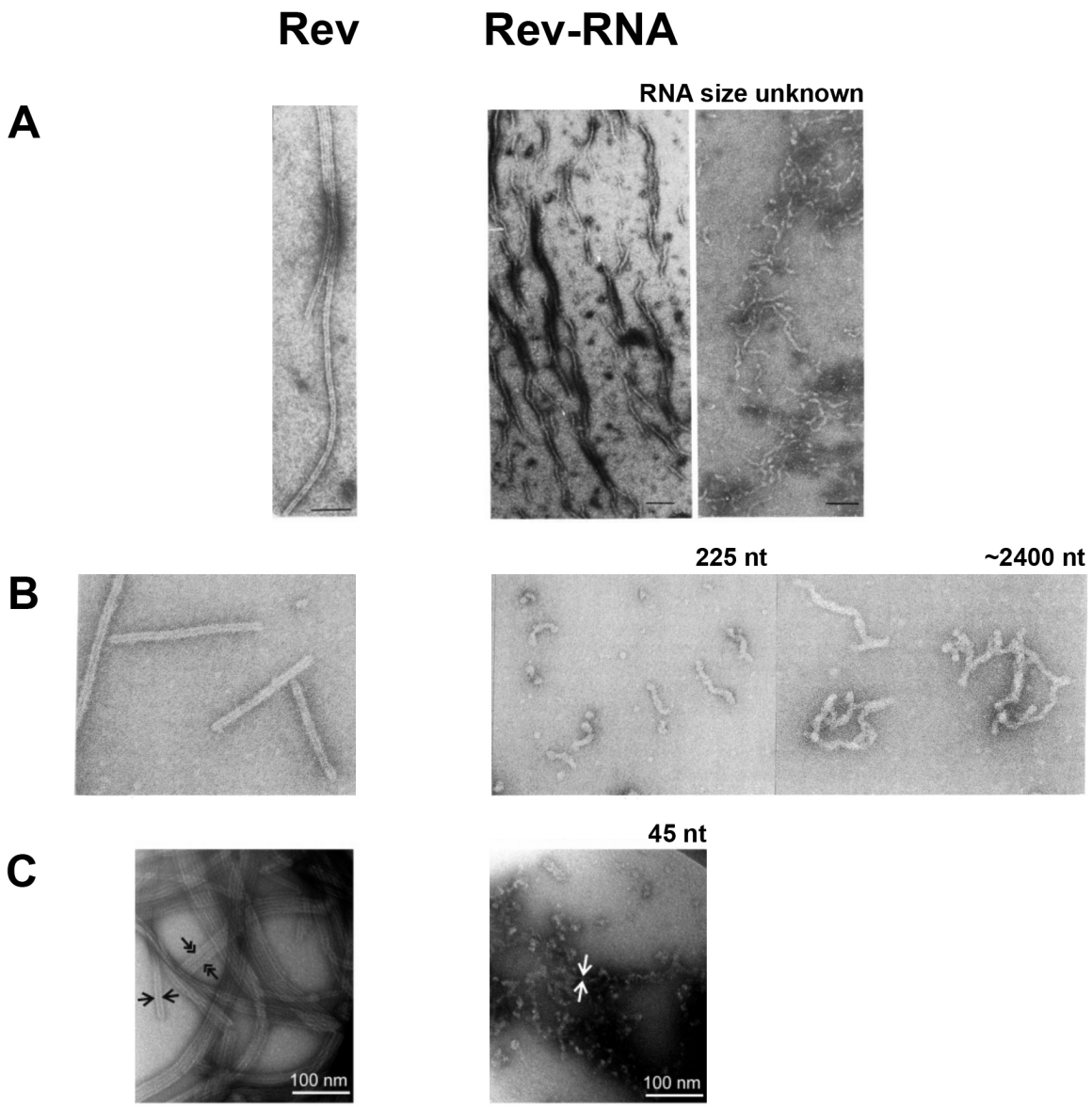


Figure 1.10: Negative-stain EM of Rev and Rev-RNA assemblies Rev filaments are shown in the left column and Rev-RNA assemblies in the right. Images reproduced from three studies (A) Wingfield et al. (1991); (B) Heaphy et al. (1991); and (C) Havlin et al. (2007). The size of RNA that was complexed with Rev is indicated above the panel, where known. Rev filaments are hollow and regular in diameter and Rev-RNA filaments appear less regular with a length proportional to size of RNA. Scale bars = 100 nm.

1.1.6 Scope of Investigation

1.1.6.1 Questions to Consider

Although much is known regarding the function of Rev as a modulator of nuclear export of HIV mRNAs, many mechanistic questions still remain unanswered:

1. What are the structural features of HIV-1 Rev that permit it to simultaneously multimerise, bind RNA, and interact with other host cell proteins?

2. Does Rev multimerisation exhibit head-to-head and tail-to tail interactions, or simply head-to-tail interactions (i.e. is there one Rev-Rev interface that propagates asymmetrically or two, symmetrically)?

3. How is binding of Rev onto RRE-containing RNAs so highly cooperative?

4. Why is Rev multimerisation requisite for Rev-mediated nuclear export, if as we know, one Rev molecule can bind the RRE high-affinity stem loop IIB?

5. What are the structural arrangements of Rev subunits in *in vitro*-assembled Rev filaments, and how might they relate to supramolecular forms Rev adopts *in vivo* to engage the RRE for nuclear export, and subsequently, importin-β for nuclear import?

6. What structural platform(s) might be present, if any, on Rev and Rev-RNA assemblies that would allow for engagement of the many host cell proteins that aid Rev in effecting nucleocytoplasmic transport?

1.1.6.2 Potential for antiviral development against Rev function

Despite decades of study, many unanswered questions remain regarding the mechanisms that HIV implements at each step of its replication cycle to successfully infect an individual. Investigation of the structural basis for these mechanisms may aid understanding of how HIV interacts with, and seizes various host cell systems in order to propagate itself. Additionally, development of novel antiviral therapeutics against these processes can be facilitated with new structural data. The antiviral drugs developed thus far have targeted the viral enzymes and PR, IN, RT, and the envelope glycoproteins, yet significantly, none have been developed against Rev or Tat, the regulatory proteins that are also essential for HIV infection.

1.2 Hepatitis B virus (HBV)

1.2.1 Epidemiology

Hepatitis B virus (HBV) infection remains a major source of acute and chronic liver disease worldwide. More than 360 million people have chronic HBV infection, which results in one million deaths annually, primarily due to cirrhosis and liver cancer (World Health Organization). HBV is an enveloped, non-cytopathic virus with a circular, partially- double-stranded DNA genome. While an effective vaccine against HBV infection has been in existence for over 20 years, the spread of the virus remains a public health problem on an immense scale, especially in areas of Asia and sub-Saharan Africa where it is endemic (Maddrey, 2000). Furthermore, the prophylactic vaccine is not always effective in eliciting an immune response and is not a viable treatment for established chronic infections. Regarding treatment, a number of antiviral drugs have been developed based on nucleoside analog reverse transcriptase inhibitors; however, they have been minimally successful in eliminating the virus. There is a clear need for improvement of clinical treatment options for chronic HBV infection (Seto et al., 2012).

1.2.2 The enigmatic e-antigen

Over the four decades since the discovery of HBV, striking advances have been made in our understanding of the molecular biology, immunology, and pathogenesis of infection. However, certain aspects of HBV biology remain elusive. While HBeAg has no demonstrated role in the viral replication cycle (Chang et al., 1987; Chen et al., 1992; Schlicht et al., 1987), the antigen has long been a key clinical marker for viral replication, infectivity, disease severity, and response to treatment (Elgouhari et al., 2008). Furthermore, HBeAg (or an equivalent) exists in all members of the *Hepadnaviridae* family,

suggesting an evolutionarily conserved and therefore important function (Revill et al., 2010).

1.2.3 Comparison of core and e-antigens

The HBV capsid protein (HBcAg; core antigen) comprises a 149-residue assembly domain and a 34-residue arginine-rich domain at the C-terminal end. The assembly domain forms dimers with a central four-helix bundle and flanking α-helices that assemble into icosahedral capsids of two sizes, with the four-helix bundles projecting as spikes (Wynne et al., 1999; Packianathan et al., 2010) (**Fig 1.11**). HBeAg consists of the ten N-terminal residues (the propeptide: SKLCLGWLWG) appended to the assembly domain with the C terminus at residue 149 (Takahashi et al., 1983; Ou et al., 1986; Standring et al., 1988). Translation of the *C* gene from an alternative upstream start codon yields a protein with a 29-residue signal peptide that routes it to the endoplasmic reticulum, where it is processed to the 10-residue propeptide (Standring et al., 1988). However, despite possessing an intact assembly domain, HBeAg does not assemble into capsids and is secreted by infected liver cells in non-particulate form.

1.2.4 HBeAg as a modulator of the immune system

Much evidence suggests that HBeAg can modulate the host immune response to favor chronic infection following perinatal transmission (the most common form of HBV transmission worldwide) and prevent severe liver injury during adult infections (Ou, 1997; Milich and Liang, 2003a; Chen et al., 2004, 2005; Yang et al., 2006; Visvanathan et al., 2007). The epidemiologic evidence is persuasive: more than 90% of infants born to mothers who are HBeAg-positive HBV carriers also develop chronic infection, whereas those born to HBeAg-negative mothers rarely progress to chronicity (Terazawa et al., 1991). While the molecular mechanisms underlying these processes are unclear, it has been shown that HBeAg

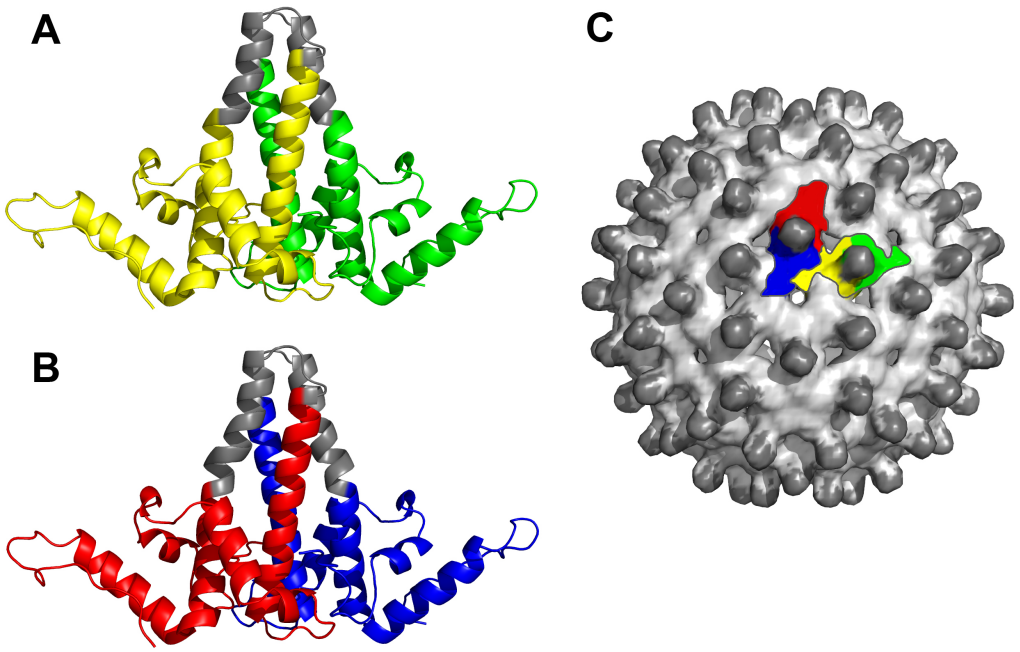


Figure 1.11: HBV core antigen.
(A, B) The HBV capsid, also known as core antigen (HBcAg), comprises a 149-residue assembly domain and a 34-residue arginine-rich domain. The assembly domain forms dimers with a central four-helix bundle and flanking α-helices that assemble into icosahedral capsids of two sizes (T=3 and T=4). The first 142 residues are ordered in the crystal structure; the rest are disordered. The T=4 capsid is shown in (**C**). Four HBcAg polypeptides comprise the viral asymmetric unit (coloured in yellow, green for the first dimer; red, blue for the second). Their arrangement in the T=4 capsid is depicted (**C**). It has remained a puzzle how HBeAg also contains the polypeptide corresponding to the assembly domain, but is incapable of assembling into HBV capsids as shown. The extra 10 residues at the N-terminus, termed the propeptide, have been posited to play a role in this structural distinction. (Conway et al., 1997; Bottcher et al., 1997; Wynne et al., 1999) **(C)**

can downregulate the inflammatory response directed at HBcAg, while itself averting robust immunogenicity (Milich and Liang, 2003a; Chen et al., 2004, 2005). Furthermore, HBeAg (but not HBcAg) can cross the placenta from mother to child (Schodel et al., 1993), consistent with data suggesting that HBeAg may induce clonal tolerance against HBcAg and HBeAg *in utero* (Milich et al., 1990; Chen et al., 2004). While the connection between HBeAg and chronic infection is not fully understood, infections with HBV strains that do not express HBeAg (due to mutation in pre-C region) lead to much higher frequency of fulminant hepatitis and acute liver failure (Fagan et al., 1986; Liang et al., 1991). In these circumstances, the lack of immune modulation by HBeAg is thought to lead to an unregulated and overwhelming immune response to HBcAg. To develop a fundamental understanding of the functional distinction between HBcAg and HBeAg, knowledge of their respective structures is essential.

Chapter 2

Experimental Procedures

2.1 HIV-1 Rev

Rev's strong tendency to polymerise and/or aggregate (Wingfield et al., 1991) has previously thwarted attempts to solve its structure either by X-ray crystallography or by NMR spectroscopy. So far, detailed structural information has been limited to an NMR structure of a 22-residue polypeptide corresponding to the Rev ARM in complex with stem loop IIb (Battiste et al., 1996). The approach used herein was to complex Rev with an antibody Fab fragment that would stabilise Rev in a soluble, non-aggregated form. Such Fabs have been used to aid in the structural study of many proteins that are otherwise refractory to crystallisation, in particular membrane proteins (Ye et al., 2008; Zhou et al., 2001; Uysal et al., 2009). The section below details the methods performed, beginning with the isolation of a high-affinity anti-Rev Fab, and continuing through crystallisation and structure determination of the Rev-Fab complex, finishing with the structural analysis.

2.1.1 Isolation of anti-Rev antibody Fab fragments

Two rabbits were immunised with purified recombinant HIV-1 Rev, after which the spleen and bone marrow were collected and processed for total mRNA extraction. A phage display library containing approximately 2×10^8 chimeric rabbit/human antibody Fab clones was prepared in three steps:

1. RT-PCR amplification of rabbit Fab variable domain (V_K, V_Λ, V_H) encoding sequences;

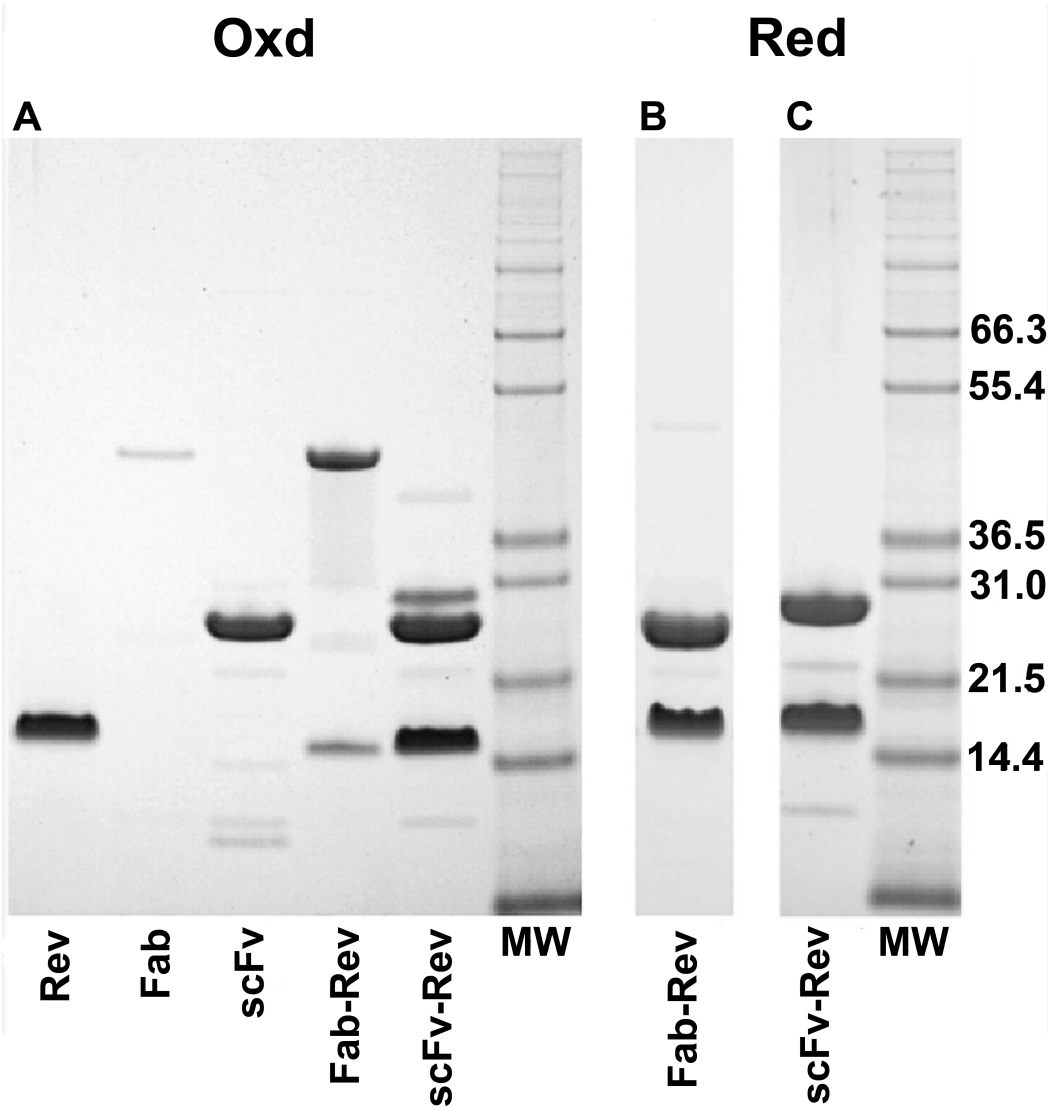
2. insertion of the sequences into a V_K-C_K-V_H cassette assembly; and
3. asymmetric SfiI ligation into phage display vector pC3C (Stahl et al., 2010).

In order to select for Fab fragments that bind Rev with sub-micromolar affinity, a panning apparatus was arranged in which recombinant Rev protein was biotinylated at the C-terminus and subsequently immobilised onto streptavidin-coated plates. The Fab clones were screened over the Rev-immobilised plates, and after four or five rounds of panning, all remaining clones were sequenced and found to be identical. Initial characterisation of these clones was done using ELISA, DNA fingerprinting, and DNA sequencing; see Stahl et al. (2010) for full details on protocol. The remaining clone was selected for further biophysical characterisation, in particular for its apparent ability to bind Rev. For clarity, this anti-Rev Fab, and its corresponding single-chain variable domain, will be simply referred to as "Fab" and "scFv", respectively, in all further relevant sections regarding HIV-1 Rev. The Fab variable domain sequences (V_K and V_H) were found to be unique when compared against all other known Fab sequences and they are presented in **Fig 2.1**, with the hyper-variable loops highlighted in a different colour.

2.1.2 Preparation of Fab-Rev and scFv-Rev complexes

2.1.2.1 Protein expression

Both HIV-1 Rev and the anti-Rev Fab were expressed individually in *E. coli*(as described previously for Rev (Wingfield et al., 1991)) and subsequently mixed to form a complex. Due to Rev's propensity to oligomerise even at low concentrations (threshold conc. of 80 µg ml^{-1}), Rev was diluted in 6 M urea to a concentration of 70 µg mL^{-1} and subsequently folded by dialysis against 50 mM sodium phosphate pH 7.0, 150 mM sodium chloride, 600 mM ammonium sulfate, 1 mM EDTA, and 1 mM DTT at 4 °C. A second round of dialysis was performed against 50 mM sodium phosphate pH 7.0 and 150 mM



D

Light Chain - V_K

ELVMTQTPSSVSEPVGGTVTIKCQASQSISSWLSWYQQKPGQPPKLLIYDASNLASGVPSRF
MGSGSGTEYTLTISGVQREDAATYYCLGGYPAASYRTAFGGGTELEII

Heavy Chain - V_H

QEQLVESGGRLVTPGTALTLTCKVSGFSLSGFWLNWVRQAPGKGLEWVGAIYRGSGSEWYAS
WAKGRFTISDTSTTVTLKLTSPTTEDTATYFCAADTTDNGYFTIWGPGTLVTVSS

Figure 2.1: SDS PAGE and sequence of Anti-Rev Fab variable domain. SDS Page of Rev, Fab, scFv, and the immune complexes using 4-12% acrylamide gels on **A** not reduced, and **B, C** reduced proteins. The purity of Rev, Fab and scFv (alone and in complex) is indicated. **(D)** Variable domain sequences of anti-Rev Fab, both light (V_K) and heavy (V_H) chains. The complementarity-determining regions (CDRs) are highlighted in different colours. [Figure modified from Stahl et al. (2010)]

sodium chloride. The Rev sample was then filtered and stored frozen at -80 °C until complexing with Fab. The Fab was also expressed in *E. coli*, with the aid of an expression cassette with two N-terminal signal sequences (pelB and ompA), which direct the heavy and light chains into the periplasmic space where inter- and intra-disulfide bonds can appropriately form. A C-terminal His tag on the Fab heavy chain (comprising 6 His residues) allowed for affinity purification via two cycles of Ni-Sepharose chromatography. The final purification step involved gel filtration using a Superdex S200 column in the presence of 1 M urea to prevent Fab aggregation, while maintaining proper conformation. Using the expression and purification procedure described above, a corresponding single chain Fab (scFv)–containing only the variable domain light and heavy chains connected by a serine-glycine linker–was also produced.

2.1.2.2 Complex formation

Fab-Rev and scFv-Rev complexes were prepared by mixing purified Fab with several-fold molar excess of Rev, followed by purification with N-Sepharose chromatography, again utilising the C-terminal His tag present on the Fab heavy chain. Final purification (polishing) of the Fab-Rev and scFv-Rev complexes could not be achieved by gel filtration due to the strong adsorption of Rev to gel-filtration matrices. Therefore, the samples were centrifuged at 100,000 *g* for 2 h to remove any protein aggregates. Final yields of the complexes were typically 0.5-1.0 mL at 5.0 mg mL^{-1}.

2.1.2.3 SDS PAGE

Once the Fab and scFv fragments were complexed with Rev, the resulting complexes were analysed by SDS-PAGE, which indicated high purity (despite the inability to use size-exclusion chromatography as a final purification step) and an apparent 1:1 stoichiometry between binding partners (optical

densitometry indicated molar ratios of 1:0.9 between Fab/scFv and Rev, for both complexes) (**Fig 2.1**). However, analysing the stoichiometry of protein-protein interactions using SDS-PAGE alone is imprecise and limited by the assumption that the binding partners have equal Coomassie dye binding capacity. Further, the possibility of having a mixture of different oligomeric species cannot be ruled out here.

2.1.2.4 Sedimentation velocity and equilibrium analysis

Sedimentation velocity analysis was also performed to ensure monodispersity of the complexes and to estimate the molecular mass of the complexes. The resulting data showed a single moving boundary, indicative of a monodisperse system. The sedimentation coefficient distribution plot also indicated a single species (not shown), and fitting of the data allowed the diffusion coefficient to be calculated, giving an estimate of the molecular mass of the complex (~63 kDa). The predicted mass of a 1:1 complex between Rev and Fab is 60.8 kDa, a value consistent with the sedimentation velocity analysis. In order to verify the stoichiometry of the Fab-Rev interaction, sedimentation equilibrium analysis was also performed. The molecular weights under native conditions of the Fab-Rev and scFv-Rev complexes were determined to be 68.1$\pm$0.70 kDa and 40.4$\pm$0.90 kDa, respectively.

2.1.2.5 Negative-stain electron microscopy

Addition of equal molar ratios of Fab molecules to Rev monomers was found by negative-stain electron microscopy to depolymerise the filaments, while forming small, uniformly sized complexes (**Fig 2.2**). Therefore, it was deduced that Rev has greater affinity for this Fab than for another Rev monomer. As the association constant for the addition of a Rev monomer to oligomerised Rev is approximately 1.0 μM, it was concluded that the Fab had at least micromolar affinity for Rev.

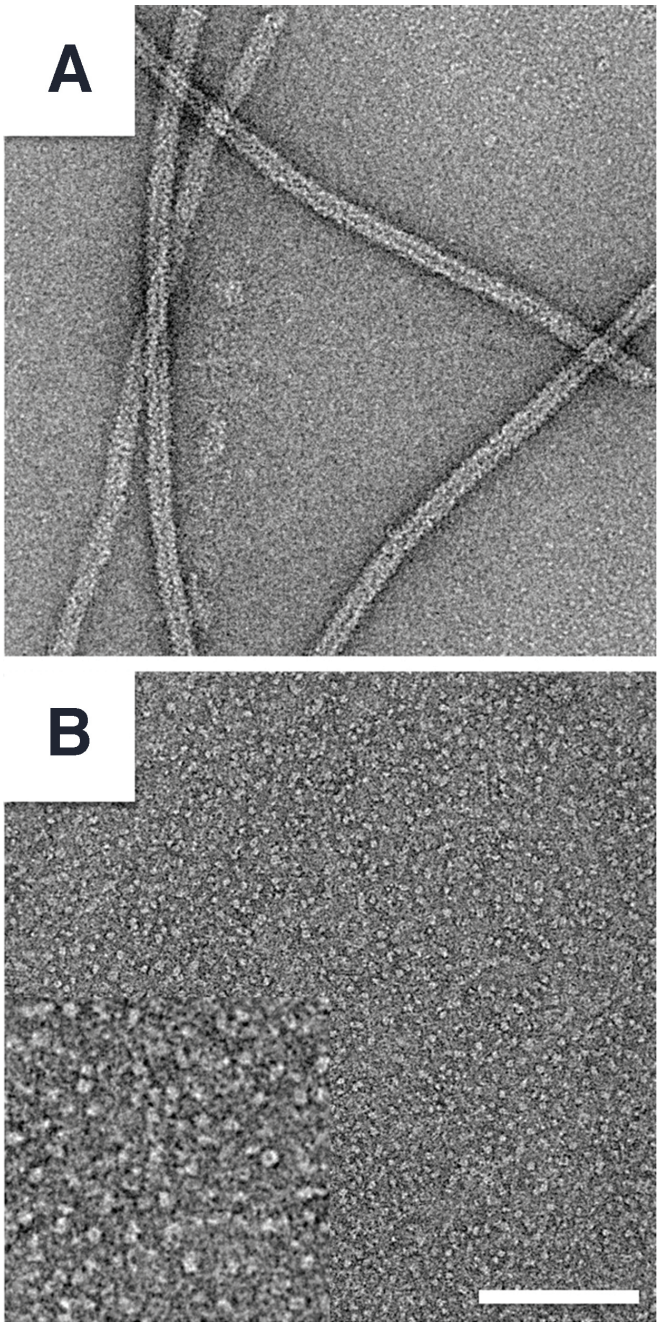


Figure 2.2: The anti-Rev Fab depolymerises Rev filaments. (A) Negatively stained electron micrograph showing filaments of HIV Rev. The filaments readily form *in vitro* at Rev concentrations as low as 80 $\mu g\,ml^{-1}$ They are observed to be long with a hollow lumen (central darker region is occupied by negative stain). **(B)** When the Rev filaments are mixed in an equimolar ratio with the high affinity anti-Rev Fab, the Fab-Rev interaction induces rapid depolymerisation of the filaments into stable, discrete, and uniformly-sized complexes. The scale bar equals 100 nm (50 nm in the inset; bottom-left quadrant). [Figure modified from (Stahl et al., 2010)].

2.1.2.6 SPR analysis

Surface plasmon resonance (SPR) was used to quantitatively determine the kinetics of the interaction between Fab and Rev. Rev was immobilised on the chip and Fab was used as the analyte. The on and off rates for the Fab-Rev interaction were measured to be 2.2e5 M^{-1} s^{-1} and 0.8e-5 s^{-1}, respectively, resulting in a calculated dissociation constant (K_D) of approximately 40 pM, or 0.04e-9 nM (data not shown) (Stahl et al., 2010).

2.1.3 Crystallisation and X-ray data collection of Rev immune complexes

2.1.3.1 Fab-Rev complex

Crystallisation trials of the Fab-Rev complex were performed at a protein concentration of 9.6 mg mL^{-1} in 20 mM HEPES pH 8.0 at 21°. Hanging drops were prepared containing 200 nL protein and 200 nL precipitant solution equilibrated against 50 µL reservoirs in 96-well plates. Primary crystallisation screening was performed with pre-formulated screens (~2000 reservoir conditions). Microcrystals grew in 12% PEG 6000, 100 mM diammonium phosphate (DAP), and 100 mM Tris (pH 8.5). Optimisation of this hit was performed, varying protein-to-reservoir volume ratio, reservoir component concentrations, and pH, as well as the presence of various additives (Hampton Research). Optimised crystals were grown in 9-14% PEG 6000, 100-200 µM DAP, 100 mM Tris (pH 8.5) and with 50 µM spermidine added to the drop. These crystals were cryoprotected by a quick pass through reservoir solution supplemented with 25% v/v ethylene glycol before flash cryocooling in a cold (100 K) stream of nitrogen gas. Diffraction data on Fab-Rev were recorded from a single crystal to a resolution of 3.2 Å at the Diamond Light Source, Beamline I02, Didcot, UK. In total, 318° of diffraction data were collected (space group: *P*1) with an oscillation angle of 1.0°. on an ADSC Q315 CCD detector and were subsequently integrated and scaled using HKL2000 (Otwinowski and Minor, 1997).

2.1.3.2 scFv-Rev complex

Purified scFv-Rev was concentrated to 5.0 mg mL^{-1} and crystallisation trials were set up at 21° in sitting drops containing 100 nL protein and 100 nL precipitant solution equilibrated against 95 µL reservoirs in 96-well plates. Crystals of scFv-Rev were initially grown in 20% PEG 3350, a variety of different salts (200 mM sodium sulfate, sodium bromide or ammonium phosphate dibasic), and pH's ranging from

6.5 to 8.5. Optimised crystals based on the initial hit condition—varying protein-to-precipitant drop ratios and reservoir concentrations—were cryoprotected by a quick pass through reservoir solution supplemented with 30% v/v ethylene glycol before flash cryo-cooling in a cold (100 K) stream of nitrogen gas (**Fig 2.3**).

Five unique diffraction data sets of scFv-Rev were recorded, each from a single crystal. The crystal lattices for each data set belonged to a different space group. The crystal that diffracted to the highest resolution was an optimised crystal grown in 20% PEG 3350 and 200 mM sodium sulfate, from which diffraction data were collected to 2.3 Å resolution and subsequently processed using XDS and XSCALE (Kabsch, 2010). A total of 1800 frames were collected with an oscillation angle of 0.1° on a Pilatus P6M detector at the Diamond Light Source, Beamline I24 (microfocus), Didcot, UK.

2.1.3.3 SDS PAGE on resolubilised crystal

In order to check for proteolysis of Rev during the crystallisation growth phase, a collection of Fab-Rev crystals were harvested and centrifuged at low speeds (1,500 rpm) to separate from soluble proteins. The crystals were then washed five times with reservoir solution to remove any loose protein on crystal surfaces and in solution. The washes along with the crystals were run on SDS-PAGE to check for banding of appropriate sizes corresponding to Fab and Rev. Fab and Rev bands appear in equimolar ratios, with minor contaminants running below ~5% (**Fig 2.4**)

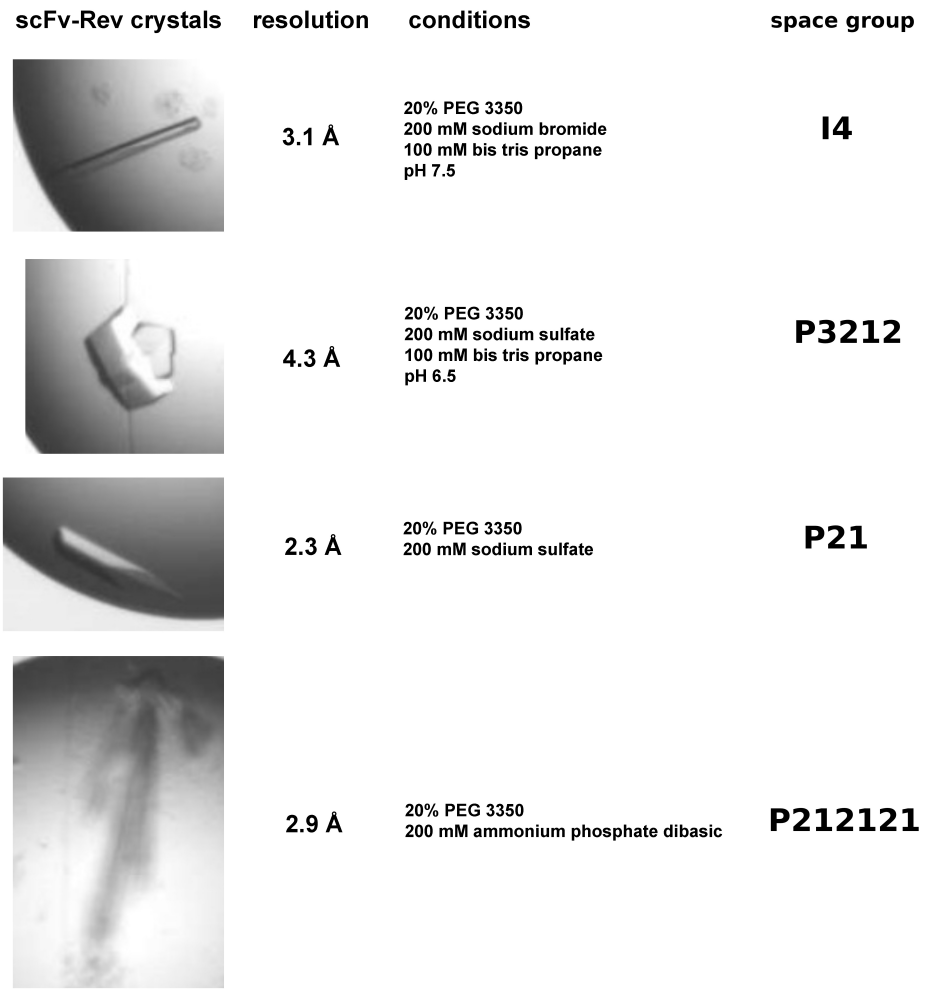


Figure 2.3: Example crystals of scFv-Rev complex. Shown above are the scFv-Rev crystals from which diffraction data were collected. Once the data were indexed, the space groups were found to be different for each crystal and the resolution of the data is reported for each, as well as the crystallisation conditions from which each crystal grew. From top, the space groups for each crystal lattice were $I4$, $P3_212$, $P2_1$, $P2_12_12_1$. The image of the crystals corresponding to the $P2_12_12_1$ space group (bottom) is blurry due to the crystals being out of the focal plane for automated imaging, as they are sitting in the bottom of the crystallisation drop.

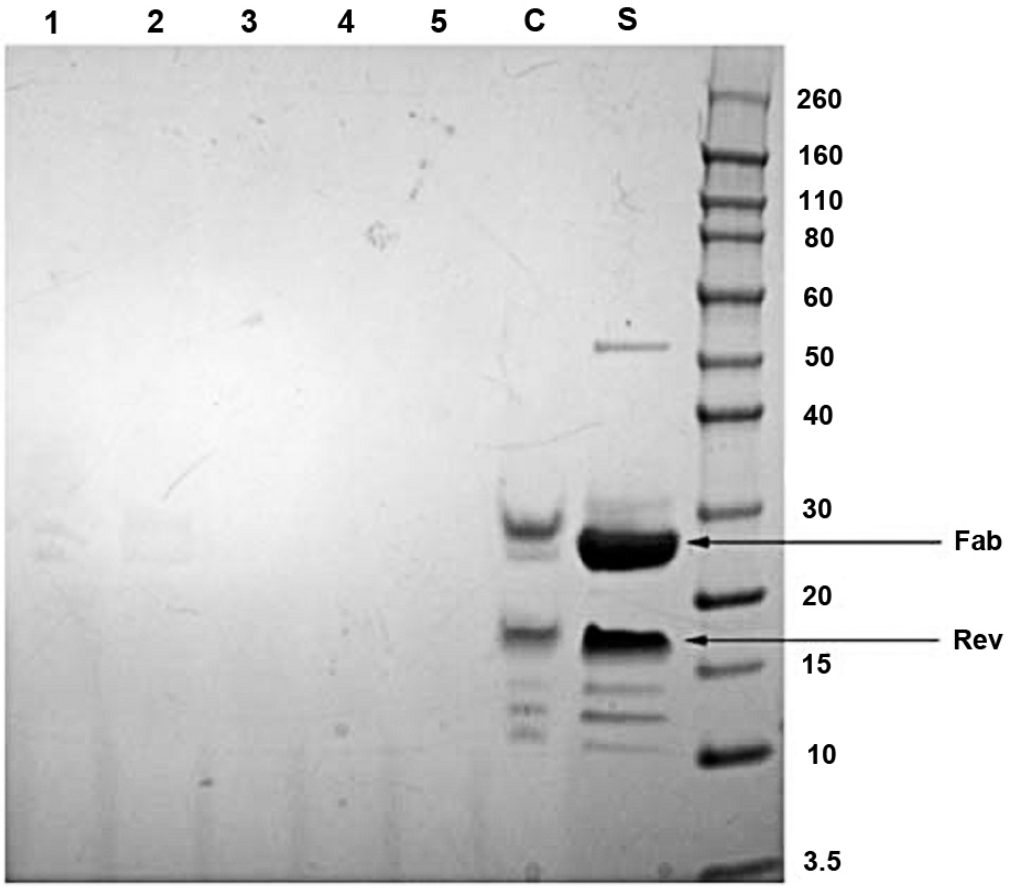


Figure 2.4: SDS-PAGE of resolubilised Fab-Rev crystal. Fab-Rev crystals were spun down at low centrifugation speeds (1,500 rpm) and washed five times with reservoir solution (lanes 1-5) to remove any loose protein on crystal surfaces and in solution. Lane C is the contents of the Fab-Rev crystal and Lane S is the Fab-Rev sample used for crystallisation. Fab and Rev bands appear in equimolar ratios, with minor contaminants running below (~5%).

2.1.4 HIV-1 Rev structure determination

2.1.4.1 Structure solution

Initial phase information for the Rev-Fab diffraction data was obtained via an automated Phaser molecular replacement (MR) search using a set of published Fab structures as search models (McCoy et al., 2007). The Fold and Function Assignment System (FFAS03) web server was used to identify the 10 light and heavy chain Fab structures in the PDB with the highest sequence similarity to the corresponding sequences of the anti-Rev Fab (Jaroszewski et al., 2005). The correct solution was found by superimposing 18 Fab structures (1JPS, 2QR0, 2R8S, 3BDY, 1TZH, 1ZA3, 2R0K, 2V7N, 3DVG, 1B2W, 1BEY, 1TZI, 2FJF, 2QQN, 1B4J, 2FGW, 2FJH, 3D85) on both the variable and constant domains and using them as one search ensemble in Phaser (McCoy et al., 2007). The space group is $P1$ with six Fab molecules and six Rev subunits (three dimers) in the asymmetric unit.

2.1.4.2 Refinement

The molecular replacement solution was rigid body refined in Phenix with each Ig domain treated separately. Positional refinement and group B-factor refinement (one group per residue) were carried out iteratively in both Phenix (Adams et al., 2010) and AutoBuster (Blanc et al., 2004). Care was taken throughout refinement not to overfit the data. The Rev structure was manually built *de novo* by first placing two α-helices in the strong tubular densities, using Coot (Emsley and Cowtan, 2004). The directionality and connectivity of the two helices were determined by first placing the bulky side-chains of F17, Y21, W45, and Y63 (Fig 2.5) in the N-terminal half. Final NCS-restrained (sixfold) refinement was performed in AutoBuster. No biochemical data on Rev was used to guide Rev building, except the primary sequence. The Molprobity server (Davis et al., 2007) and the validation tools in Coot informed

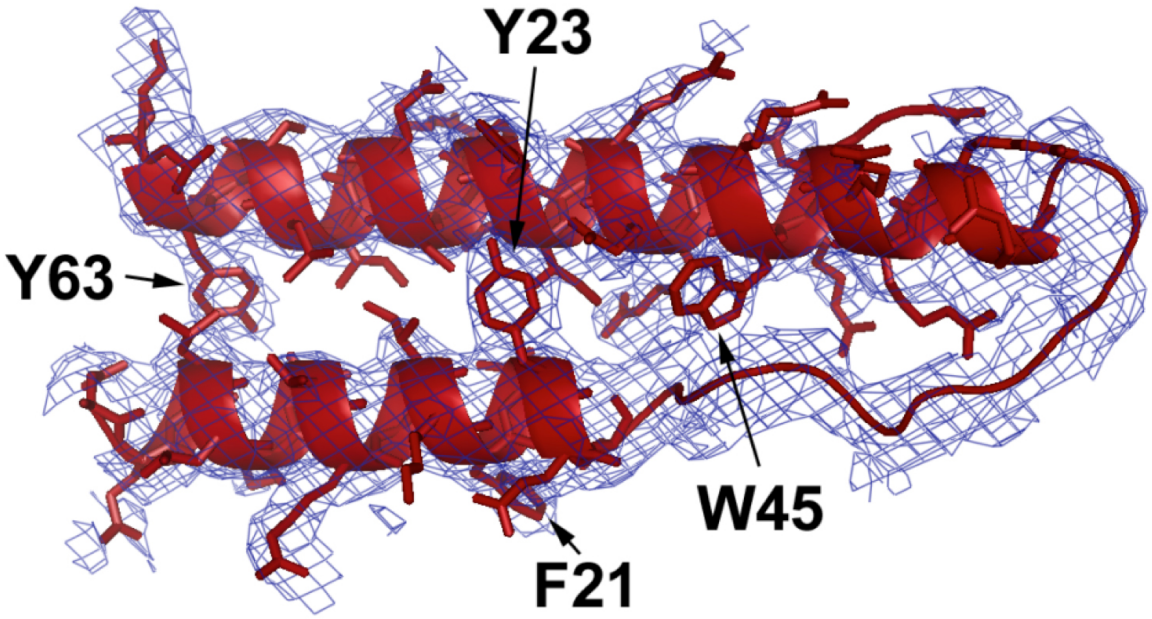


Figure 2.5: Omit map of HIV Rev at 3.2 Å resolution. Rev omit map shows unbiased density into which the Rev NTD polypeptide was manually built. Strong electron density corresponding to the side chains of bulky aromatic residues (F21, Y23, W45, Y63) aided initial structure building. While the full-length Rev polypeptide was present in the crystal (residues 1-116), density was observed only for the NTD. The CTD is either disordered or conformationally heterogeneous.

the quality of the structure refinement process. Refinement statistics are given in **Table 2.1**, and final refined coordinates and structure factors have been deposited with the PDB with accession code 2X7L.

The Fab-Rev structure was used as an initial model for scFv-Rev structure solution and refinement proceeded as was performed with the Fab-Rev data. Water molecules were modelled in the highest resolution structure (2.3 Å - space group $P2_1$; see **Table 2.2**). The lowest resolution structure (4.3 Å - space group $P32_12$) could not be refined (including via target restraints to a high resolution structure) due the limited number of structure factors available for the given parameters requiring refinement (see **Table 2.2**).

Table 2.1: Fab-Rev X-ray diffraction data collection and refinement statistics

Data collection	
Space group	*P*1
Crystal asymmetric unit	3 Rev dimers and 6 Fab molecules
Unit cell dimensions	
a, b, c (Å)	87.7, 87.7, 176.3
$\alpha, \beta, \gamma(^\circ)$	94.9, 95.5, 104.6
Resolution (Å) *	48.8–3.2 (3.3–3.2)
R_{sym} (%)	11.5 (100)
$I/\sigma I$	9.9 (1.1)
Completeness (%)	99.0 (98.8)
Redundancy	3.4 (3.4)
Refinement	
Resolution (Å)	48.8–3.2
No. reflections	83,636
R_{work} / R_{free}	23.5 / 25.0
No. protein atoms	22,554
Average protein B-factor (Å^2)	144
Root mean square deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.3
B-factor of bonded atoms (Å^2)	18
NCS related Rev coords (Å)	0.4
NCS related Rev B-factors (Å^2)	14
NCS related Fab coords (Å)	0.2
NCS related Fab B-factors (Å^2)	33

***Highest resolution shell is shown in parentheses.**
5% of the total reflections were excluded from data derived from a single crystal. These R-factors are derived from Buster (Blanc et al., 2004), taking account of disordered unmodelled portions of the structure.

Table 2.2: scFv-Rev X-ray diffraction data collection and refinement statistics

Data Collection				
Space group	$P2_1$	$P2_12_12_1$	$I4$	$P3_212$
Beamline[a]	Diamond I24	Diamond I02	Diamond I02	ESRF ID23-EH2
Unit cell dimensions				
a, b, c (Å)	43.9, 81.9, 98.8	44.5, 87.1, 166.7	121.1, 121.1, 89.3	48.5, 48.5, 264.5
α, β, γ(°)	90.0, 101.1, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å)	48.5-2.30 (2.42-2.30)	50.0-2.90 (2.95-2.90)	60.8-3.10 (3.20-3.10)	50.0-4.30 (4.37-4.30)
Wavelength (Å)	0.9779	0.97930	0.97930	0.87260
Completeness (%)	100.0 (100.0)	99.5 (99.2)	99.8 (97.6)	98.3 (98.6)
R_{sym} (%)	10.3 (84.6)	20.1 (100.0)	15.1 (90.1)	5.5 (71.7)
$I/\sigma(I)$	11.5 (1.6)	10.3 (1.1)	15.1 (2.7)	20.4 (1.1)
Mean multiplicity	8.1 (5.0)	13.3 (13.3)	7.5 (7.6)	3.1 (2.8)
Refinement				
No. of reflections used in refinement	30,611	14,917	11,655	2,588
R_{work} / R_{free} (%)	18.9 / 23.0	24.6 / 27.3	29.1 / 30.0	- / -
Reflections used for R_{free} ($\|F_o\| > 0$) (%)	5.0	5.0	5.0	5.0
Ramachandran Plot (%)				
Most favored	96.9	92.6	95.7	-
Allowed	3.1	6.9	3.6	-
Outliers	0.0	0.5	0.7	-
No. of protein / water atoms	4344 / 180	4381 / 0	4444 / 0	2157 / 0
Average B-factor (Å^2)	58	70	72	-
Wilson B-factor (Å^2)	50	81	80	-
RMSDs				
Bond lengths (Å)	0.010	0.010	0.008	-
Bond angles (°)	1.15	1.21	1.14	-

[a]Data were collected at beamlines Diamond I02 and I24 at the Diamond Light Source (Didcot, UK) and at Beamline ID23-EH2 at the ESRF (Grenoble, France).
Ramachandran plots were calculated with Molprobity (Davis et al., 2007).

2.1.5 Fab-Rev and scFv-Rev structural analysis

The PISA Interface Web server was utilised for buried surface area and interacting residue analysis of the Fab-Rev epitope, the Rev-Rev *A-A* dimer interface, and the residues contributing to structural integrity of the Rev monomer (Krissinel and Henrick, 2007). The inter-axial dimer angle between two Rev subunits was determined manually by calculating the dot product of the vectors parallel to $\alpha 2$ of each dimer subunit. Molecular graphics were produced using Chimera (Pettersen et al., 2004) and Pymol (DeLano Scientific LLC).

2.1.6 Biophysical analysis of Rev L64A

2.1.6.1 Analytical ultracentrifugation

Analytical Ultracentrifugation (AUC) was carried out using a Beckman (Palo Alto, CA) Optima XL-I AUC operating in velocity mode. Data were collected on Rev WT and Rev L64A using 280 nm absorbance optics at a speed of 40,000 rpm with one scan taken every 6 min for a total of 60 scans. The sedimenting species were analysed using the c(s) method with the SedFit software package (Brown and Schuck, 2006). All samples were in 100 mM sodium phosphate pH 6.5, 150 mM NaCl, 100 mM potassium sulfate, 1 mM DTT, 1 mM EDTA.

2.1.6.2 Negative-stain electron microscopy

Samples of Rev WT and Rev L64A were applied to glow-discharged, poly-lysine coated carbon grids at a concentration of ~0.2 mg mL^{-1}, stained with 1% uranyl acetate and observed at x35,000 magnification in a Philips CM-120 electron microscope.

2.1.6.3 Dynamic light scattering

Dynamic light scattering measurements were conducted on a Protein Solutions DynaPro instrument at 22 ÂřC at 0.2 mg mL^{-1} in 100 mM sodium phosphate pH 6.5, 150 mM NaCl, 100 mM potassium sulfate, 1 mM DTT, 1 mM EDTA. Sample was centrifuged at 13,000 rpm for 5 min before being placed in the cuvette.

2.2 Cryo-electron microscopy of HIV-1 Rev filaments

Structural information on the filament assemblies that HIV Rev forms *in vitro* remains limited. The seminal study on the matter (Watts et al., 1998) characterised the structure of the tubes using the Fourier-Bessel reconstruction method, complemented by various biophysical techniques. The Fourier-Bessel inversion method involves indexing the layer lines of the computed diffraction patterns from the micrograph images. This approach works quite well when a helical polymer is highly ordered over long distances, is rigid and straight, and has no Bessel overlap (more than one Bessel function present on any one layer line). Rev filaments do not conform to such ideality, offering weak layer lines on diffraction patterns and often having a significant degree of curvature, as well as more recent obversations of fluctuating diameters, even within a given filament (see **Fig 4.4**). Therefore, the approach I took here was to treat segments of the filaments as single particles (i.e., as in single-particle analysis), while taking advantage of the rotational and helical symmetries as powerful constraints.

2.2.1 Data collection

5 µL drops of a suspension of Rev filaments were applied to holey carbon grids, blotted to a thin film, and vitrified by plunge-freezing in liquid ethane. Specimens were observed with a CM200-FEG electron microscope (operating at a nominal magnification of x38,000 with defocus values in the range of -1.5 to 2.7 µm (as determined after CTF fit). A data set of 10 focal pair micrographs was recorded on film and digitisation was done at a sampling rate of 1.84 Å pixel^{-1}.

2.2.2 Filament segmentation

The contrast transfer function (CTF) was partially corrected for each micrograph by phase-flipping (Fig 2.6). For averaging experiments, Rev filaments that appeared relatively straight (arc of curvature visually negligible over box length) were selected and cut into segments using a box size of 256 x 256 pixels. Segments were selected with 90% overlap between adjacent boxes in order to aid particle alignment in further processing steps. In total, 30,426 overlapping particles were selected from all of the micrographs.

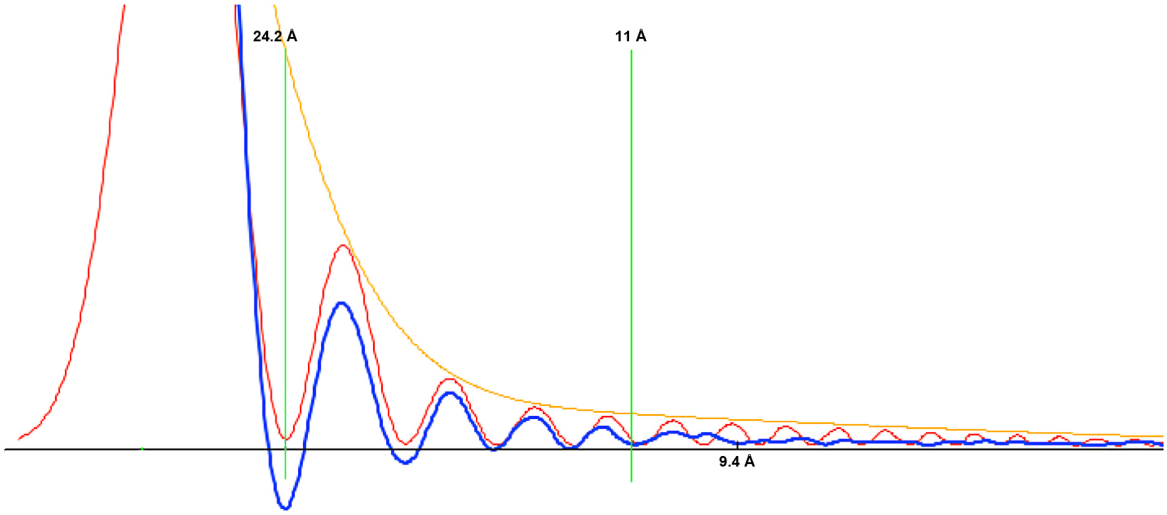


Figure 2.6: Example power spectrum of Rev filament EM micrograph. The power spectrum of a recorded cryo-EM micrograph of Rev filaments is calculated by first tiling the micrograph into sub-images (of size 512 by 512 pixels). The Fourier transform of these sub-images is then computed, squared, and averaged. The resulting average spectrum is plotted above, resolution (x-axis) vs. amplitude (y-axis). The red curve represents the idealized diffraction pattern, which is fitted to the blue (experimental) curve by optimising the parameters of the Contrast Transfer Function (CTF). The CTF is the function that modulates the amplitudes and phases of the electron diffraction pattern formed in the back focal plane of the objective lens. The resolution of the micrograph can be estimated by observing where the Thon rings (local maxima) fall to the background noise level, here, approximately 11-12 Å resolution (second green line). The first zero is at ~24 Å (first green line). Without CTF-correction, information beyond the first zero is not recovered. Here, the phases have been flipped according to the CTF-corrected (blue) curve.

2.2.3 Classification

The Rev filaments appear qualitatively heterogeneous with varying diameter, even within a given tube (see **Fig 4.4**). Principal component analysis (PCA) was performed for classification of the aligned images using SPIDER. PCA is a mathematical procedure that can reduce the large dimensionality of data space, as direct clustering on images with thousands of pixels is computationally prohibitive; therefore, the problem must be simplified before clustering. In other words, a set of observations of *possibly* correlated values (set of filament segments) are reduced to a set of values of uncorrelated variables called principal components (PCs). In analysis of cryo-EM particles, it is often used to obtain an immediate inventory of the different components of variability within the data, e.g. presence or absence of a domain, rocking of a molecule, or conformational variability. These components of variability are embedded into the PCs: the first PC has the greatest inter-image variance possible, that is, it accounts for as much of the variance in the data as possible. Each succeeding PC in turn has the highest variance possible given the constraint that it be uncorrelated with (orthogonal to) the preceding components. The PCs, or eigenimages, generated for the Rev filament data set are shown in **Fig 2.7**. The data are ultimately clustered over a variety of scales and is typically presented as a hierarchical tree or dendrogram. This tree does not represent a single set of clusters, but rather a multilevel hierarchy where clusters at one level are joined into a larger cluster at the next level. For the purposes of Rev filament classification, the threshold was selected to give the largest number of clusters possible before reaching the noise level.

The final classification result by PCA (i.e. the number and features of each class average) is sensitive to the choices made during the analysis, e.g. which PCs are selected for use and which dendrogram threshold is selected for clustering during hierarchical classification. This is a known consequence,

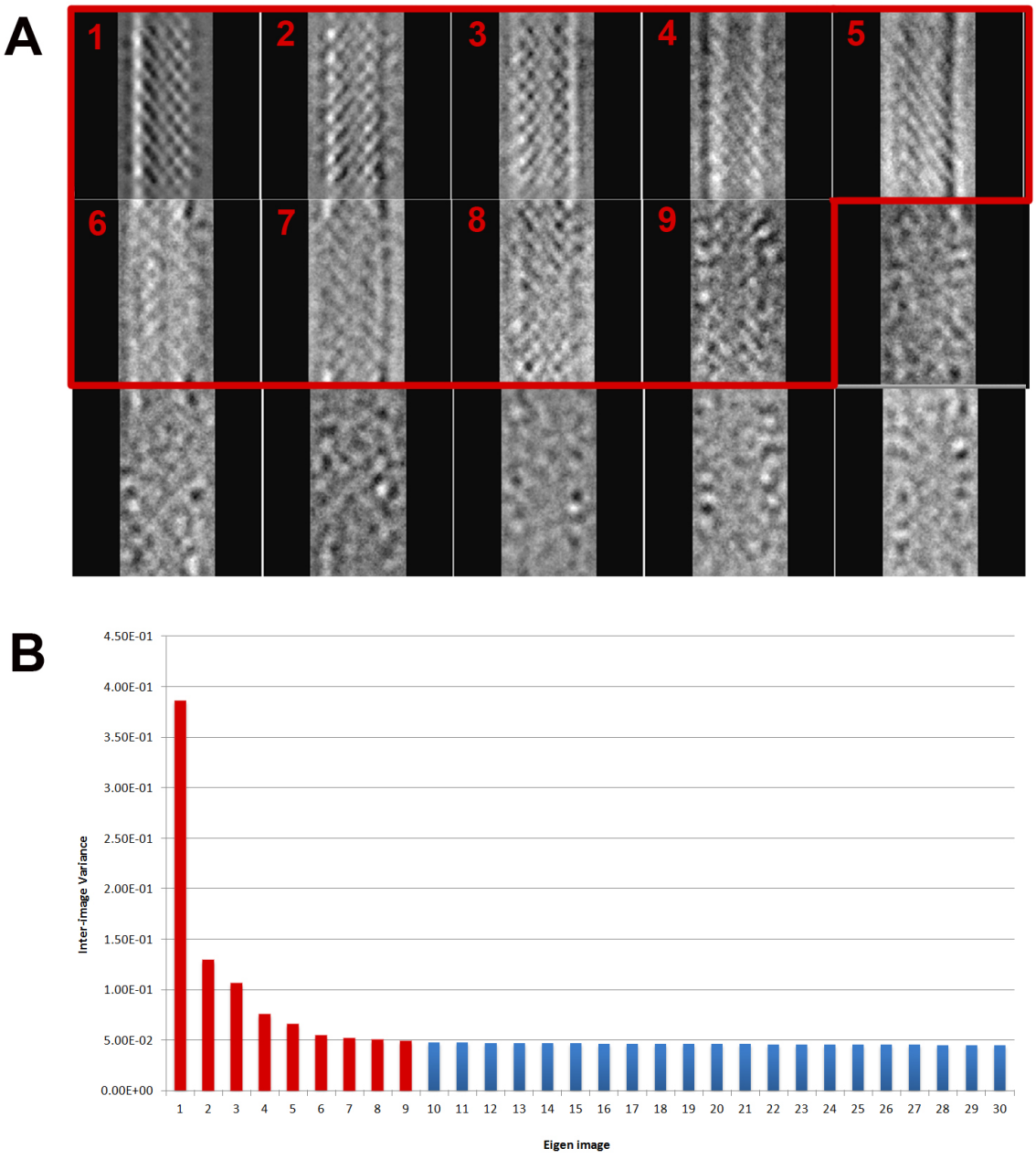


Figure 2.7: Rev filament principal components Principal components of Rev filaments represent the components of variability within the data. The first eigenimage accounts for the greatest inter-image variance and each successive represents the highest variance possible given that it is uncorrelated with the preceding components. The first nine PCs, shown in red, were used for classification of Rev filaments.

but is unhelpful when attempting to determine which size filaments represent real phenomena versus artifacts of the classification scheme.

Consequently, final PCA runs were performed in which the particles were over-classified by using the top 9 PCs (representing virtually all of the inter-image variance within the data) and a liberal histogram threshold for hierarchical classification. The aim of such "over-classification" was to separate the particles to such a degree that the probability of homogeneity within a bin was as high as possible. All in all, 26 bins were generated. The observation of several bins containing less than 100 particles confirmed that the particles were separated to their limit within the context of PCA.

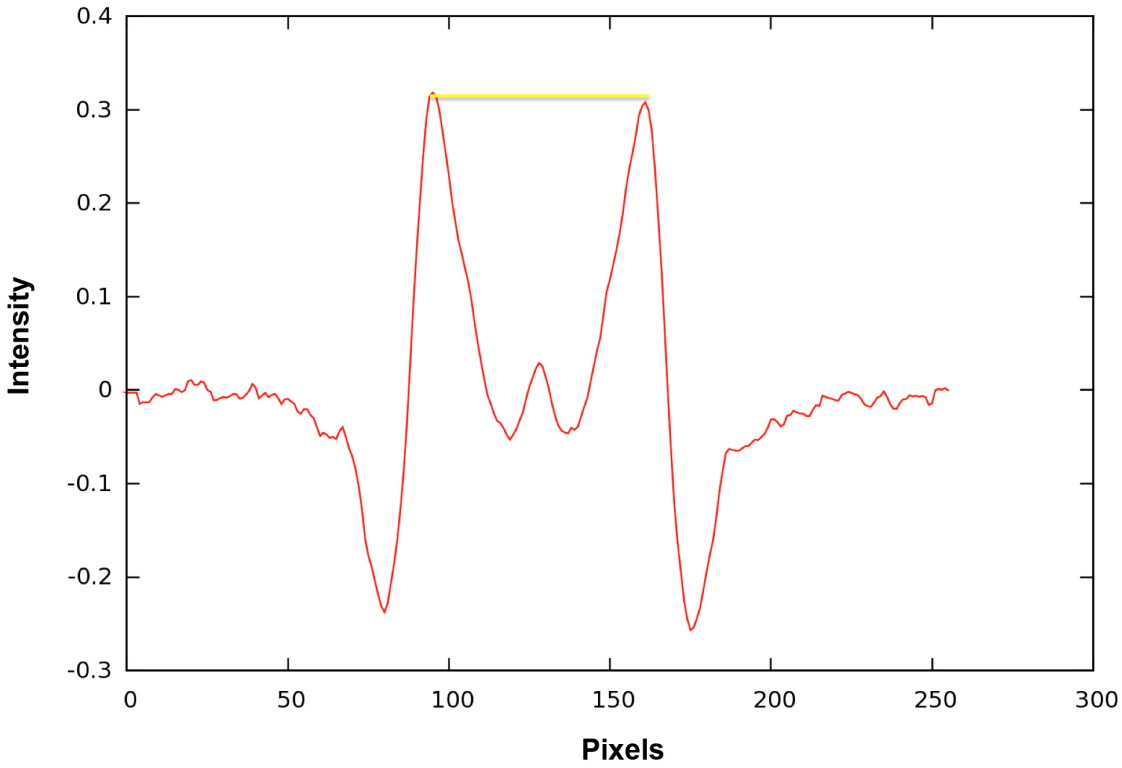


Figure 2.8: Example transverse density plot of Rev filament class average. Shown is an example one-dimensional transverse density plot of a class average of Rev filament segments. Peak-to-peak diameter was measured (yellow bar) and used as one of several criteria for classifying the different Rev filaments into distinct species.

These 26 bins were characterised in order to assess which could be recombined to form larger,

homogeneous classes. The goal of this separation-recombination is to learn about the full breadth of conformational variability within Rev filaments, while minimising the final number of classes to those that represent unique conformational states. Diameter was the first measure used for characterisation of each bin. A two-dimensional class average of each bin was calculated and the diameter of each class was estimated by projection of the average along the y-axis, resulting in a transverse density plot (Fig 2.8). The Fourier transform of each class average was also calculated and the apparent layer lines were analysed. However, the visible layer lines were insufficient for direct determination of the helical symmetry of each class (not shown). Helical parameters were the second measure used for characterisation of the bins. Accordingly, the particles of each class were aligned and reconstructed (as described below) in order to obtain the helical parameters to which each volume converges. Classes containing similar diameter and helical parameter values were recombined into one of five Classes (I-V).

2.2.4 Iterative helical real-space reconstruction

The flexibility and heterogeneous width of Rev filaments hamper structural analysis by Fourier methods. Accordingly, a "single-particle" approach was employed wherein the helical symmetry of the filament is used to guide the structural analysis to convergence to the correct solution (IHRSR: iterative helical real space reconstruction) (Egelman, 2007).

BSOFT subroutines were used to implement the IHRSR method (Heymann and Belnap, 2007). In this process, a starting model is needed (although a solid cylinder can be used) of which reference projections are generated by rotating the starting model azimuthally around the filament axis and projecting every 4°. The raw images are then aligned (translation and rotation) and cross-correlated with

the set of reference projections using the *borient* and *brefine* operations in BSOFT. A 3D reconstruction is then generated by back-projection using *breconstruct*. The helical symmetry of this volume is then determined using two different methods in parallel: one is a real-space least squares fitting algorithm (Egelman, 2000) and the other is a reciprocal-space cross-correlation algorithm (*bhelix*) (Heymann and Belnap, 2007). The determined helical parameters are then imposed on the reconstruction and this newly symmetrised volume is then used as the reference map for the next round of orientation search and alignment.

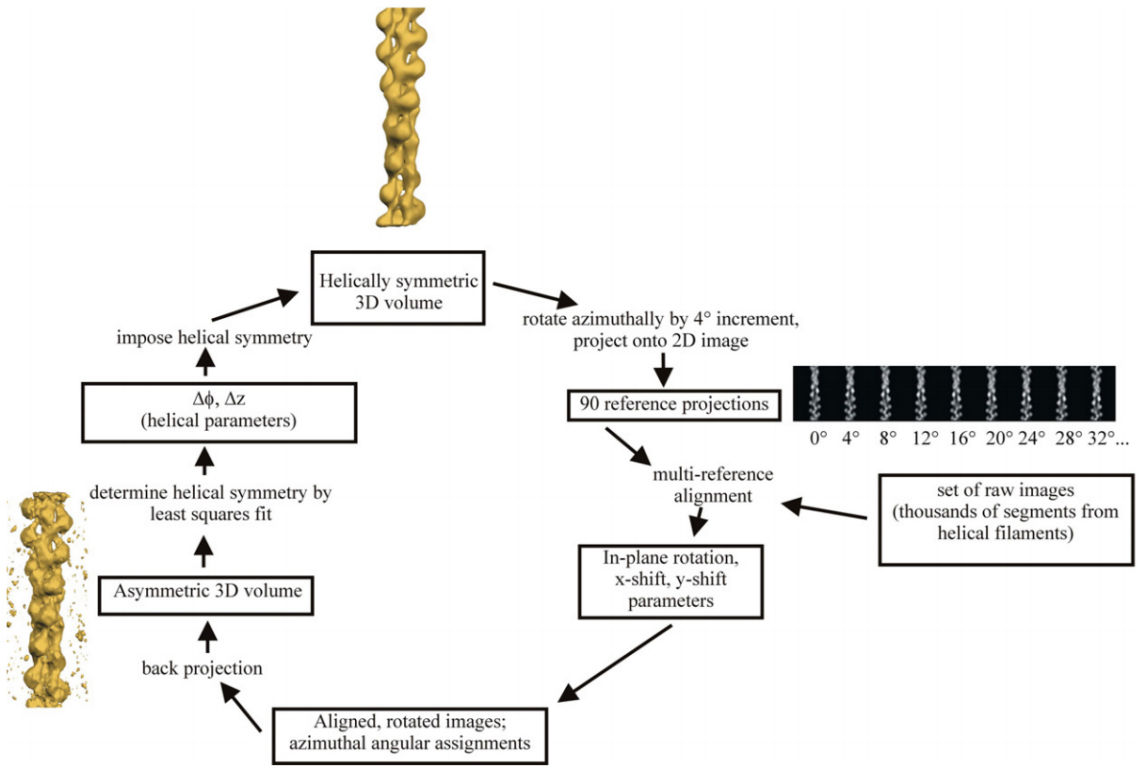


Figure 2.9: Iterative helical real space reconstruction method. The IHRSR method is an approach to the structure determination of helical assemblies by cryo-EM single particle analysis. The technique works well in situations where EM of the assembly shows heterogeneity and disorder, as well as poor diffraction in reciprocal space. The technique differs from a classical single particle approach in that once the aligned images are back-projected into a 3D volume, a real-space search of the volume is done for the helical symmetry. The helical parameters are then imposed onto the volume and the cycle repeats using this volume as the new reference. Figure adapted from Egelman (2007).

2.2.5 Estimation of helical symmetry

A limitation of the IHRSR method is that the starting helical parameters must be approximately known (reasonably close to the actual values). In order to estimate the helical parameters for Rev filaments, the particles belonging to Class I were successively aligned and reconstructed without imposition of any helical symmetry. Rotational (C6) symmetry was imposed, as it has been previously determined that Rev filaments are a six-start helix (Watts et al., 1998). A cylinder was used for the initial model and multiple runs were initiated with randomly selected subsets of particles (>500). After approximately 30 rounds of reconstruction, one of the runs gave rise to an electron density map with sufficient features for direct estimation of the helical symmetry. The *bcylinder* program in BSOFT was used to unwrap the map into its cylindrical projection and perform an auto-correlation search for peaks corresponding to the helical register of the map (**Fig 2.11**) (Heymann and Belnap, 2007). For the Class I filament, the initial subunit rise and rotation were estimated to be 27.6 Å and 30.0°, respectively. These helical parameters were sufficiently close for convergence to a unique solution. This map, with known helical symmetry, was subsequently used as a starting model for particle alignment and reconstruction of the other size classes.

2.2.6 Molecular docking

UCSF Chimera was used to dock the Rev dimer crystal structures (as presented in (see **Chapter 3**) into the different Rev filament reconstructions (Class I-V) (Pettersen et al., 2004). A Rev dimer structure (from scFv-Rev crystal structure $P2_1$ - crossing angle of 115°) was placed into the map at the globular density cluster thought to be successive Rev subunits. A local optimisation algorithm was performed using Chimera with the "Fit in Map" command, which adjusts the model position in real-space to

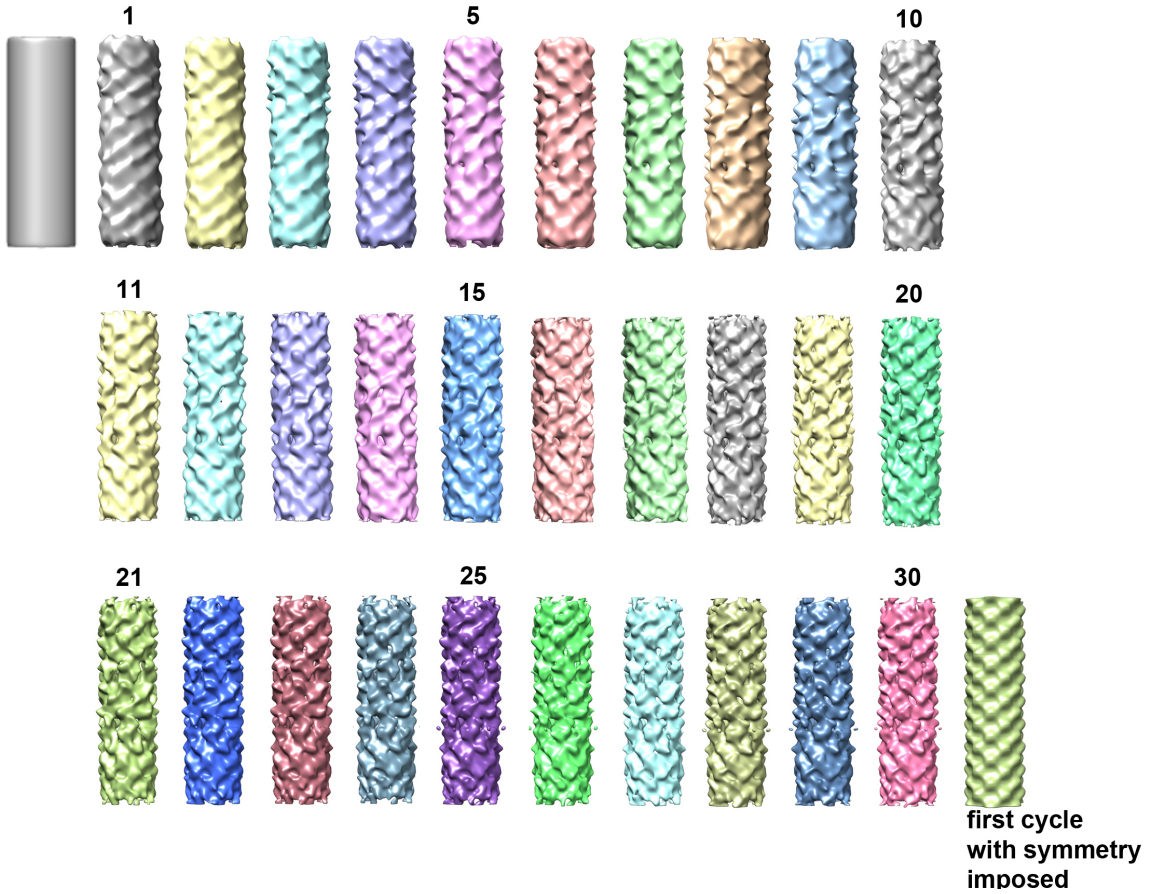


Figure 2.10: Generation of initial model. Rev filament segments belonging to Class I were aligned and reconstructed without imposition of helical symmetry. After 30 iterations, enough features were present to directly detect the helical register of the map in real space.

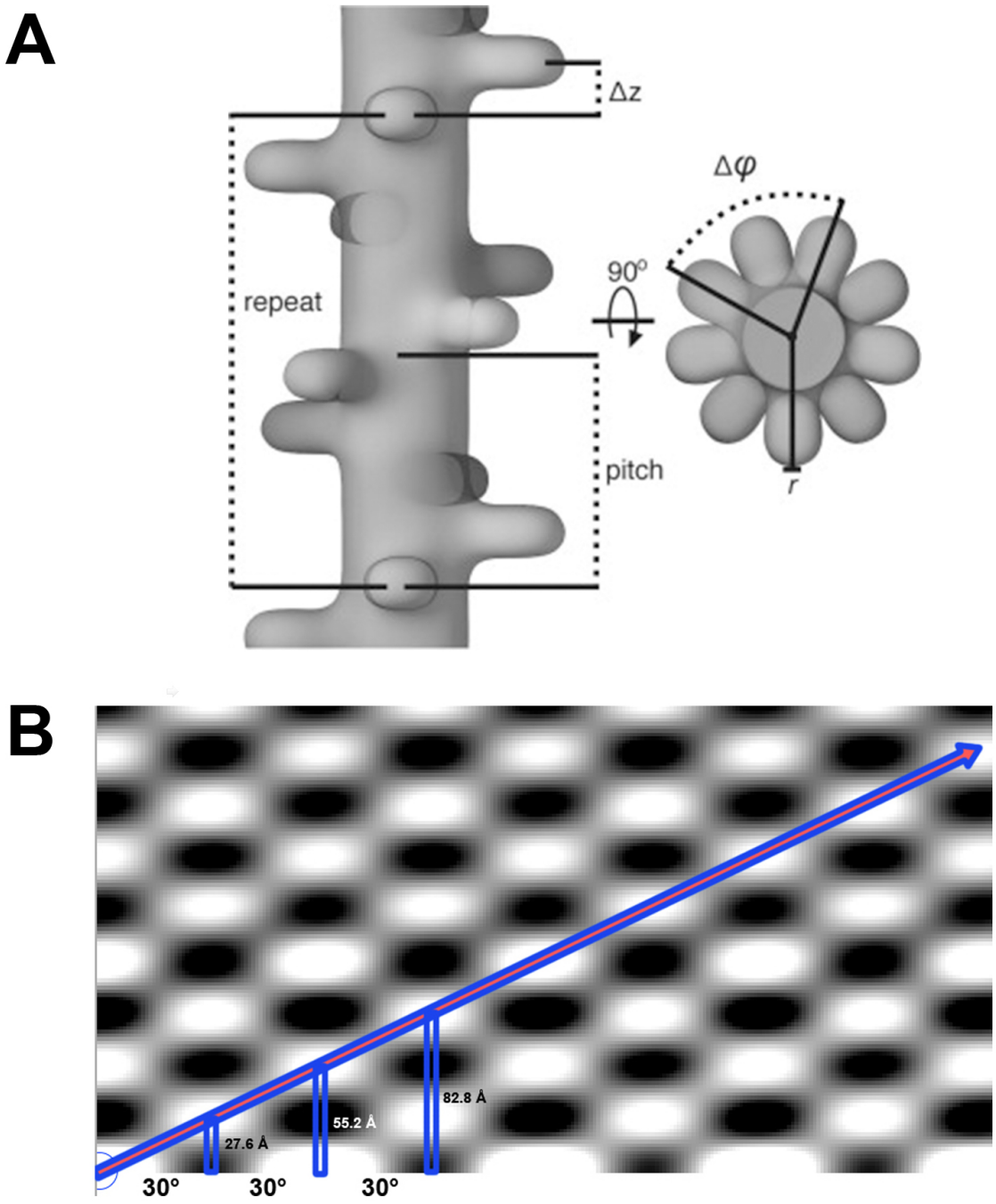


Figure 2.11: Auto-correlation search to determine helical symmetry.
(A) Helical geometry can be defined in terms of distance (Δ z; subunit rise) from one subunit to the next along the azimuth and rotation ($\Delta\varphi$; subunit rotation) from one subunit to the next around the azimuth. Figure adapted from Behrmann et al. (2012).
(B) Thirty cycles of single particle analysis of Rev filaments were performed without imposing helical symmetry (only C6 rotational symmetry imposed). The best possible such reconstruction was unwrapped along the filament axis into its cylindrical project. An auto-correlation search was then performed to locate self-peaks that may correspond to the helical register of the tube. Helical symmetry was apparent with subunit rise and rotation of 27.6 Å and 30.0°, respectively.

maximise the overlap of the atoms with the map. As an objective measure to determine the fit, a global real-space search with the Rev dimer structure was also performed in Chimera, producing 100 docking searches. All solutions that had the *A* and *B* surfaces facing the filament interior or exterior were discarded, given biological data supporting their involvement in Rev oligomerisation. Additionally, it is unlikely that the *A* and *B* surfaces are solvent-exposed due to their high hydrophobicity content. Only one set of orientations remained, which was consistent with the manual docking described above.

Once the initial orientation of the Rev dimer was determined, the rotational and helical symmetry of each map (C6; φ: 30°; z: 26.6-34.0 Å) was incorporated into the docking in order to aid fitting. Upon generation of symmetry-related Rev dimer subunits, the *A*, *B*, and *C* interfaces were in close register. Manual adjustment of the Rev dimer was done to optimise the dimer orientation and crossing angle of the subunits across the dimer, in order to bring the three oligomerisation interfaces into proper register.

Difference maps were calculated by subtracting the calculated structures corresponding to the fitted NTD model from those of the experimental cryo-EM density map.

2.3 HBV e-antigen

2.3.1 Preparation of Fab-HBeAg complex

The construct used for recombinant expression of HBeAg contained the 10 residues of the propeptide and the full assembly domain, or Cp(-10)149. The construct differs from wildtype in that it contained two Ala substitutions at C48 and C107, residues that do not form disulfide bridges in HBcAg or HBeAg given their position near the interior of the folded protein (Wingfield et al., 1995). The HBeAg used for crystallization included the aforementioned mutations plus G123A (Watts et al., 2011). In brief, the expressed protein was extracted from *E. coli* with 3 M urea, 100 mM sodium bicarbonate, pH 9.6 and subjected to gel filtration on a Superdex 200 column in the same buffer plus 2 M urea. The HBeAg pool was further fractionated by anion exchange chromatography on Sepharose Q in 50 mM Tris-HCl, pH 8.0 (0-0.5 M NaCl gradient). Finally, the protein was re-chromatographed on Superdex 200, pH 9.6 buffer (minus urea). This material was dimeric (~36000 Da) as determined by sedimentation equilibrium analysis of analytical ultracentrifugation data. Analyzed by SDS-PAGE, the protein was monomeric with a lower apparent molecular weight under non-reducing versus reducing conditions, consistent with the presence of an intra-molecular disulfide C(-7)-C61 (Watts et al., 2011). Fab e6 and the HBeAg-Fab e6 complex were produced as described in Watts et al. (2010). HBeAg dimers and e6 Fab molecules were mixed at an expected stoichiometry of 1:2, and excess Fab was removed by size-exclusion chromatography.

2.3.2 Sequencing of e6 Fab

The protein sequence of e6 variable domain regions (heavy: V_H and light: V_L) were determined using RT PCR (Qiagen) to sequence the mRNA encoding the mouse e6 mAb. Using previously reported primers (Morrison et al., 2001), the coding sequence for the V_H region of Fab e6 was ascertained, but not that of the V_L region. The amino acid sequence of the N-terminus of V_L was determined by Alphalyse Inc. (Palo Alto, CA) to be NIMMTQSP and an RT PCR primer for the 5' end of the V_L coding region was synthesized based on this sequence, which fully and exclusively matched mouse V gene IGKV8-27 (www.imgt.fr). A second oligonucleotide primer was subsequently synthesized that matched the 5' end of the secretion sequence associated with gene IGKV8-27. Both of these oligonucleotides primed DNA synthesis from the mouse mRNA preparation and enabled sequencing of the entire V_L region.

2.3.3 Crystallization and X-ray data collection of Fab-HBeAg complex

Crystallization trials were performed using the sitting drop vapor diffusion method at a protein concentration of 5.3 mg mL^{-1} in 20 mM HEPES pH 8.0 at 21 °C. Sitting drops were formed by mixing 100 nL of protein solution and 100 nL of reservoir solution (Walter et al., 2005). Plate-shaped crystals (dimensions ~70 µm x 80 µm x 30 µm) appeared in 5 days and grew to full size within 10 days with reservoir solution containing 20% PEG 6000 and 100 mM bicine pH 9.0. Rod-like crystals of Fab e6 (dimensions ~230 µm x 40 µm x 40 µm) appeared in ~3 months using reservoir solution containing 20% PEG 3350, 100 µM bis-Tris propane pH 6.5, and 200 µM potassium thiocyanate. Fab e6 crystals serendipitously grew from a preparation of the complex from which excess Fab had not been removed after complex formation. Both HBeAg-Fab e6 and Fab e6 crystals were cryo-protected in 25% glyc-

erol, frozen in liquid nitrogen, and screened using synchrotron radiation at the Diamond Light Source, Beamline I24, Didcot, UK. Diffraction data for the complex and Fab e6 were collected from single crystals to a minimum Bragg spacing of 3.3 Å and 2.5 Å resolution, respectively. Diffraction data were integrated and scaled using xia2 (Winter, 2010).

2.3.4 HBV e-antigen structure determination

2.3.4.1 Structure solution

Initial phase information for the Fab e6 data were obtained via an automated Phaser (McCoy et al., 2007) molecular replacement (MR) search using a library of 334 Fab structures as search models. The Fab library was designed to include models that represent the full range of Fab elbow angles observed in the PDB (in 1-2° increments) (Stanfield et al., 2006). The rationale for this design is that subtle differences in this angle can dramatically change the low-resolution shape of a Fab molecule, and is arguably more important in a six-dimensional MR model search than the Fab sequence itself. The correct solution was found using the structure of a monoclonal fab fragment raised against HIV-1 protease (PDB ID 2HRP), which shares 77% sequence identity with Fab e6. The space group is C2 with 4 molecules in the asymmetric unit.

2.3.4.2 Refinement

The MR solution was rigid body-refined with each Ig domain treated separately, hyper-variable regions deleted, and backbone converted to poly-alanine. Positional, TLS, and individual isotropic B-factor refinement on the Fab e6 structure were carried out in BUSTER (Blanc et al., 2004) using fourfold NCS restraints, followed by iterative rebuilding of the hyper-variable loops and correction of the se-

quence according to the separately sequenced Fab e6 light and heavy chains. This Fab e6 structure, refined against 2.5 Å resolution data, was used as an MR model to obtain initial phase estimates for the HBeAg-Fab e6 data. Two Fab e6 molecules were placed in the $P1$ unit cell and refinement used local structure similarity restraints (LSSR) with the 2.5 Å resolution Fab structure as a target function (Smart et al., 2012). The HBeAg subunits were placed by first building three α-helices into the tube-like density regions in the map adjacent to each Fab e6. An HBcAg monomer (from PDB ID 1QGT) was unambiguously superposed onto each α-helical framework, which provided the initial orientation of the dimer subunits. Additional electron density was clearly observed extending from C61 to near the N-terminus of the HBcAg model. Positional, group B-factor (one group per residue), TLS refinement were used for structural refinement with twofold NCS-restraints and local structure similarity restraints (LSSR) were applied to the Fab structure to mitigate the limited resolution of the data. The Molprobity server (Davis et al., 2007) and validations tools in Coot (Emsley and Cowtan, 2004) informed the quality of the structure refinement process. While the resolution is limited, the key features of HBeAg are firmly established (**Fig 2.12**). Refinement statistics are given in **Table 2.3**, and final refined coordinates and structure factors have been deposited for Fab e6 and HbeAg-Fab e6 with the PDB with accession codes 3V6F and 3V6Z, respectively. The crystal packing depicts how the Fab fragment facilitated crystallization (**Fig 2.13**).

2.3.5 HBeAg structural analysis

The PISA interface web server was utilized for buried surface area and interacting residue analysis of the HBeAg dimer interface and the HBeAg-Fab e6 epitope-paratope interface (Krissinel and Henrick, 2007). The inter-axial angle between two HBeAg subunits was determined manually by calculating the

Table 2.3: Fab-HBeAg X-ray diffraction data collection and refinement statistics

Crystal Data	**HBeAg-e6 Fab Complex**	**Fab e6**
Space group	Triclinic *P*1	Monoclinic *C*2
Unit cell dimensions		
a, b, c (Å)	66.7, 75.8, 88.7	124.1, 68.2, 236.6
α, β, γ(°)	96.8, 103.8, 116.0	90.0, 96.3, 90.0
HBeAg dimer molecules per ASU	1	0
e6 Fab molecules per ASU	2	4
V_M (Å^3 Da^{-1})	2.92	2.60
Solvent content (%)	57.9	52.7
Data Collection		
Resolution[a] (Å)	46.2-3.34 (3.43-3.34)	31.1-2.52 (2.58-2.52)
R_{sym} (%)	11.6 (73.7)	9.6 (91.1)
$I/\sigma(I)$	6.4 (1.0)	7.7 (1.3)
Completeness (%)	92.6 (88.8)	99.5 (99.6)
Mean multiplicity	1.8 (1.6)	3.4 (3.5)
No. of total reflections	34,840	223,579
No. of unique reflections	19,664	66,469
Refinement		
No. of reflections used in refinement	19,651	66,468
R_{work} / R_{free}[b] (%)	23.3/23.4	18.1/21.9
Reflections used for R_{free} ($\|F_o\| > 0$) (%)	5.0	5.0
Ramachandran Plot (%)		
Most favored	96.0	96.9
Allowed	3.7	3.0
Outliers	0.3	0.1
No. of protein atoms	9,076	13,480
Average B-factor (Å^2)	111	62
Wilson B-factor[c] (Å^2)	76	58
RMSDs		
Bond lengths (Å)	0.008	0.010
Bond angles (°)	1.19	1.27
B-factor of bonded atoms (Å^2)	0.8	1.7
NCS related HBeAg coordinates (Å)	0.03	-
NCS related HBeAg B-factors (Å^2)	6.6	-
NCS related fab e6 coordinates (Å)	0.09	0.64
NCS related Fab e6 B-factors (Å^2)	7.5	14.1
Targeted refine. Fab coords[d] (Å)	0.50	-
Targeted refine. Fab B-factors[d] (Å^2)	13.0	-

[a]Highest resolution shell is shown in parentheses.
[b]R_{work} is calculated taking account of disordered, unmodeled portions of the structure (Blanc et al., 2004). R_{free} is calculated identically to R_{work} but corresponds to 5% of data omitted from refinement for cross-validation. The e6 Fab and HBeAg-e6 Fab structures were each determined from data derived from a single crystal.
[c]The Wilson B factor is an initial measure of the isotropic overall B-factor, as determined from the slope of a linear fit to a plot of the average reflection intensity vs. resolution (the Wilson plot).
[d]RMSD between coordinates (or B-factors) of refined structure and high-resolution structure used as target function for LSSR refinement (Smart et al., 2012).

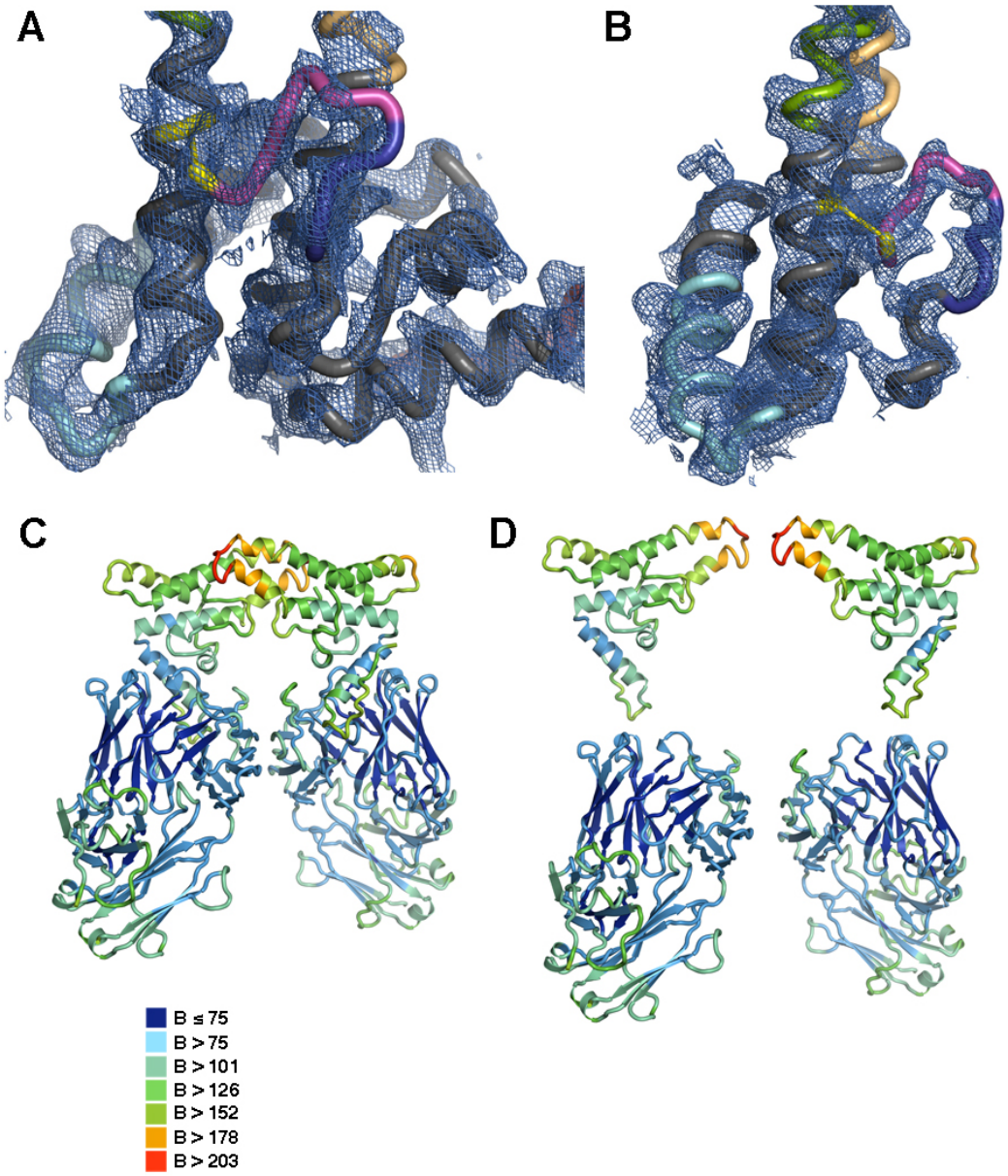


Figure 2.12: HBeAg structure validation. (A and B) Electron density map (2Fo-Fc) of HBeAg monomer. **A** depicts front view, as blue mesh and contoured at 1.0σ. **B** shows the same representation, but the HBeAg monomer is rotated 60° to show the electron density for α2 and the propeptide. **(C and D)** B-factor distribution of HBeAg-Fab e6 crystal structure asymmetric unit. The ribbon representation is colored by B-factor per residue, ranging from 75 (navy blue) to 203+ (red) (in $Å^2$). B-factors are a measure of atomic displacement of an atom from its mean position, which is affected by thermal vibrations and static disorder of atomic position in different unit cells. HBeAg has higher B-factors than the Fab, particularly at its spike apex.

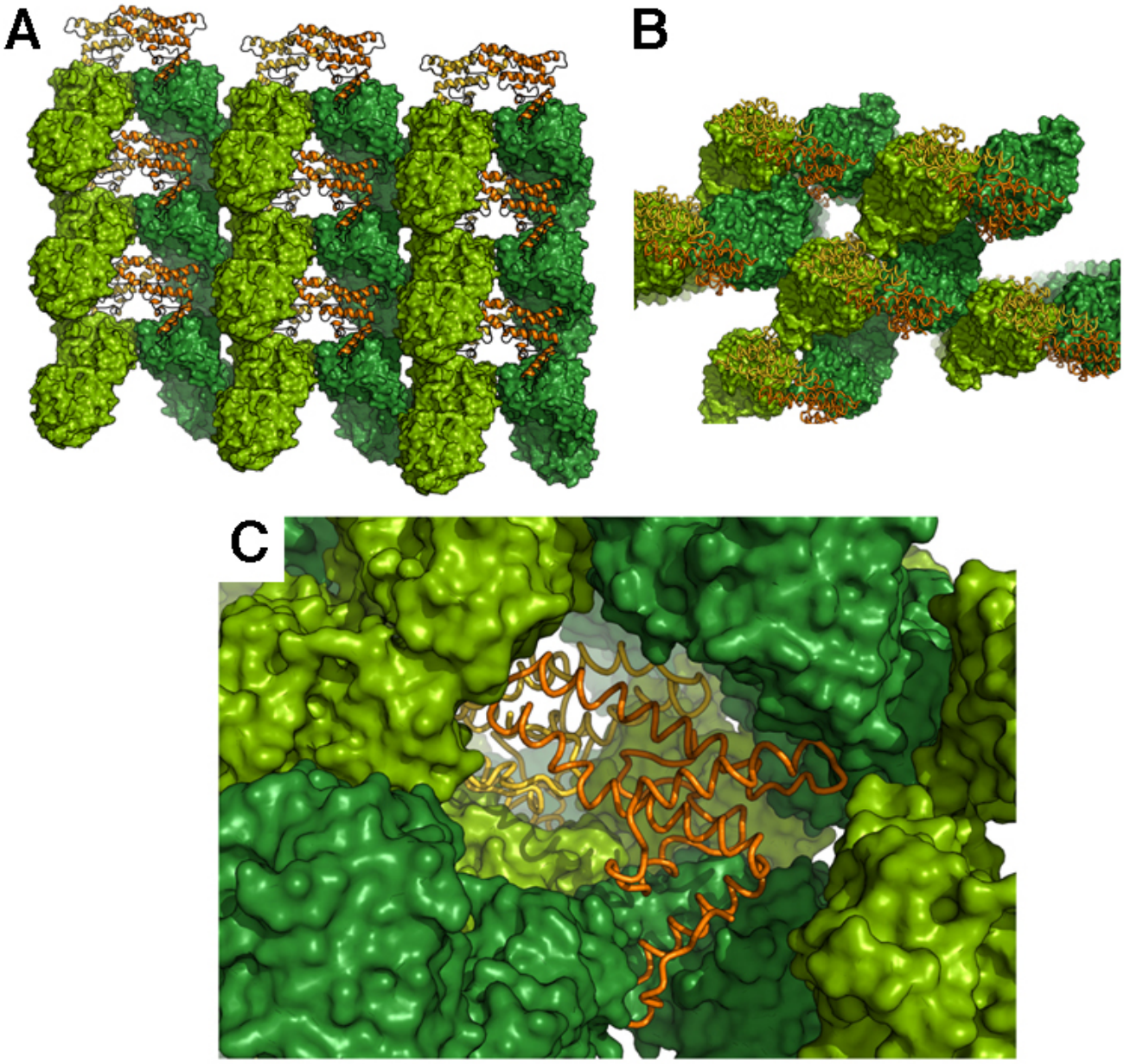


Figure 2.13: Crystalline lattice of HBeAg-Fab e6 complex. **(A)** HBeAg is depicted in ribbon (orange and yellow) and Fab e6 molecules are surface-rendered in shades of green. The lattice arrangement shows how the Fabs aided in crystallization of HBeAg, where previously the molecule had been prone to aggregation. All of the crystal contacts are mediated by the Fabs. **(B)** 90° rotated view of lattice from view in (A). **(C)** Close-up of HBeAg molecule (orange, yellow) surrounded by Fab molecules (green) in the crystal lattice.

dot product of the vectors parallel to the central hairpin of each subunit. The Rapido server was used to determine structurally similar sub-domains within HBeAg and HBcAg monomers (Mosca et al., 2008). A structurally similar region between the antigens was found and used for all superpositions between HBcAg and HBeAg (residues 5-34, 49-65, 93-124). Secondary structure assignment of HBeAg was done using the consensus of two programs: DSSP (Frishman and Argos, 1995) and Stride (Heinig and Frishman, 2004). Molecular graphics were produced using Pymol (Delano Scientific LLC).

2.3.6 Sedimentation velocity analysis and negative-stain electron microscopy

HBeAg was dialyzed against phosphate buffered saline (PBS) pH 7.2, plus 300 mM NaCl (total NaCl 450 mM). A sample treated with 10 mM DTT was also dialyzed against the same buffer, plus 2 mM DTT. Under either of these conditions, dimeric HBcAg (Cp149) readily and efficiently forms capsids. Following dialysis for ~24 h, samples were analyzed by sedimentation velocity. Measurement of the height (UV absorbance) of the sedimenting boundaries allows the concentrations of the various species to be determined. The oxidized and reduced HBeAg samples were also applied to glow-discharged, poly-lysine coated carbon grids at a concentration of ~0.25 mg mL^{-1}, stained with 1% uranyl acetate and observed at x35,000 magnification in a Philips CM-120 electron microscope.

Chapter 3

The Structure of HIV Rev

3.1 Fab-Rev crystal structure

Historically, structural investigation of HIV-1 Rev has been difficult due to the protein's strong propensity to form filaments or otherwise aggregate *in vitro*. In order to obtain crystals, Rev was complexed with a high-affinity Fab antibody fragment that depolymerizes Rev filaments into soluble, stoichiometric Fab-Rev complexes (**see Chapter 2**). Crystals of this complex were successfully grown and X-ray diffraction produced 3.2 Å resolution data. The structure was solved using molecular replacement, with Fab structures from the PDB serving as MR candidate models (**see Chapter 2**). The space group of the crystal was determined to be *P*1, with its unit cell containing six Fab fragments and three Rev dimers, arranged as heterotetramers, each comprising a Rev dimer flanked by two Fabs (**Fig 3.1**). The Fabs' variable domains provide most of the crystal contacts. The crystal structure of the Fab-Rev complex is the first high-resolution structure determination of HIV-1 Rev.

3.1.1 Structure of Rev monomer

The structure is α-helical, with the N-terminal domain (NTD; residues 9-65) (**Fig 3.2A**) adopting a helix-loop-helix motif (**Fig 3.2B**). Although full-length Rev was expressed, electron density for the C-terminal domain (CTD; residues 66-116) was not observed. In order to verify that proteolytic cleavage did not result in crystallisation of merely a fragment of the Rev polypeptide, SDS-PAGE was performed

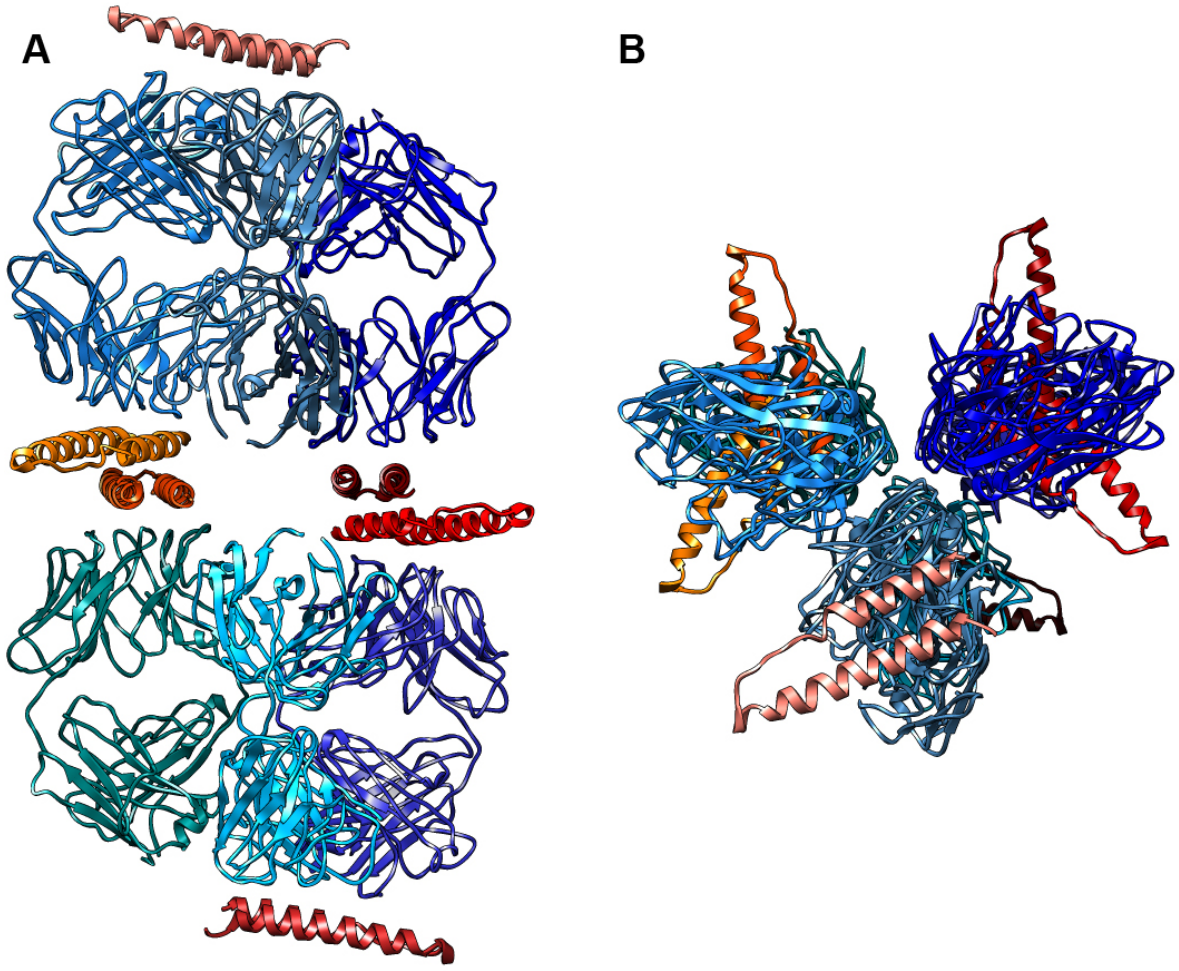


Figure 3.1: Fab-Rev crystal symmetric unit. **(A)** Ribbon structure of the Fab-Rev *P*1 unit cell, occupied by six Fab molecules (coloured in shades of blue) and six Rev molecules (coloured in shades of red) organized as molecular dimers. **(B)** Overhead view.

on a re-solubilized Fab-Rev crystal. The resulting gel showed unaltered Rev mobility for the crystal, relative to a control Rev sample, with contaminants at ~5% (**Fig 2.4**), indicating that the CTD is disordered in the context of the Rev-Fab crystal packing environment. This is consistent with other reports indicating structural disorder of the CTD (Auer et al., 1994; Daugherty et al., 2010a).

One prong of the helix-loop-helix motif (herein termed a helical hairpin) consists of an α-helix (α1; residues 9-24). This is followed by a loop region from residue 25 to 33 and a longer α-helix (α2; residues 34-65) that contributes the second prong (**Fig 3.2B**). The Rev NTD structure was found to be consistent with previous proposals suggesting the NTD forms a helix-loop-helix motif (Auer et al., 1994; Jain and Belasco, 2001; Blanco et al., 2001; Havlin et al., 2007). However, the α-helices of the hairpin form a planar structure, not a coiled-coil, different from a previously proposed model (Jain and Belasco, 2001). Helix α1 and the C-terminal half of α2 are amphipathic, with hydrophobic and polar surfaces on opposing sides of the helices. The N-terminal half of α2 is the RRE-binding arginine-rich motif (ARM) and is densely populated with Arg and Glu residues that interact with the RRE RNA (**Fig 3.2B**). Non-polar residues cluster around the apposed surfaces of α1 and α2 to form a hydrophobic core that serves at least two purposes:

1. stabilization of the hairpin;
2. and presentation of hydrophobic surface patches, one on each face of the hairpin (**Fig 3.2B, C**) that mediate interaction with other Rev monomers.

The residues within the core identified by buried surface area analysis as major contributors to stabilizing the helical hairpin are L12, I19, L22, Y23, W45, I52, I59, L60, and Y63. Of these, four—I19, L22, I52, and I59—have been proposed on other grounds to contribute to an intra-molecular interface (Jain and Belasco, 2001). Mutations at these positions reduce the affinity of Rev monomers for stem

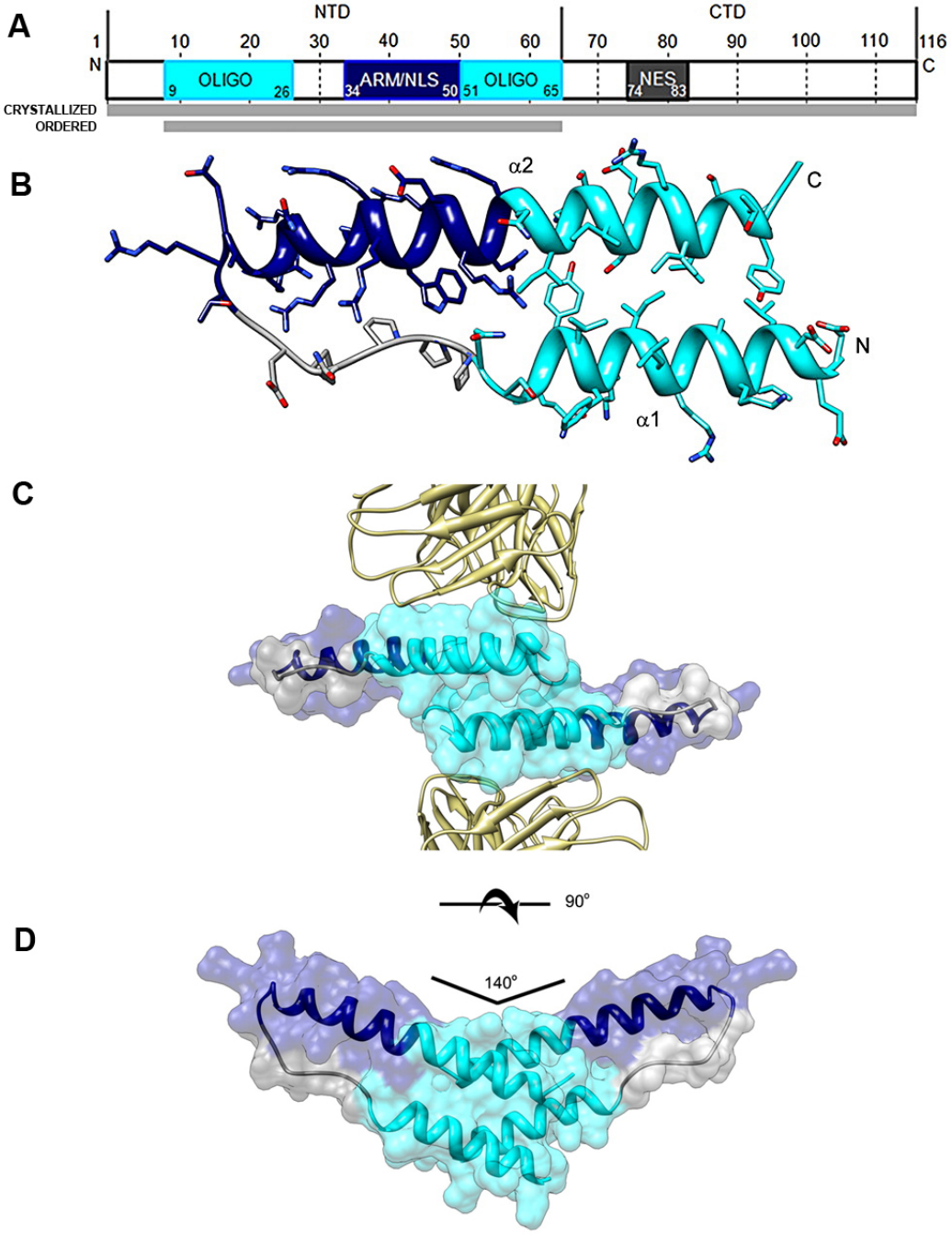


Figure 3.2: Rev Monomer and Dimer Structures. **(A)** Domain organization of Rev. Oligomerisation motifs depicted in cyan, arginine-rich motif (ARM)/nuclear localization sequence (NLS) in navy blue, and nuclear export sequence in dark gray. Although full-length Rev was crystallised, only the N-terminal domain was ordered. **(B)** Ribbon representation of Rev monomer (top view) showing the helix-loop-helix motif with coplanar helices. **(C)** Semi-transparent surface-rendering of Rev dimer, with ribbon backbone shown underneath. The Rev dimer is flanked by two Fab molecules in gold. **(D)** View of Rev dimer showing the 140° inter-axial angle (orthogonal to view in panel **C**).

loop IIb of the RRE, despite their lying outside the RNA-binding domain. It is likely that hydrophobic-to-hydrophilic mutations in these residues unhinge the helical hairpin and disorder the ARM, preventing the initial RRE-binding event. The inter-helix loop region probably also contributes to structural integrity, due to three consecutive prolines (sequence NPPPNPEGT) that form a rigid, concave stretch that is stabilized by the side-chain of W45. Supporting this, the electron density for this region is stronger than for the rest of the loop.

3.1.2 Structure of Rev *A-A* dimer

The Fab-Rev crystal structure reveals molecular dimers of Rev flanked by Fabs (**Fig 3.2C**). The Rev dimer interface is formed by a hydrophobic patch at the pronged end of the hairpin, where the subunits overlap by two helical turns (**Fig 3.2D**). The two subunits are related by an almost exact twofold axis (between 177° and 180°) and rest across each other at an angle of ~140° (**Fig 3.2D**).

In the dimer, identical *A* surfaces mate (i.e., in a head-to-head as opposed to a head-to-tail interaction) (**Fig 3.3A**), while the *B* surfaces interact with Fabs (**Fig 3.3B**). This result clarifies a prior ambiguity as to the manner in which Rev molecules oligomerise. An early proposal suggested the existence of just one oligomerisation interface that formed a heterotypic (or, asymmetric) interaction with adjacent subunits (Thomas et al., 1998). More recent reports suggested a model in which there are two oligomerisation surfaces, each defined compositely by residues upstream and downstream of the ARM (Jain and Belasco, 2001; Edgcomb et al., 2008). It is this latter model that is consistent with the Rev dimer structure.

It has been previously reported that Rev self-association is mediated predominantly by hydrophobic interactions (Jain and Belasco, 2001). Indeed, the dimer structure confirms this inference: the dimer-

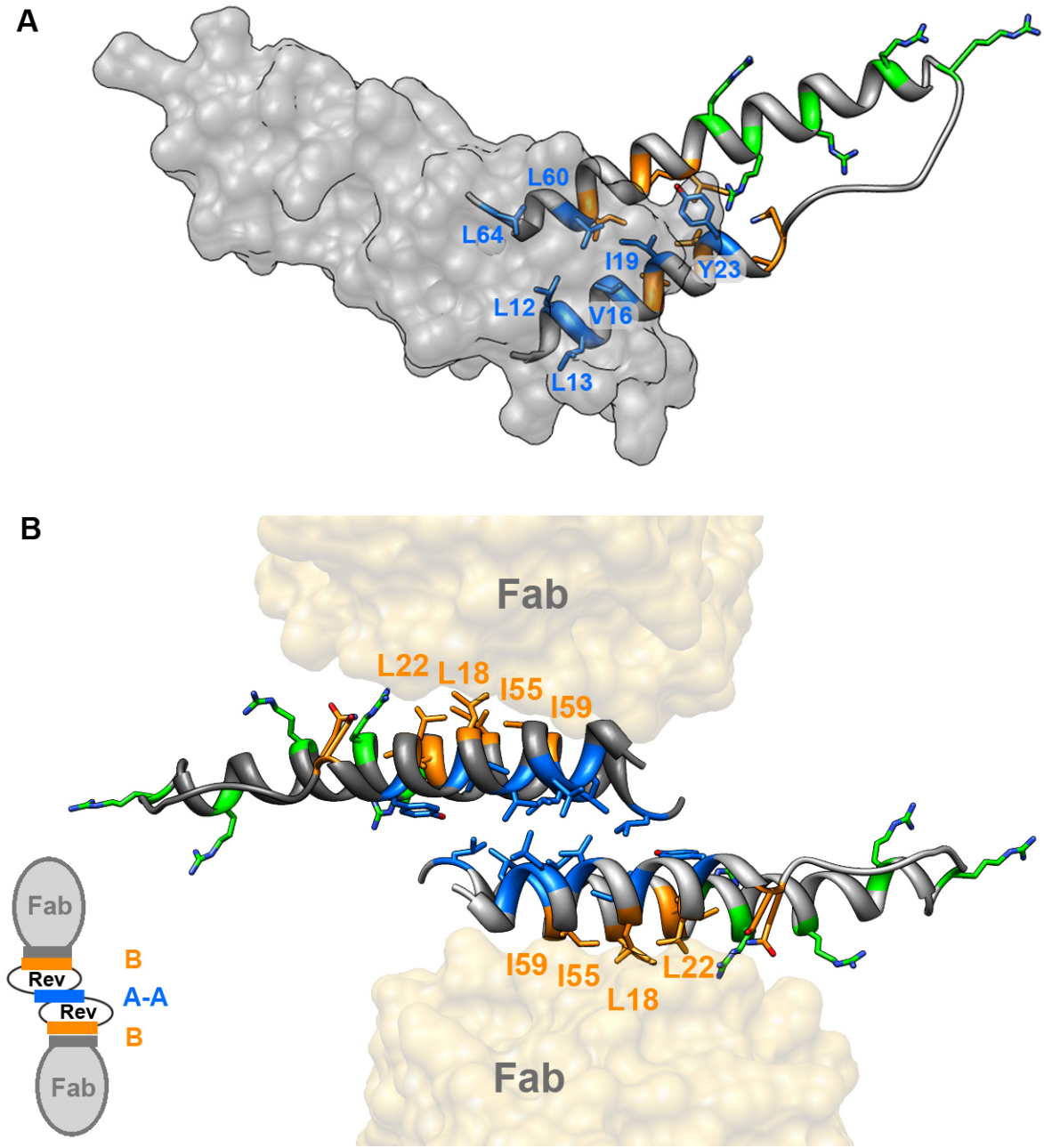


Figure 3.3: Rev Dimer Structure. **(A)** Rev dimer depicted with foreground subunit as a rendered isosurface and background subunit in ribbon representation. Residues involved in forming the *A-A* interface are coloured blue, and residues implicated in *B-B* oligomerisation and RRE binding shown in orange and green, respectively. **(B)** Ribbon representation of Rev dimer, orthogonal to view in **A**, depicting the *A-A* dimer interface and the predicted *B* oligomerisation faces contacting the Fabs. The oligomerisation surfaces and ARMs are segregated.

ization footprint on surface *A* (**Fig 3.4C**) matches closely with the location of its hydrophobic patch (**Fig 3.4A**). It is noteworthy that the hydrophobic patch on surface *B* (**Fig 3.4B**) coincides with the Rev epitope for this Fab (**Fig 3.4D**). One can thus envision that in the absence of Fab, Rev dimers could assemble further via *B*-*B* interactions as well. The hydrophobic patches required for dimerization, and (by inference) higher-order assembly, localize to the pronged end of the Rev monomer and are thus segregated from the RNA-binding ARM at the other end.

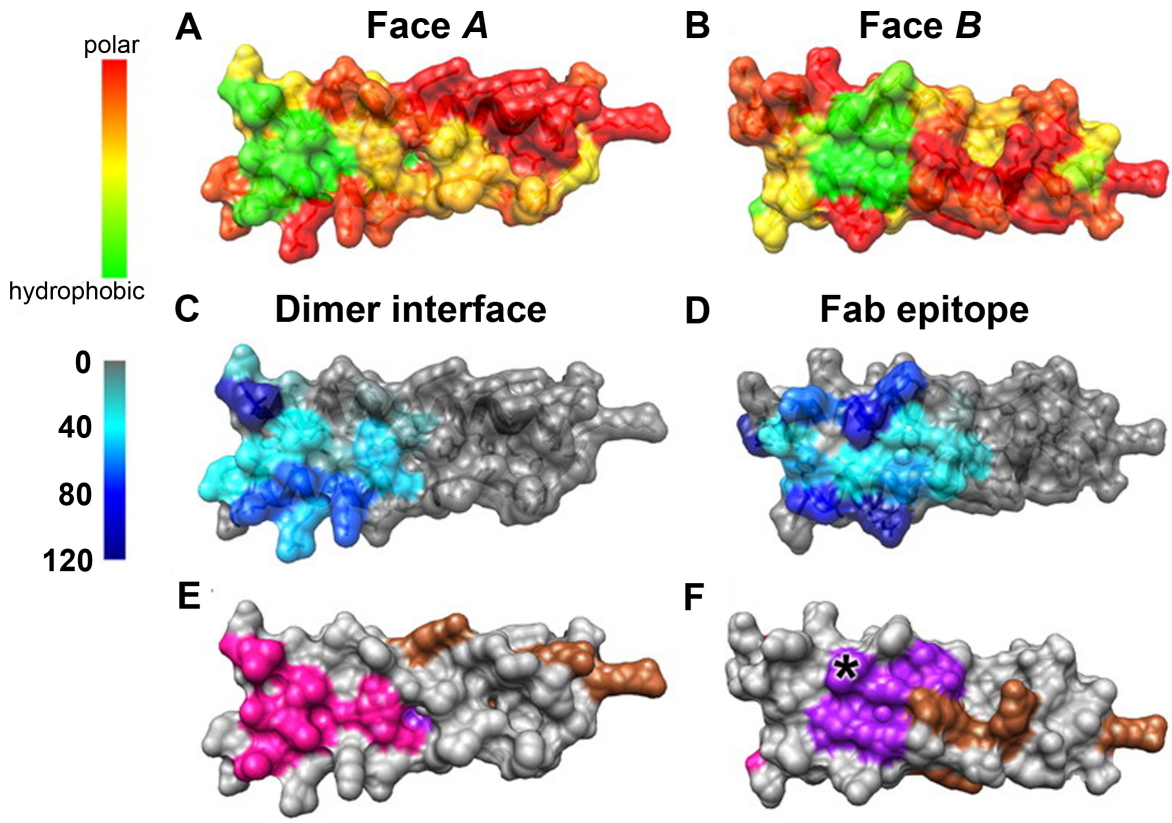


Figure 3.4: Rev oligomerisation faces. Rev monomers are isosurface-rendered and coloured according to: **(A and B)** surface hydrophobicity; **(C and D)** buried surface area (on a per-residue basis) at the Rev *A*-*A* dimer and Fab-Rev Face *B* interfaces, respectively, in Å^2; **(E and F)** residues shown to be involved in Rev multimerisation in pink and purple for Faces *A* and *B*, respectively. Residues important for RRE interaction are depicted in brown. L18 is marked with an asterisk to show its lateral position within the *B* face.

3.1.3 Structural analysis of the *A-A* dimer interface

Detailed analysis of the residues involved in the *A-A* interface is presented in (**Fig 3.3A**). As the hydrophobic residues L12, L13, V16, I19, L60, and L64 consistently bury more than 50% of their surface areas at the interface, it is clear that they make key stabilizing interactions. The dimer structure is consistent with a wealth of biochemical and genetic data on the functional relevance of specific residues in Rev oligomerisation. Jain and Belasco identified L12, V16, L60 and L18, L55 as key residues mediating Rev oligomerisation (Jain and Belasco, 2001), and these sets of residues map exactly to the *A* and *B* surfaces, respectively (**Fig 3.3A, B**). However, Edgcomb *et al.* further analyzed the effect of aliphatic-to-polar mutations in residues V16, L18, L55, and L60 and found that the mutants' reduced RRE-binding and nuclear export activities are more closely correlated with de-stabilization of Rev tertiary structure than simply the alteration of oligomerisation faces *A* and *B* (Edgcomb et al., 2008). Interestingly, they determined that of the residues investigated, polar substitutions at only residue 18 disrupt Rev oligomerisation without altering Rev structure (Edgcomb et al., 2008). While the other residues form the core of Rev, L18 sits at the edge of the hydrophobic cluster where it can affect oligomerisation without compromising structure. Additionally, the lateral position for L18 may explain Rev's putative role in HIV-1 latency: by reducing the affinity for Rev self-association somewhat but not entirely, Rev function is diminished, leading to less Gag expression and greater possibility for infected T cells to escape anti-Gag cytotoxic T cell responses. Mutations of the polar residues Y23, S25, and N26 also disrupt oligomerisation, resulting in an increased cytoplasmic concentration of Rev (Malim et al., 1989b; Szilvay et al., 1997; Trikha and Brighty, 2005). Y23 makes significant contacts both within the Rev monomer and at the *A-A* interface, explaining the functional phenotype; S25 and N26 are on the *B* surface and so support the imputed role of this surface in Rev oligomerisation.

3.1.4 Proposed Rev-RRE interaction model I

The Rev dimer structure may provide insight into Rev-RRE interactions, as the hydrophobic oligomerisation faces are buried upon Rev multimerisation in a manner that dictates an extended orientation of the RNA-binding ARMs. The NMR structure of the stem loop IIB-ARM complex (PDB ID: 1ETF) (Battiste et al., 1996) was modelled onto the Rev crystal dimer by superposition of the corresponding Rev regions between the structures. The RNA helix is positioned such that there are no steric clashes with the Rev dimer and there is ample clearance for the binding of additional dimers. While the ARM-SLIIB interaction involves double-stranded RNA, it is not known whether additional Rev subunits interact with single-stranded or double-stranded RNA. The two ARMs per dimer are at opposite ends of this elongated molecule, and when one subunit is modelled as RRE-bound, the other subunit clearly cannot bind to an adjacent groove of the same RNA helix. As such, two stem loops were modelled onto the dimer, the dyad axis aligning the two RREs into a track upon which additional Rev molecules might bind. In this model, stacked Rev dimers present arrays of ARMs spaced 20-25 Å apart, consistent with such an array making repetitive interactions with successive major grooves along the same side of the RNA helix. Because it has been shown that the ARM can bind RNA with different helical faces (Daugherty et al., 2008), Rev has some flexibility with which to bind additional RNA grooves as it multimerises. This arrangement probably imparts Rev with the needed spatial specificity for association with the RRE. The rather obtuse crossing angle of the Rev dimer (~140°) suggests that RNA may bind in a two-track model along successive, oligomerised Rev subunits (**Fig 3.6**). However, more data are needed in order to corroborate this proposed model. This model is very different from an earlier one, proposed in the absence of the Rev crystal structure, in which the angle between interacting Rev monomers is far more acute (Jain and Belasco, 2001). There is evidence that subsequent to the

initial RNA-Rev binding event, additional Rev molecules add into the complex by binding along stem loop I (Mann et al., 1994; Pond et al., 2009). While the structure of the RRE is largely unknown, it is plausible that stem loop I provides an opposing second RNA helix for binding on the opposite side of dimer from the initial stem loop IIb interaction.

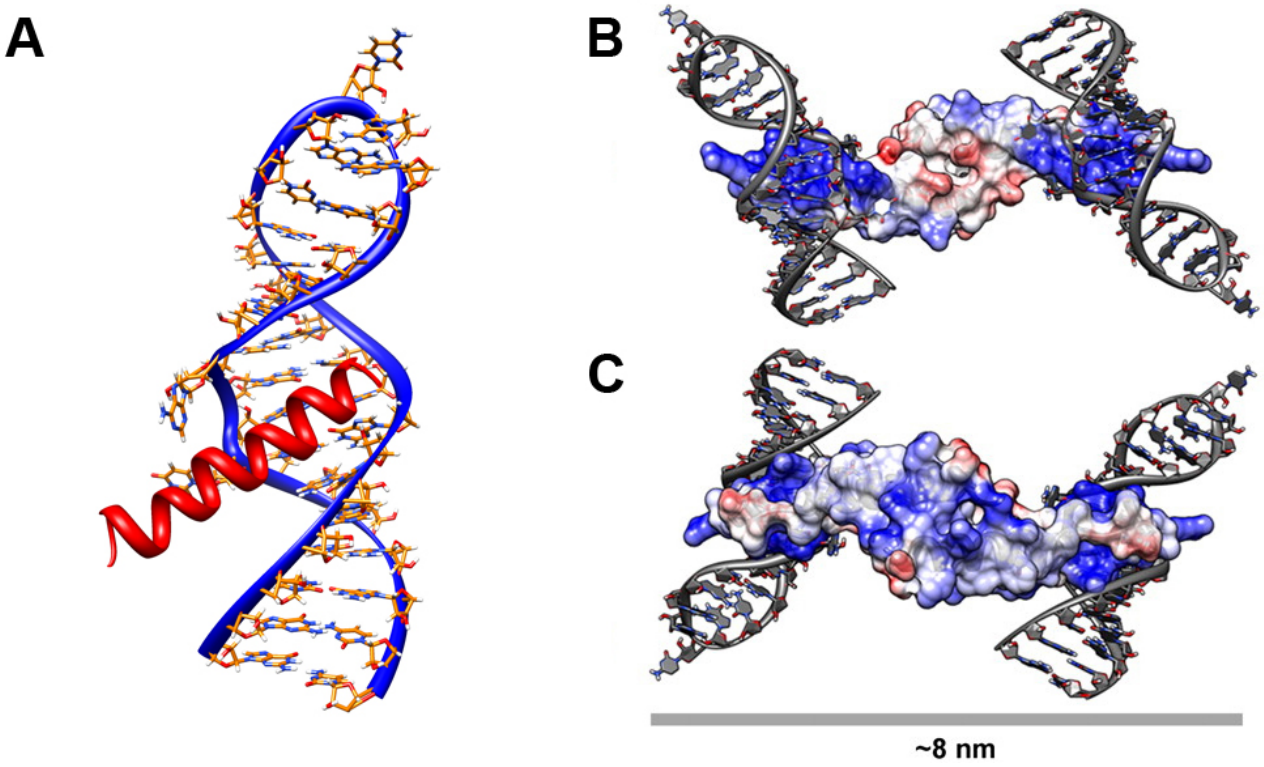


Figure 3.5: Modeling of Rev-RRE interaction. (A) Previously reported NMR structure of Stem loop IIb (high-affinity RRE-binding site) bound to synthetic Rev ARM (α-helix; residues 33-55) (Battiste et al., 1996). **(B and C)** NMR structure in **A** modelled onto Rev dimer by superposing the synthetic ARM onto the corresponding region in the Rev dimer (rmsd of 0.8 Å over main chain atoms of residues 37-55). The two views are related by a 180° rotation about a horizontal axis. Molecules are coloured according to surface charge, from acidic (red) to basic (blue). The two ARMs per dimer are at opposite ends of this elongated molecule, and when one subunit is modelled as RRE-bound, the other subunit clearly cannot bind to an adjacent groove of the same RNA helix. As such, two stem loops were modelled onto the dimer.

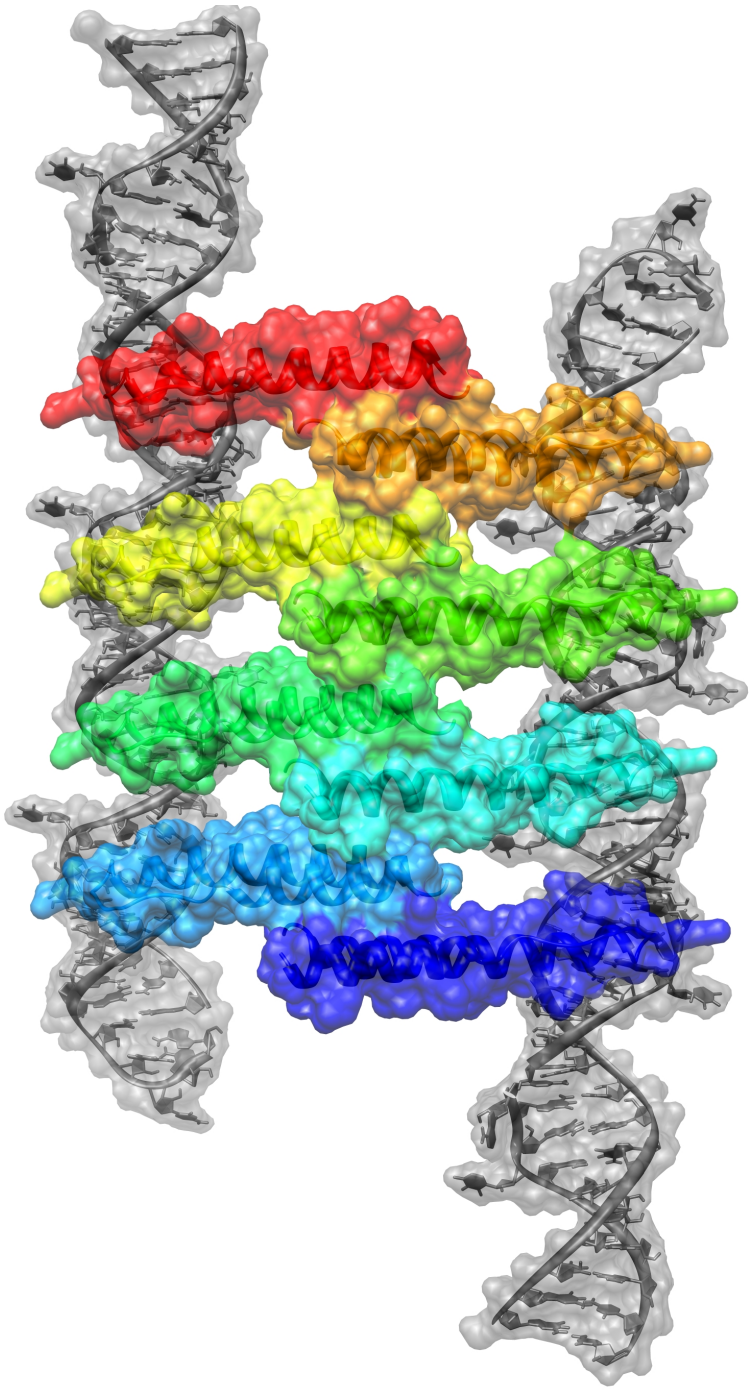


Figure 3.6: Proposed Rev-RRE association model I. Oligomerised Rev dimers are arranged with a dyad axis aligning the two RREs into a track upon which additional Rev molecules might bind. In this model, stacked Rev dimers present arrays of ARMs spaced 20-25 Å apart, consistent with such an array making repetitive interactions with successive major grooves along the same side of the RNA helix.

3.2 Structure of Rev *B-B* dimer

Subsequent to the Rev *A-A* dimer structure determination (DiMattia et al., 2010), Daugherty et al. reported a Rev crystal structure confirming the *B-B* interface proposed above (Daugherty et al., 2010b) (**Fig 3.7**). Instead of using an Fab to facilitate crystallisation, two hydrophobic-to-hydrophilic mutations at the *A-A* interface (L12S and L60R) were made. These mutations partially disrupted the *A-A* interface such that Rev oligomerisation was inhibited enough that well-ordered 3D crystals could form. Oligomerisation was not inhibited entirely, as the *A-A* interface was also observed (although in a distorted form, see below). The crystal asymmetric unit contained four Rev subunits, with two *B-B* interfaces and one *A-A* interface present. In addition to the engineered mutations, the disordered CTD (46 residues from the C-terminus) was removed to promote more stable crystal contact formation. The ability of $Rev_{1\text{-}70}$ to bind and oligomerise onto the RRE, at least up to a dimer, was verified, suggesting that the species retains a strong, and probably intact, *B-B* interface (Daugherty et al., 2010b). The space group of their crystal was $P6_422$, different from all space groups observed in the Fab-Rev and scFv-Rev structures (see **Chapter 2**). In this crystal structure, the *A-A* dimer interface is offset by one helical turn, presumably as a compensatory structural rearrangement due to the presence of the large, charged arginine side chain (L60R) in an otherwise entirely hydrophobic environment.

The *B-B* interface comprises the flat, hydrophobic patch opposite the *A* surface, as predicted in the Fab-Rev structural analysis. From the $Rev_{1\text{-}70}$ dimer of Daugherty et al., five key hydrophobic residues are recognised as comprising the *B-B* interface: L18, F21, L22, I55, and I59. Consistency with the residues observed as forming the epitope for the anti-Rev Fab can be observed by comparison with **Fig 3.3B**. Notably, the *A-A* and *B-B* interfaces are not aligned symmetrically along the axis of *A-A*

B-B A-A oligomerisation. While it is a visually subtle feature when observed directly for 3-4 subunits (**Fig 3.7**), the *A-A* interface clearly forms at the far edge of the Rev subunit (**Fig 3.4A**), while the *B-B* interface forms inward from the edge by about one helical turn (**Fig 3.4B**). Such an offset may be important in the context of helical Rev multimerisation: asymmetric propagation is a classic feature inducing helix formation.

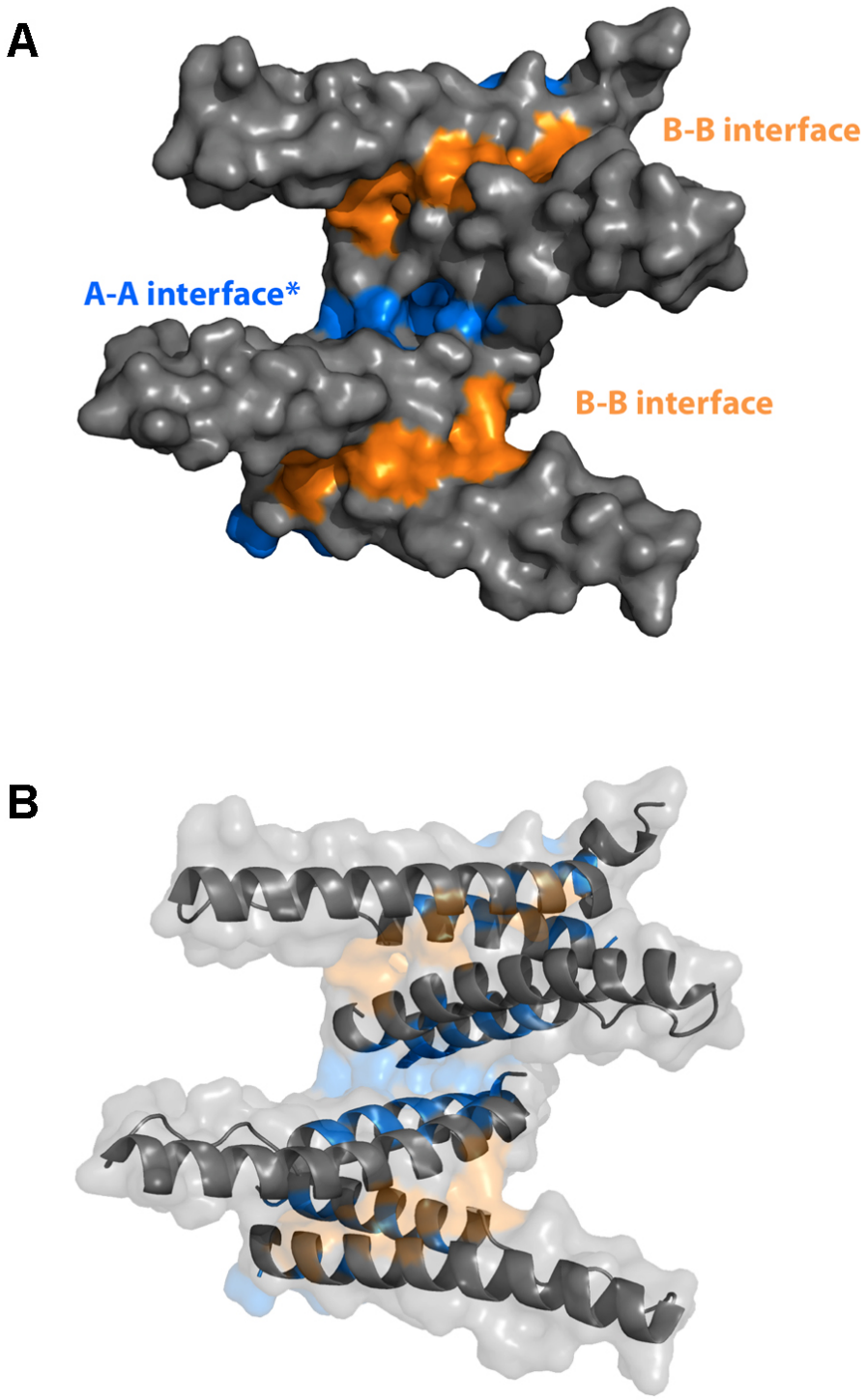


Figure 3.7: Rev dimer *B-B* interface. Asymmetric unit of $Rev_{1\text{-}70}$ crystal structure (Daugherty et al., 2010b). Four Rev chains are resolved, mediated by one *A-A* interface and two *B-B* interfaces. The *B-B* interface reveals that Rev oligomerisation propagates in a similar fashion to the *A-A* interface, as shown in the Fab-Rev crystal structure. In the structure above, $Rev_{1\text{-}70}$ crystals were obtained by introducing two hydrophobic-to-hydrophilic mutations at the *A-A* interface: L12S and L60R. These substitutions sufficiently disfavored the thermodynamics of Rev oligomerisation to allow for crystallisation to occur.

3.3 scFv-Rev crystal structures

Subsequent to the crystallisation and structure determination of the Fab-Rev complex, a variant of the same complex using a single chain variable fragment (scFv) instead of the full Fab, was crystallised. The scFv-Rev complex promoted tighter crystal packing (**Fig 3.8A**) and more efficient crystallogenesis than the Fab-Rev complex, resulting in higher resolution data (namely to 2.3 Å) (**Fig 3.8B**). Crystals of the scFv-Rev complex were grown under various conditions that gave rise to four unique space groups, which allowed for separate structure determination of each crystal setting (**see Chapter 2**).

3.3.1 Structural integrity of the helical hairpin motif

Unlike the previous 3.2 Å Fab-Rev data, the 2.3 Å data allows analysis of hydrogen bonding within the Rev monomer. Notably, there is a hydrogen bonding network that forms within a Rev monomer between the indole nitrogen of W45 and the backbone carbonyl of N26 and P27, as well as the amide oxygen of N26. As this network forms at the inter-helix region between the concave proline-rich loop and W45, the proper alignment of the α1 and α2 helices is maintained. Notably, contact between a poly-proline stretch and tryptophan (even in the same manner: concave proline-rich loop encasing a Trp) has been shown to be a very stable interaction, able to stabilize structural elements in proteins and protein complexes (Biedermannova et al., 2008). An additional hydrogen bond pair forms between Y23 and H53 between the α1 and α2 helices themselves. All of these residues participating in hydrogen bonding within the Rev monomer are highly conserved among all HIV-1 isolates, highlighting their role in maintaining the integrity of the *A* and *B* oligomerisation surfaces. Furthermore, mutation of W45 and/or the nearby residues of the proline-rich loop disrupts assembly of Rev onto the RRE, even though

it is not strictly part of either oligomerisation surface (Jain and Belasco, 2001; Daugherty et al., 2008). Once again, the biophysical requirement for Rev to oligomerise, particularly via proper formation of its oligomerisation surfaces, is coupled to RNA binding and therefore nuclear export function.

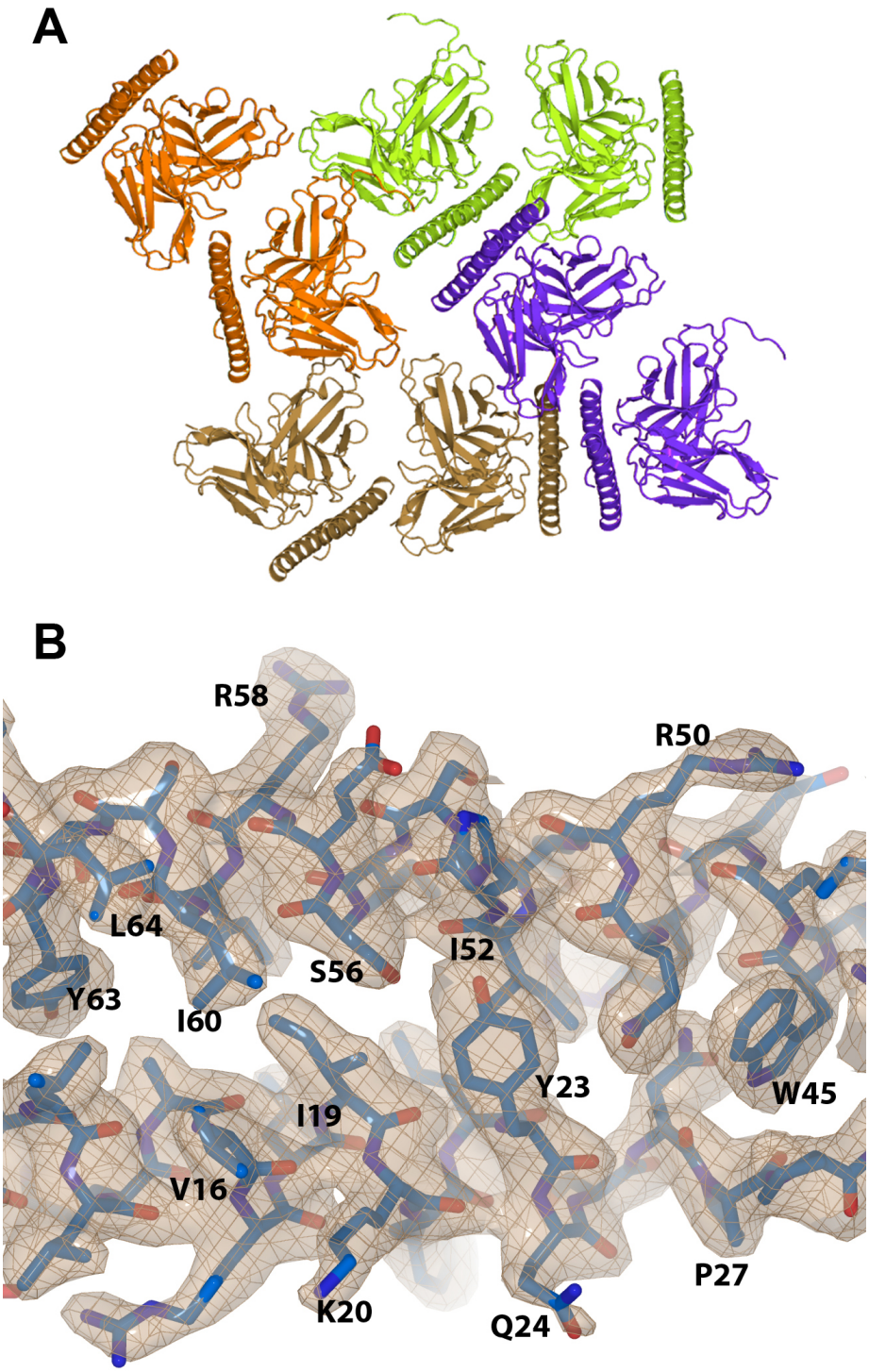


Figure 3.8: HIV-1 Rev electron density map. **(A)** Crystal packing of an illustrative scFv-Rev structure. The scFv-Rev complexes packed closer together than the Fab-Rev complexes, giving rise to more stable crystal lattices and accordingly higher resolution of X-ray diffraction data. **(B)** Sample HIV-1 Rev electron density map (2Fo-Fc) at 2.3 Å resolution and contoured at 1.5σ.

3.3.2 Pivot of the Rev *A-A* dimer

The scFv-Rev crystal structures revealed that there is considerable variability in the crossing angle of the Rev dimer (from 115° to 140°) (**Fig 3.9**). This was determined by superimposing all of the Rev dimer structures, using only one of the subunits for superposition. The crossing angle for each dimer was then calculated (**Fig 3.9**). In observing this phenomenon, a question arose: do the Rev subunits move via a ball-and-socket model–wherein there is a specific pivot point about which the dimer hinges– or via lateral sliding of one subunit across the other, as is plausible given the predominantly hydrophobic interaction surface. To answer this question, the spatial variance (SV) was calculated for each *A-A* dimer interface residue. Given two subunits of the Rev dimer, arbitrarily termed X and Y, if all dimers are superposed onto subunit X, spatial variance is defined as the degree to which a residue "moves" when comparing different subunit conformations of Y. This measure is attractive in that if the Rev dimer subunits do move by lateral sliding, the SV values of the interacting residues should all be roughly the same. However, this is not what was observed. At the *A-A* dimer interface, some residues had low SV values, i.e. they are close together in all of the different conformations: V16, I19, K20, Y23, S56, and L60 (**Fig 3.9, 3.10**). Other residues had large SV values, leading to the conclusion that Rev dimers pivot via a ball-and-socket model, with the spatially-conserved residues forming the socket around which the diver pivots. A careful examination of the pivot point structure reveals that V16, I19, Y23, and L60 form a largely non-polar socket flanked by polar groups of K20 and S56 (**Fig 3.11B**). The "ball" of the socket, residing in the other subunit, is predominantly formed by one residue, L64 but L63 and L60 contribute as well (**Fig 3.11B**).

Buried surface area (BSA) calculations of the residues at the Rev dimer *A-A* interface were performed for all five dimer structures, providing a multiplicity of data for assessment of which residues

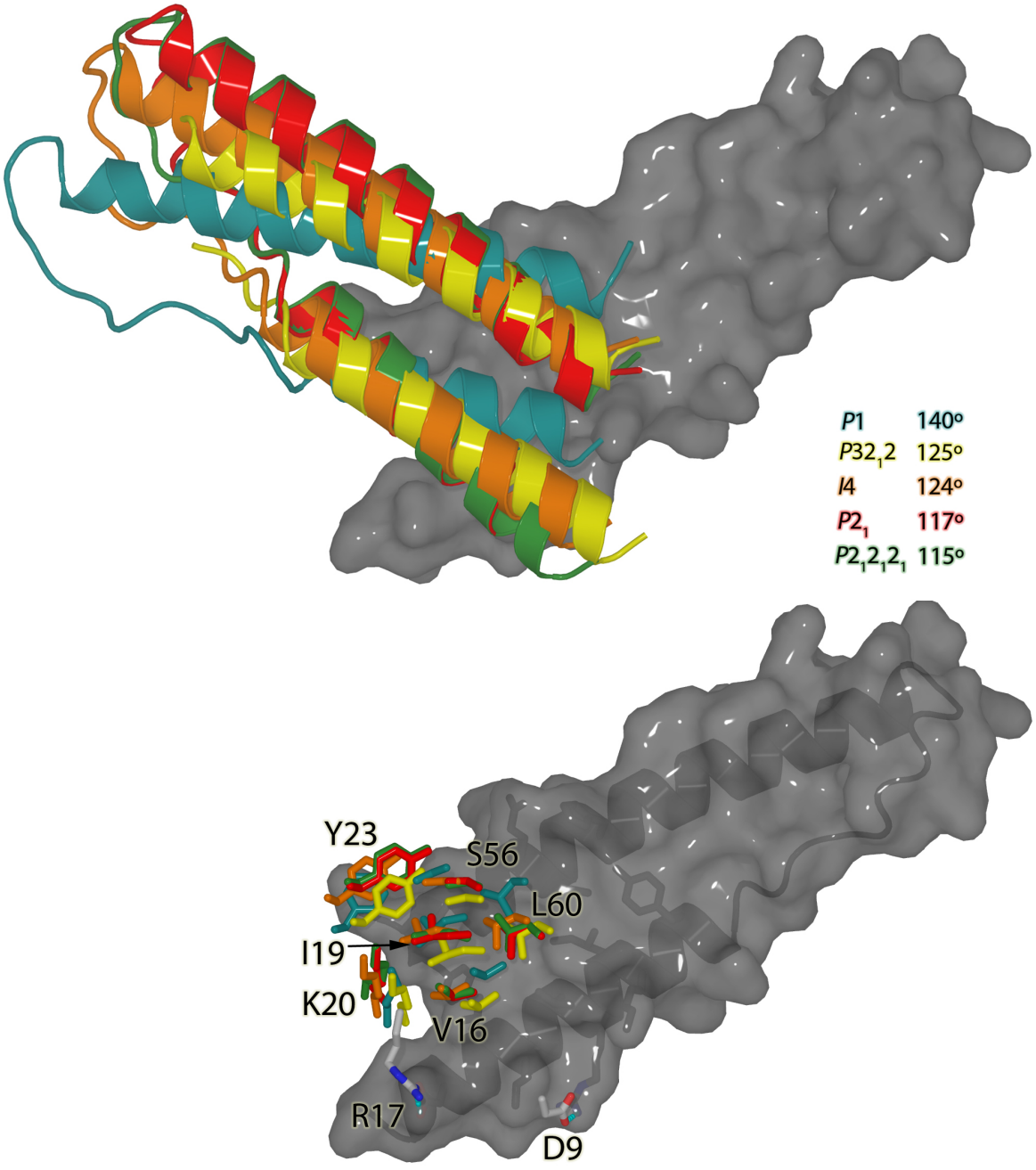


Figure 3.9: Pivoting of the Rev *A-A* dimer. (A) All five Rev dimer *A-A* structures superposed using the α-helical regions (residues 11-23, 45-63) of the isosurface-rendered Rev subunit (in gray). Structures are differentiated by colour of apposed subunit (see legend with crossing angles). **(B)** Rev dimers showing "socket" residues of the ball-and-socket model through which Rev hinges at its interface. Residues of top subunit are coloured as in **A**. Homotypic salt bridges form in the $P2_1$ (red) and $P2_12_12_1$ (green) structures between residues R17 and D9.

are buried in all instances, and are therefore likely to be more important for Rev dimerization. The aforementioned hydrophobic residues, L12, L13, V16, V19, L60, and L64 consistently bury the most surface area at the dimer interface. With the 2.3 Å resolution data, these six residues are clearly observed to form a non-polar cluster with the same six residues of the apposed subunit, forming the large majority of the contact between Rev subunits. In two of the structures, the side chains of D9 and R17 are positioned to form salt bridges, possibly stabilizing the interface in those instances.

It is unclear to what degree, if at all, the observed pivoting of the Rev dimer is biologically relevant. As a purely structural feature, it probably reflects the dynamic nature of Rev-Rev interactions, akin to the dynamism seen in many protein-protein interactions (Jones, 2012). While 25° does appear to be a rather large angular shift, the overall V-shape is maintained from one extreme to another. It is probable that the small nature of the Rev NTD, relative to the size of most protein dimers, exaggerates the observed flexibility. This dynamism may also reflect the fact that the Rev dimer, within the spectrum of observed homodimers, is a relatively weak interaction. The dissociation constant for Rev self-association is in the micromolar range (Cole et al., 1993). Furthermore, the dimer interface buries 1160 $Å^2$, which is less than the average BSA for weak dimers 1620 $Å^2$ according to a recent mining study of the PDB (Dey et al., 2010). However, the Fab-Rev inteface buries 1420 $Å^2$ and forms a tight interaction, so the BSA cannot be directly correlated with binding affinity. Notably, the dimer interface is almost entirely hydrophobic: with virtually no stability provided by polar interactions, the Rev dimer has greater freedom to pivot. One can conjecture that Rev employs this flexibility in order to adopt appropriate different conformations during different functional states, however at this time there is no additional structural or biochemical data to support this hypothesis.

It should also be noted that the dimer crossing angles observed in the six crystal structures (including the structure determined by Daugherty et al.) (115°, 117°, 120°, 124°, 125°, 140°) do not necessarily

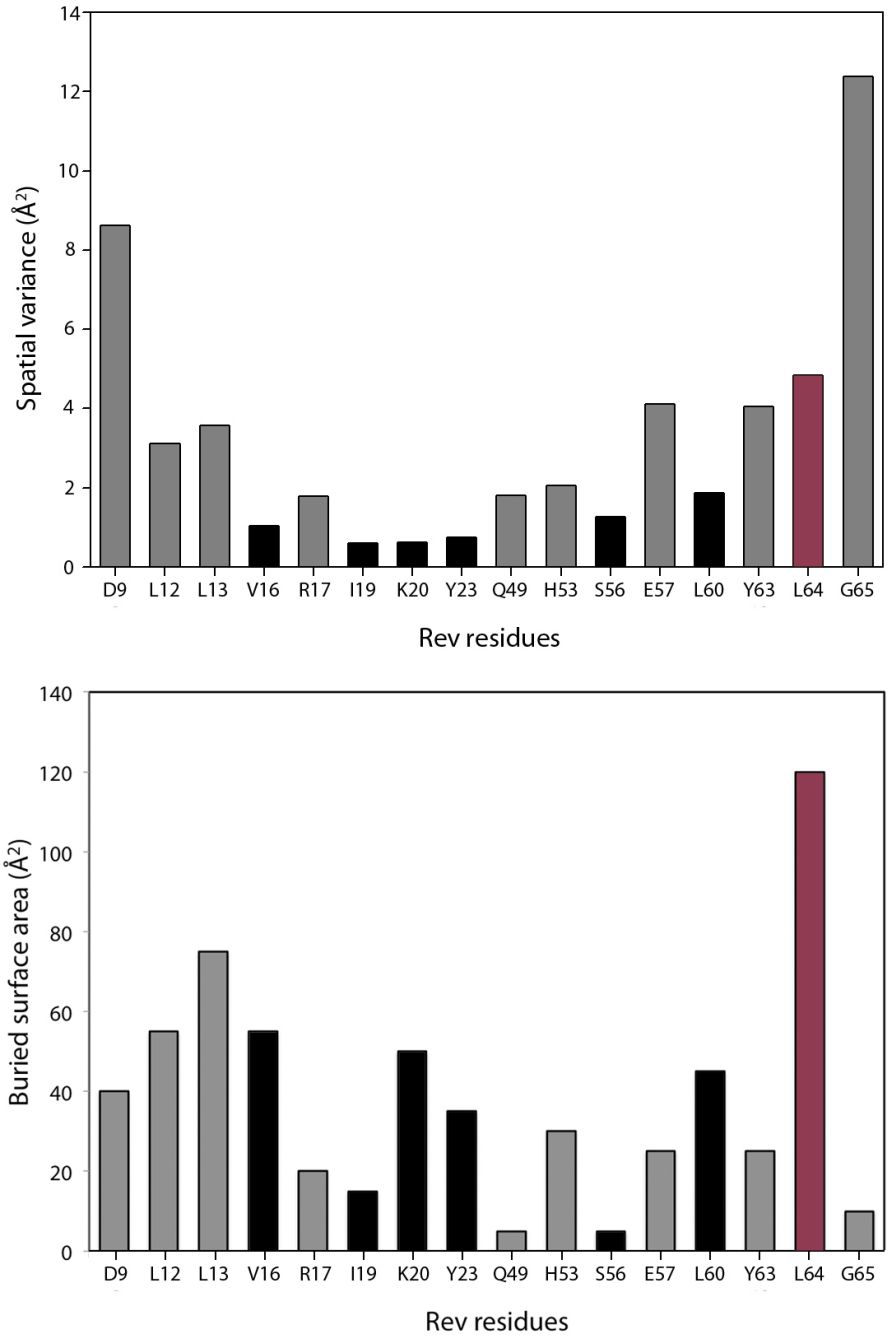


Figure 3.10: Biophysical statistics on Rev residues comprising the *A-A* dimer interface. (A) Spatial variance (SV; Å^2) of *A-A* dimer interface residues reflects positional shift when comparing different superposed Rev dimer structures. Residues in black form the "socket" of the Rev ball-and-socket pivot model. The residues' low SV values suggest that the residues are the stationary hinge about which the remainder of the Rev subunit pivots. Residue L64A is shown in maroon to highlight its role as the opposing "ball" that interacts with the "socket" to form the dimer hinge point. **(B)** Buried surface area (BSA; Å^2 of *A-A* dimer interface residues. Each BSA value is a per residue average of the five Rev dimer structures (Fab-Rev and 4 scFv-Rev).

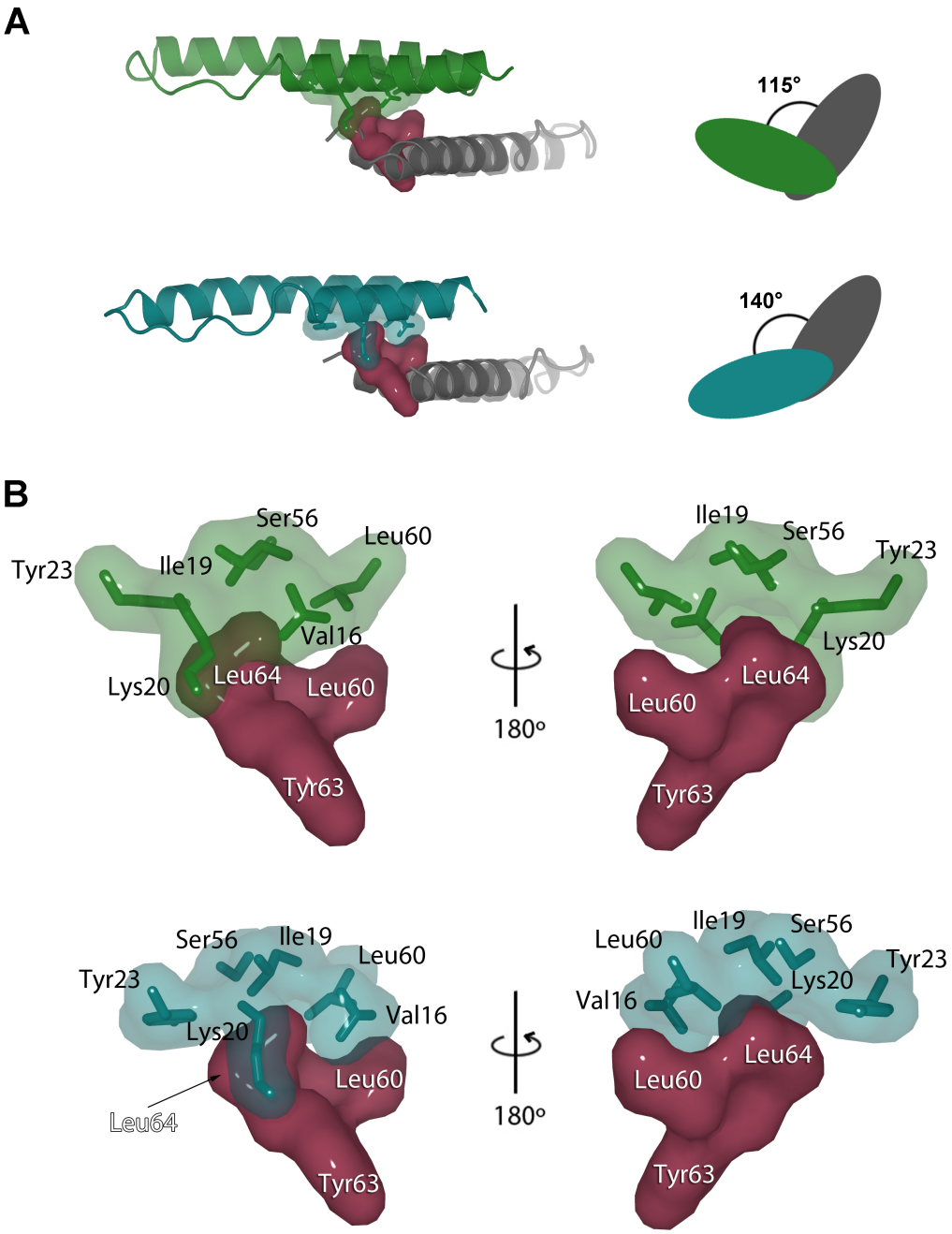


Figure 3.11: Ball-and-socket of Rev *A-A* dimer. **(A)** Shown are views of two Rev dimers approximately along the twofold axis: the green dimer corresponds to the most acute crossing-angle observed (115°) and the blue dimer the most obtuse (140°). The purple isosurface represents the "ball" of the ball-and-socket model. **(B)** At left are close-up views of the ball and socket of the Rev dimer hinge points. Also shown rotated 180°. L64 can be seen nestling into the socket formed by I19, Y23, V16, and L60 and flanked by polar residues K20 and S56. L64 buries nearly twice as much surface area as any other Rev dimer *A-A* interface residue, consistent with its position within the hinge pocket. The purpose of depicting this region for both dimer crossing angles is to reveal the stationary nature of the socket from one angular extreme to the other.

reflect the full possible range of angles through which Rev can dimerize. The molecular packing in each crystal lattice is distinct, requiring a different dimer crossing angle to propagate the lattice contact in each case. However, we do know that the Rev dimer is dynamic enough to accommodate at least a 25° range in its *A-A* crossing angle. Is is also unclear which conformation of the dimer is the most thermodynamically favorable, or even if there is a favored state. However, amidst a sea of hydrophobicity at the dimer interface, hydrogen bonding is only observed in the two most acute structures ($P2_1$ and $P2_12_12_1$), between the side chains of D9 and R17 (**Fig 3.9**). It is possible that the formation of these two-fold symmetric salt bridges may preferentially stabilize the Rev dimer with a crossing angle ~115-117°.

3.3.3 Mutation of *A-A* pivot point inhibits Rev oligomerisation

As detailed above, L64 was observed to be a key residue in the ball-and-socket pivot model about which the Rev *A-A* dimer hinges. Unsurprisingly, L64A buries the most surface area out of all *A-A* participating residues, nearly twice as much as the next most buried residue (**Fig 3.10**). Therefore, it was hypothesized to be an essential residue for formation of the *A-A* dimer interface, and by extension, Rev oligomerisation at large. In order to test this hypothesis, a Rev mutant was generated in which L64 was mutated to Ala. Alanine was chosen over a bulky, hydrophilic group such as arginine in order to rule out steric hindrance or charge effects as possible consequences of the mutation. When observed by negative-stain electron microscopy (EM), Rev L64A did not form filaments at a concentration at which WT Rev readily does (0.2 mg mL^{-1}) (**Fig 3.12A, B**).

Dynamic light scattering (DLS) is a biophysical technique whereby intensity fluctuations of light scattered by macromolecules are used to measure polydispersity and distribution of macromolecular

size. Here, DLS was performed and while both Rev and Rev L64A samples were polydisperse, WT Rev was much more so than the mutant and also contained species with an order of magnitude larger hydrodynamic radius (R_H) (**Fig 3.12**). However, given that R_H assumes that a molecule is spherical, it is a poor parameter for analyzing Rev filament formation, which are long and thin (non-spherical). The samples were also assayed by sedimentation equilibrium experiment using analytical ultracentrifugation. The C(s) trace shown in (**Fig 3.12**) reveals that WT Rev contains a vast array of different sized species (sedimenting from 5S-150S). Rev L64A also contains many differently sized entities, particularly in the 5-30S range, with a single peak around 75S. The DLS and AUC data trend toward the notion that WT Rev forms at least an order of magnitude larger species, either quantified by R_H or Svedberg units. Nevertheless, Rev L64A retains very large species. These could simply be non-specific aggregation, as mutation of L64 to Ala does not shield the large hydrophobic areas from the hydrophilic buffer environment. Altogether, these data support the notion that Rev oligomerisation is inhibited, at least to some degree, by a rather conservative Leu to Ala mutation. These data suggest that the ball-and-socket region of the Rev A-A dimer interface, particularly centered around L64, may be a prototypical "hot spot": a region of key structural importance within a protein-protein interface (Clackson and Wells, 1995). This line of investigation may have clinical applicability, e.g. via through development of a transdominant negative phenotype *in vivo* using Rev L64A or as a starting point for fragment-based drug design of Rev oligomerisation inhibitors (Kuttner and Engel, 2012; Thanos et al., 2006).

3.3.4 Evidence for a third Rev oligomerisation interface

Thus far, the two Rev oligomerisation surfaces, *A* and *B*, have been structurally characterised. The biochemical literature exploring residues involved in oligomerisation maps precisely to the *A* and *B*

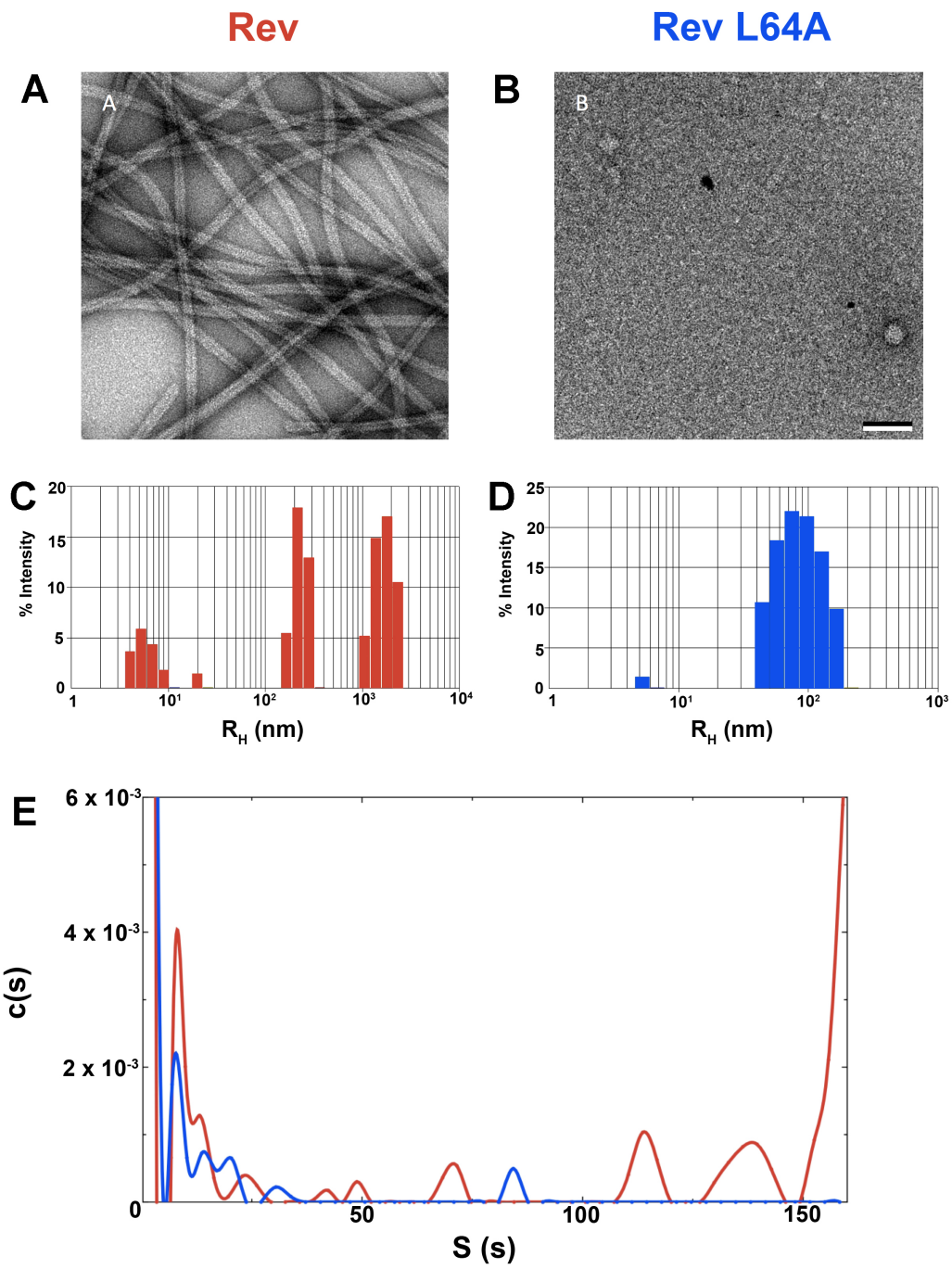


Figure 3.12: Rev L64A mutant. **(A and B)** Negative-stain electron micrographs of WT Rev at 0.2 mg mL^{-1} **A**, above the threshold concentration at which filaments readily form. In **B** is Rev L64A at the same concentration; no filaments were observed. Scale bar is 50 nm. **(C and D)** Dynamic light scattering of Rev (red) and Rev L64A (blue). Large species are observed in both samples, ~1000 nm hydrodynamic radius for Rev and ~100 nm for Rev L64A. Hydrodynamic radius (R_H), however, is a poor biophysical measure for Rev filaments, as they are non-ideally shaped into long and thin tubes rather than an ideal sphere. The large species seen in Rev L64A may simply be protein aggregate via the Rev hydrophobic platforms which are unable to properly oligomerise without the presence of L64. **(E)** Analytical ultracentrifugation (AUC) on both samples, coloured as above. Again, large species are observed in both samples, although significantly larger for Rev than Rev L64A (150S vs. 75S)

faces. As presented above, the scFv complex crystallised in four different space groups, each with its own lattice packing arrangement. Given the data redundancy, the different structures were analyzed for any notable similarities or differences. One such pattern was observed in a crystal packing interaction that occurs between Rev dimers, in three out of the four scFv-Rev crystal structures ($P2_1$, $P2_12_12_1$, $P32_12$). As shown in **Fig 3.13A**, the extended arms (or ARMs, if you prefer) of each Rev dimer interact laterally with the adjacent dimers, interfacing primarily through the proline-rich loop that is intercalated between α1 and α2. Consequently, a repeating "V-shaped" dimer propagation is observed in the crystal packing, wherein Rev dimers are formed through the *A-A* interface and are subsequently connected by neighboring inter-helix loops, termed here the *C-C* interface (**Fig 3.13B, C**).

An investigation by Zygmunt Derewenda has revealed that crystal contacts can either be cohesive and contribute to the thermodynamic stability of crystal growth (i.e. primary contacts) or they can be repulsive and forced (secondary contacts) (Derewenda, 2011). Secondary contacts occur when molecules are serendipitously brought within physical proximity during the thermodynamic "collapse" of crystal nucleation, rather than via the local formation of a thermodynamically favorable interaction. It is difficult to conclude whether a given crystal contact is primary or secondary; however, the preponderance of views of this interaction supports the notion that it is a primary contact. Remarkably, in addition to the three views of the putative *C-C* interface in the scFv-Rev crystal structures, it is also observed in the crystal packing of the $Rev_{1\text{-}70}$ structure (Daugherty et al., 2010b), forming in exactly the same manner despite very different crystal packing environments (the former contain the scFv while the latter does not). The interface is not seen in the crystal packing of the Fab-Rev structure because the large size of the Fab spatially segregated the Rev dimers, preventing them from interacting in any manner.

The different views of the *C-C* interface are presented in **Fig 3.14A, B**. Similar to the *A-A* interface, the interaction is dynamic and appears to pivot around a particular region of residues. Interestingly, one

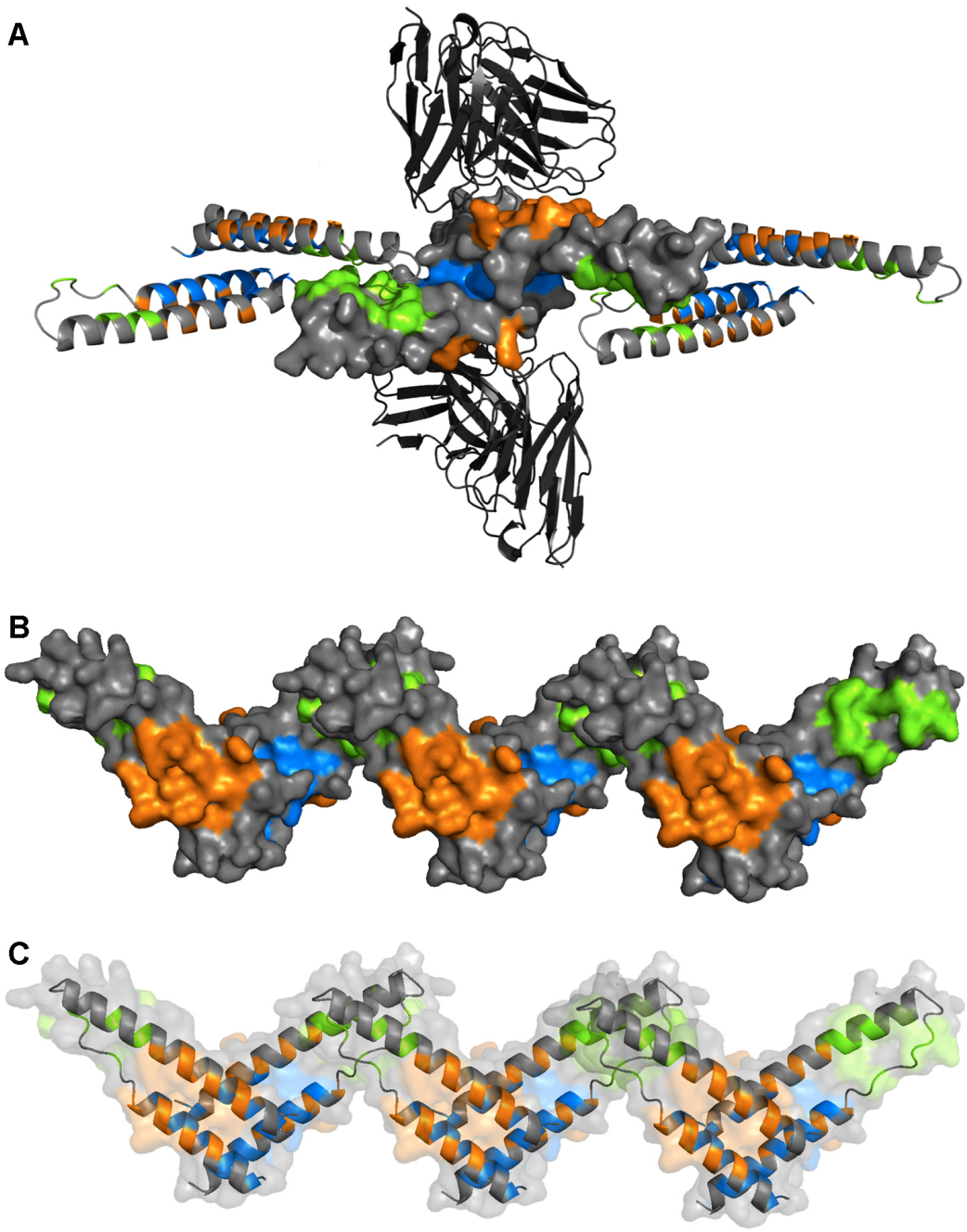


Figure 3.13: Crystal packing of scFv-Rev structure reveals *C-C* interface. **(A)** Crystal packing of scFv-Rev $P2_1$ structure. In the center is depicted a Rev dimer formed through *A-A* interface (blue) and bound to two scFv molecules via *B* oligomerisation faces. Adjacent Rev dimers in the crystal lattice are also shown (additional scFv molecules are not shown for visual clarity). These peripheral Rev dimers appear to interact laterally with the central Rev dimer, termed the putative *C-C* interface (green). **(B and C)** Rev dimers in **A** rotated 90°. The *C-C* interface can be seen, aligning Rev dimers such that successive subunits are parallel to each other. The inter-helix loop and the beginning of α2 appear to be contacting in the interface.

of the two views of the *C-C* interface in the structure of Daugherty et al. has a significantly different interaction footprint than the others (shown in red in **Fig 3.14C**). This molecular shift is probably compensating for the distorted *A-A* interface in the aforementioned structure (not shown), due to the two hydrophobic-to-hydrophilic mutations at the *A-A* interface. The other view of the *C-C* interface in the $Rev_{1\text{-}70}$ structure is consistent with those observed in the scFv-Rev structures.

Perhaps most intriguing are the specific residues that participate in the homotypic *C-C* interface. Both W45 and the proline-rich loop—already characterised as forming a rigid structure important for stabilizing the Rev monomer (see **Section 3.3.1**)—form the framework of the *C-C* interface. Here, the poly-proline loop, essentially a continuous hydrophobic strip comprising P28, P29, and P31, "hooks" into the same poly-proline groove on the opposing subunit (**Fig 3.15A**). Proline-proline interactions are common stabilizers of protein-protein interfaces due to their relative rigidity compared to other exposed (i.e. non-globular) sequences (Kay et al., 2000). Secondly, the *C-C* interface has a stacking interaction between W45 and P31, a type of interaction which has also been shown to be very strong, even without the presence of any classical hydrogen bonding (Biedermannova et al., 2008) (**Fig 3.15A**). Finally, a set of hydrogen bonds are also predicted to form between R46 and backbone carbonyl moieties of the opposing subunit. Proline-rich sequences are known to form weaker protein-protein interactions than those formed by globular domains. Interestingly, it has been suggested that such weaker binding is advantageous in situations where a structurally-defined complex is not the priority, but rather the recruitment of proteins in such a way that subsequent interactions are more probable, e.g. during cytoskeletal rearrangements and signalling cascades (Kay et al., 2000).

Whether the *C-C* interface is biologically relevant or not (i.e. its formation is involved in Rev oligomerisation *in vivo*) remains unknown. There is a dearth of biochemical data to complement this structural characterisation. However, there is a sparse collection of data indicating that mutation of W45

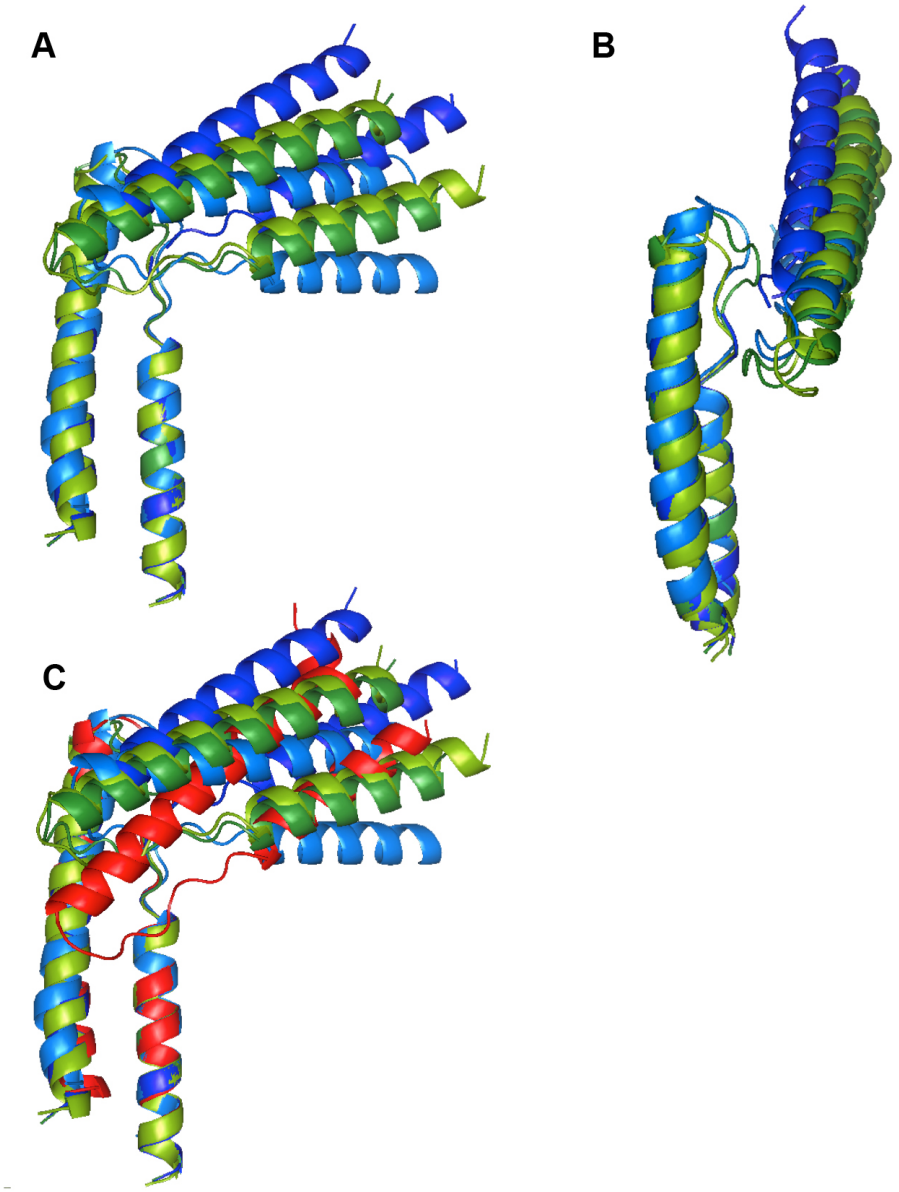


Figure 3.14: Dynamic nature of the Rev *C-C* interface. (A and B) Superposition of the *C-C* interfaces observed in the crystal packing of the scFv-Rev and $Rev_{1\text{-}70}$ (Daugherty et al., 2010b) structures. The inter-helix loops appear to "hook" into each other in a manner that allows considerable flexibility of each subunit. Repeated views of the same interaction in different crystal contexts suggest that the interface is a primary crystal contact rather than a secondary one, as defined in Derewenda (2011). **(C)** The Rev subunit shown in red corresponds to one of the two views of the *C-C* interface in the $Rev_{1\text{-}70}$ structure. Its apparent outlying conformation is probably an artificial compensation for the mutations at the *A-A* interface in that structure, which alter the alignment of the A-A interface by an entire α-helical turn. If the *C* face is utilized in an *in vivo* context, this aberrant observation is evidence that the *A*, *B*, and *C* interfaces are not static entities, but rather work in concert with each other to form larger, multi-subunit macromolecular assemblies.

and other inter-helix loop residues (which notably fall outside the *A* or *B* oligomerisation surfaces) have a deleterious effect on Rev multimerisation onto the RRE (Daugherty et al., 2008; Jain and Belasco, 2001). These mutagenesis data have been attributed previously to loss of Rev monomer structure. While probably a valid interpretation, the impact these residues may have on formation of the putative *C-C* interface, and by extension Rev oligomerisation, is as yet unexplored and cannot be ruled out.

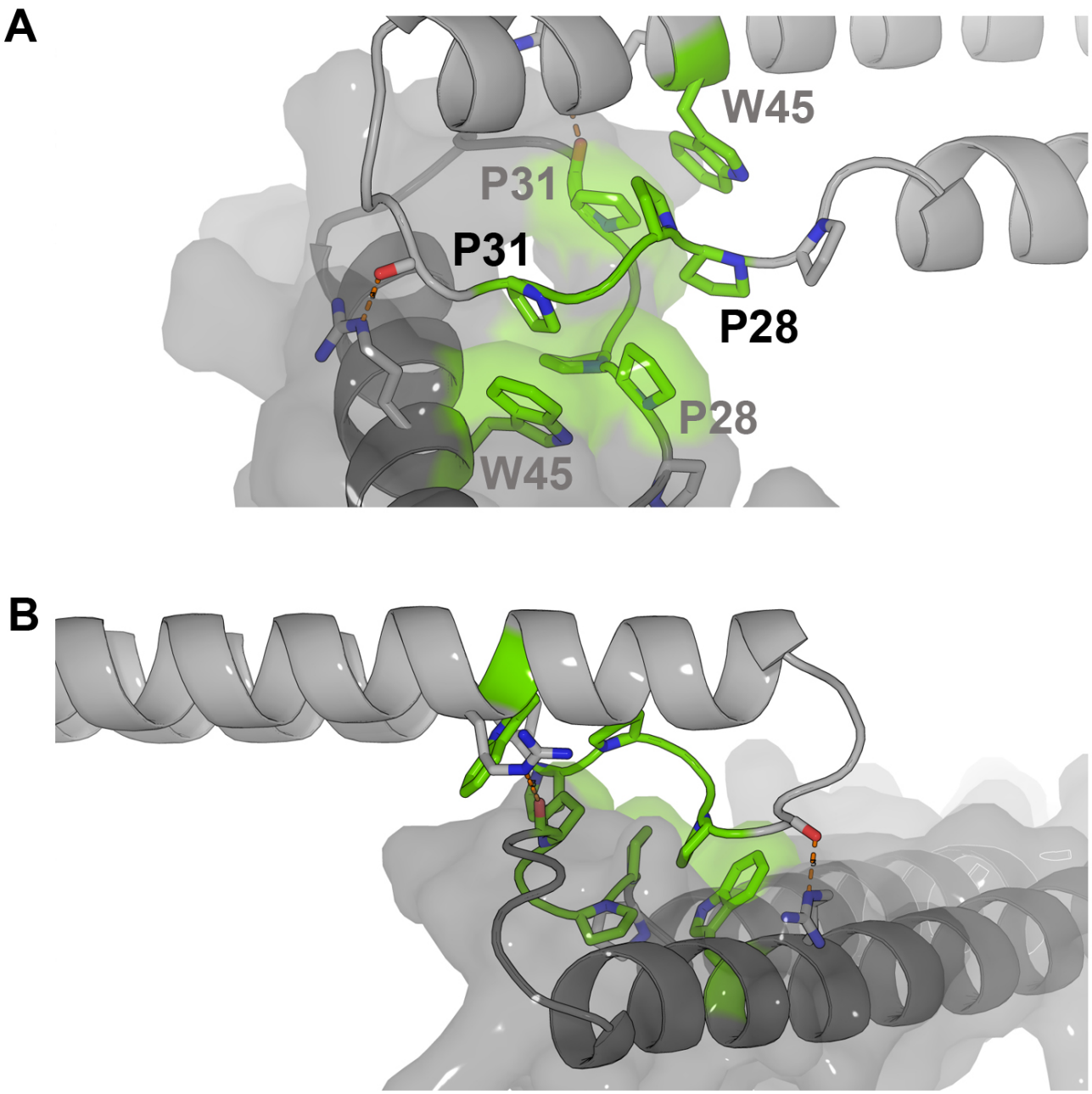


Figure 3.15: Rev *C-C* interface. **(A)** Primarily forming the Rev C-C interface are W45 and the proline-rich loop (P28, P29, P31 interact, but P27 does not). The conformationally rigid poly-proline loops hook into each other, supported by stacking interactions between W45 and P31. Both of these interactions have been shown to be a way of maximizing protein-protein interaction affinity over short stretches of amino acid residues (Kay et al., 2000). The interface is further stabilized by two hydrogen bonds between R46 and backbone carbonyl moieties of the apposed subunit. **(B)** Rotated view to show the interface from a different angle. The tryptophan-proline stacking, proline-rich loop intercalation, and predicted hydrogen bonds can be observed from this angle.

3.3.5 Disorder of the C-terminal domain

In the crystal packing arrangement described above for the Rev-Fab crystal structure, the CTD is disordered. The CTD is also disordered in all four of the scFv-Rev crystal structures, although occasionally an extra helical turn is observed extending to approximately residue 70. However, in the Fab-Rev structure at least, adjacent to the points at which the disordered residues 1-7 and 70-116 must join the ordered structure are regions of weak (~0.8σ), protein-like density without clear connectivity. There is now evidence that the CTD is intrinsically unstructured (Daugherty et al., 2010a). One possibility is that the disordered CTD becomes partly or completely folded in the context of complex formation and/or higher order Rev multimerisation, because Rev also interacts with many other cellular factors and viral proteins. Once the interfaces involved in these interactions become apparent, further insight into their functional roles should follow. In the meantime, the proposed Rev-RRE model I reveals key features of the organization of the initial Rev-RRE complex (**Fig 3.6**), including segregation of the surfaces that modulate oligomerisation from the RNA-binding ARM regions, thus allowing a tentative explanation for how the structural organization of Rev dimers drives the non-specific binding of RNA.

3.4 Conclusions

Rev-mediated nuclear export of unspliced viral mRNAs is essential for HIV replication. Decades of molecular biology work have pioneered investigation into Rev function, all of which has been performed in the absence of a crystal structure for Rev itself. In order to crystallise Rev, complexes were made of Rev with a high affinity Fab molecule that prevent aggregation and filament formation, features that are typically refractory to crystallisation. A variant of the Fab was produced as a single-chain variant fragment form (scFv) and was also complexed with Rev, facilitating Rev structure determination to 2.3 Å resolution.

Despite being a rather small protein (~12 kDa), only the N-terminal half of Rev (NTD) is ordered in all of the crystal contexts observed. The NTD forms an anti-parallel helix-turn-helix motif that presents two co-planar hydrophobic platforms through which Rev monomers oligomerise. These faces have been designated *A* and *B*, one for each side of the Rev subunit. Furthermore, how the oligomerisation surfaces interface has been structurally characterised: identical faces unite in a homotypic as opposed to a heterotypic fashion (i.e. *A*-*A* and *B*-*B*, not *A*-*A* *B*-*B* *A*-*A*). Point mutation of residues that present at the dimer interfaces severely disrupts Rev oligomerisation (and nuclear RNA export), typically through ablation of proper Rev tertiary structure. The conserved nature of these residues, their imputed role both structurally and functionally, and kinetic studies suggesting the dimer may be an early intermediate assembly of the export complex (Pond et al., 2009), altogether make the *A* and *B* oligomerisation platforms attractive targets for inhibitor design. Along these lines, structural analysis of the Rev dimer *A*-*A* interface revealed a key residue, L64, which may be a "hot spot" for Rev subunit association, and therefore a plausible starting point for structure-based drug design. The design of small molecules that

bind protein-protein interfaces, particularly those that are rather hydrophobic and flat, proves to be a challenge. However, *in silico*-binding analysis against the structure of the Rev subunit (with a spatial focus on L64), followed by fragment-based screening may prove to be a worthwhile endeavour. Biophysical experiments suggest that Rev oligomerisation is disrupted by a rather conservative mutation of this leucine to alanine.

How Rev subunits self-associate is an important question and the data presented in this chapter provide a step forward in our understanding of this process. Nevertheless, it is clear that the ultimate goal is to understand how the protein's ability to oligomerise relates to its association with the RRE. We know that protein-protein interactions between Rev monomers are essential for Rev-RRE binding specificity and RNA export, and that Rev subunit binding to the RRE is highly cooperative, ultimately forming a very specific and high-affinity RNP complex (Malim and Cullen, 1991; Mann et al., 1994; Daugherty et al., 2008). Furthermore, a single interaction between a Rev monomer and the high affinity binding site in the RRE (stem loop IIb) is not sufficient to mediate export. Further specificity between Rev and the RRE, presumably conferred by cooperative oligomerisation, is required. Accordingly, we can begin to understand how this cooperativity may occur by reviewing the highly dynamic and multi-faceted manner in which Rev subunits interact with each other:

1. The *A* and *B* oligomerisation faces are buried upon Rev multimerisation in a manner that dictates an extended orientation of the RNA-binding ARMs.

2. This arrangement probably imparts Rev with the needed spatial specificity for RRE-association, e.g. with stacked Rev dimers presenting arrays of ARMs spaced 20-25 Å apart, consistent with repetitive interactions with successive grooves along an RNA helix.

3. The *A*, *B*, and *C* oligomerisation faces can form dimers with a variety of different crossing angles,

while conserving the basic framework of Rev dimer "V-shape" topology. This crossing angle variation may allow for optimization of the fit between Rev ARMs and different helical grooves of the RRE.

4. The *C-C* interface presents extended inter-molecular contacts that are likely to be strong relative to the small interaction surface (see **Section 3.3.4**), but which are probably lower affinity than a typical globular protein-protein interface, and almost certainly lower affinity than that of the *A-A* and *B-B* interfaces. This is consistent with the conjecture that the *C-C* interface may be an auxiliary interaction that aligns successive Rev *A* and *B* faces during Rev multimerisation.

5. There is a propensity for the *A-A*, *B-B*, and *C-C* interfaces to form even under artificially-constrained conditions, such as a highly-charged arginine placed within the entirely hydrophobic *A-A* interface). Remarkably, the interfaces (in this case involving the *C-C* interface) appear to be able to compensate for each other by altering crossing-angles as necessary.

As might be expected for a small RNA-binding building block, these data altogether suggest that Rev subunits probably work in concert to multimerise in a highly dynamic fashion, able to structurally adapt as needed for RNA-association as well as other functional states. The Rev-RRE proposed model I incorporates the data of the Rev dimer having a rather obtuse (~140°) crossing angle. This limited the model to RNA binding of successive major grooves along properly spaced ARMs of oligomerised *A-A B-B A-A* Rev subunits in a two-track fashion. With the determination of five additional Rev structures, it is noted that dimer crossing angles have a large dynamic range, and it is conceivable that Rev subunits may align far more acutely, such that successive ARMs may both contact the same RNA helix. Additionally, our knowledge of the functional link between cooperativity and RNA export is limited by the lack of structural knowledge regarding the CTD, which appears to be disordered in all contexts

studied thus far. How Rev associates with host cell proteins such as Crm-1 (exportin-1) and importin-β to actually effect nuclear egress and ingress, respectively, also remains to be discovered. The structure of the Rev dimer building block clarifies that the essential next step for correlating Rev structure to function is to acquire data that reveals the higher-order assemblies Rev can form, either in the absence or presence of the RRE. The former encompasses the focus of the following chapter.

Chapter 4

The Structure of Rev filaments

4.1 Background

HIV Rev is often discussed in terms of its role in viral RNA binding and subsequent export of the RNP from the nucleus. However, it has been observed by electron microscopy that Rev readily forms polymers *in vitro* in the absence of RNA (Stahl et al., 2010). These polymers are hollow tubes composed of repeating Rev dimer building blocks that arrange into a helical lattice (Watts et al. (1998) and personal communication). Solving the structure of the Rev dimer provides an opportunity to explore a hybrid methods approach to determining the structure of polymerised Rev tubes. The power of hybrid methods—combining high-resolution X-ray crystallography data with medium-resolution cryo-EM data—to resolve the structure of large macromolecular assemblies has been widely articulated (e.g. see Steven and Baumeister (2008)).

Why determine the structure of *in vitro* assembled Rev filaments that form in the absence of RNA? It is a fair question, especially given that it is not known whether or not these filaments reflect a physiologically-relevant assembly. The following sub-sections outline the rationale and context for this study.

4.1.1 Insight into Rev-RRE assembly

Perhaps most obviously, a better understanding of how Rev subunits arrange into a large-scale macromolecular assembly may shed light on how Rev multimerises in the context of RNA binding. This notion is articulated notwithstanding two distinctly plausible limitations:

1. The Rev filaments observed *in vitro* may be purely artifactual, a consequence of high-purity storage in a test tube, as opposed to the more natural environment within the cellular milieu of HIV-infected cells.
2. The organisation of the Rev subunits in the RNP may be driven by interactions with the highly-structured RRE rather than by Rev-Rev interactions, rendering irrelevant the structure observed in the absence of RNA.

However, concern over these limitations is mitigated upon further review of the Rev literature. On the first point, Rev filament formation is driven primarily by a large negative enthalpy, occurs at a low concentration threshold of ~80 $\mu g\,mL^{-1}$, and the resulting assemblies are stable for years at 4°C (Watts et al., 1998). While in no way conclusive, the degree to which formation of this assembly is thermodynamically favorable is consistent with it reflecting a set of specific interactions that are physiologically-relevant.

Secondly, while no high-resolution structural data on Rev-RRE complexes has been reported yet, we do know from biochemical data that the association is highly cooperative and mediated primarily through protein-protein interfaces (Mann et al., 1994; Pond et al., 2009; Cook et al., 1991; Daugherty et al., 2010a). The residues at these interfaces have been biochemically characterised (Jain and Belasco, 2001; Edgcomb et al., 2008) and they map to the structurally determined *A* and *B* oligomerisation platforms (DiMattia et al. (2010); Daugherty et al. (2010b) and see **Chapter 3**). Therefore, even if

the RRE guides the Rev-RRE complex into an architecture that is profoundly different from that of the *in vitro* Rev filaments, the *A* and *B* oligomerisation faces will almost certainly remain involved. If these surfaces are also juxtaposed between Rev subunits in the filament assemblies, the *A-A-B-B-A-A* architecture observed may be related to the assembly that forms in Rev-RRE association.

4.1.2 Insight into Rev re-entry into the nucleus

Once Rev translocates its bound RNA from the nucleus into the cytoplasm, it releases the RNA and can return to the nucleus via the support of host cell karyopherin importin-β. Rev has a nuclear localisation signal (NLS)—coincident with the sequence of the RNA-associating ARM—which is used for re-entry into the nucleus. It is not known if Rev adopts a multimeric form for nuclear re-entry; however, there is cursory evidence that this is the case. Kubota et al. showed that a non-functional Rev mutant (dRev) containing a deletion in the NLS not only accumulated in the cytoplasm, but also inhibited co-expressed wildtype Rev from nuclear import (Kubota et al., 1992). Others have also shown similar cytoplasmic sequestration of wildtype Rev upon interaction with Rev NLS mutants (Szilvay et al., 1997), suggesting that hetero-oligomerisation occurs between mutant and WT in the cytoplasm prior to nuclear import. Though artificially long, the structure of the Rev filaments observed *in vitro* may reflect the macromolecular assembly that Rev forms in the cytoplasm to prepare itself for nuclear re-entry.

4.1.3 The *C-C* interface as a support brace?

The *A-A* and *B-B* interfaces are spatially offset along the length of the NTD helical hairpin (see **Fig 3.4**). As a result, propagation of Rev subunits oligomerisation—formed via the *A* and *B* faces—satisfies the necessary condition for the formation of helical geometry. A theoretical Rev assembly—constructed by

aligning the *A-A* and *B-B* oligomerisation interfaces determined in the Rev dimer crystal structures—is shown in **Fig 4.1**. This is but one example of helices that could form using the Rev *A* and *B* surfaces: altering the Rev subunit orientation relative to the helical axis, as well as the crossing angles between them, would create helices with different pitches. Further modeling using these oligomerisation interfaces suggested that it is difficult to make a one-start helical assembly with a pitch short enough for tighter packing between adjacent layers of Rev subunits.

However, it was subsequently observed that the *C-C* interface forms between Rev dimers in a way that would connect these theoretical *A-A B-B A-A* helical strands (**Fig 4.2**). Modeling of this interaction network suggests that multiple strands of *A-A B-B A-A* strands would be required to bring them in close enough register to form supporting *C-C* interfaces. As described in **Section 3.3.4**, the *C-C* interface has only been observed in crystal packing interactions in the majority of the scFv-Rev as well as the $Rev_{1\text{-}70}$ crystal structures (Daugherty et al., 2010b). If the *C-C* interface were to be observed in the context of Rev macromolecular assemblies, it would further suggest its importance in Rev multimerisation, possibly as a support brace in the manner suggested above.

4.1.4 The Rev NTD is insufficient for filament assembly

As discussed in **Section 3.3.5**, the Rev C-terminal domain is disordered in all of the scFv-Rev crystal structures determined thus far. Multiple groups have corroborated this CTD disorder when Rev is in solution (Auer et al., 1994; Daugherty et al., 2010a). Accordingly, a construct of Rev was made containing only the Rev NTD (residues 1-69) to test the *in vitro* phenotype of Rev absent its CTD. When viewed by negative-stain EM, this sample was observed to be incapable of making recognisable Rev filaments (**Fig 4.3**). This result suggests that the CTD, or at least a subset of it, is necessary for

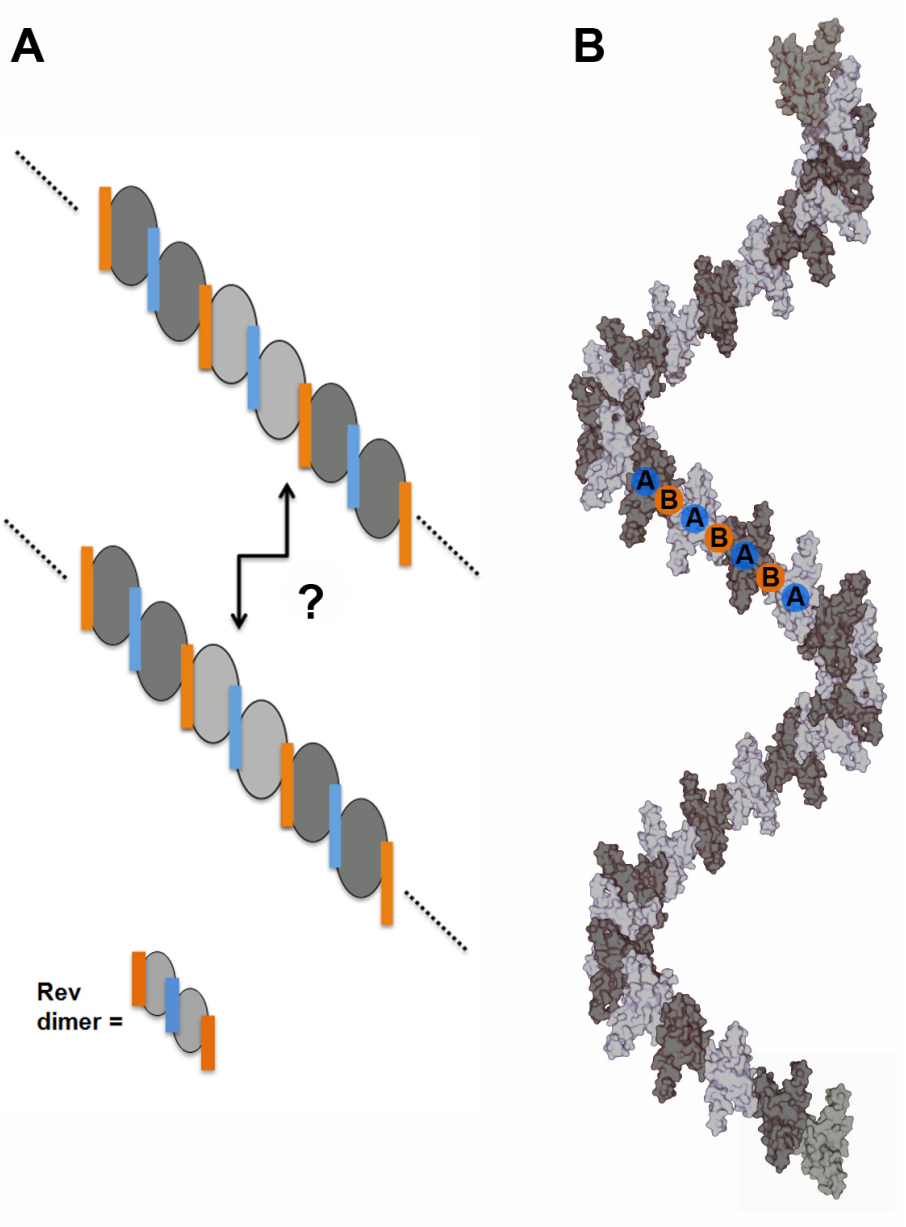


Figure 4.1: Theoretical Rev *A-A B-B A-A* assembly. **(A)** Cartoon schematic of a theoretical Rev assembly, constructed by aligning the *A-A* and *B-B* oligomerisation interfaces determined in the Rev crystal structures. **(B)** A three-dimensional model of the Rev *A-A B-B A-A* assembly. The crossing angles between the Rev subunits and the offset of the oligomerisation platforms (relative to each other) lend a natural helical curvature to theoretical Rev assemblies. Helical assemblies of Rev- and Rev-RRE filaments observed by EM do not appear to have the large gaps between helical strands as seen in this construct. Therefore, there is probably another interaction between Rev subunits to complete the interaction network. This unknown interaction is indicated by the question mark in (**A**).

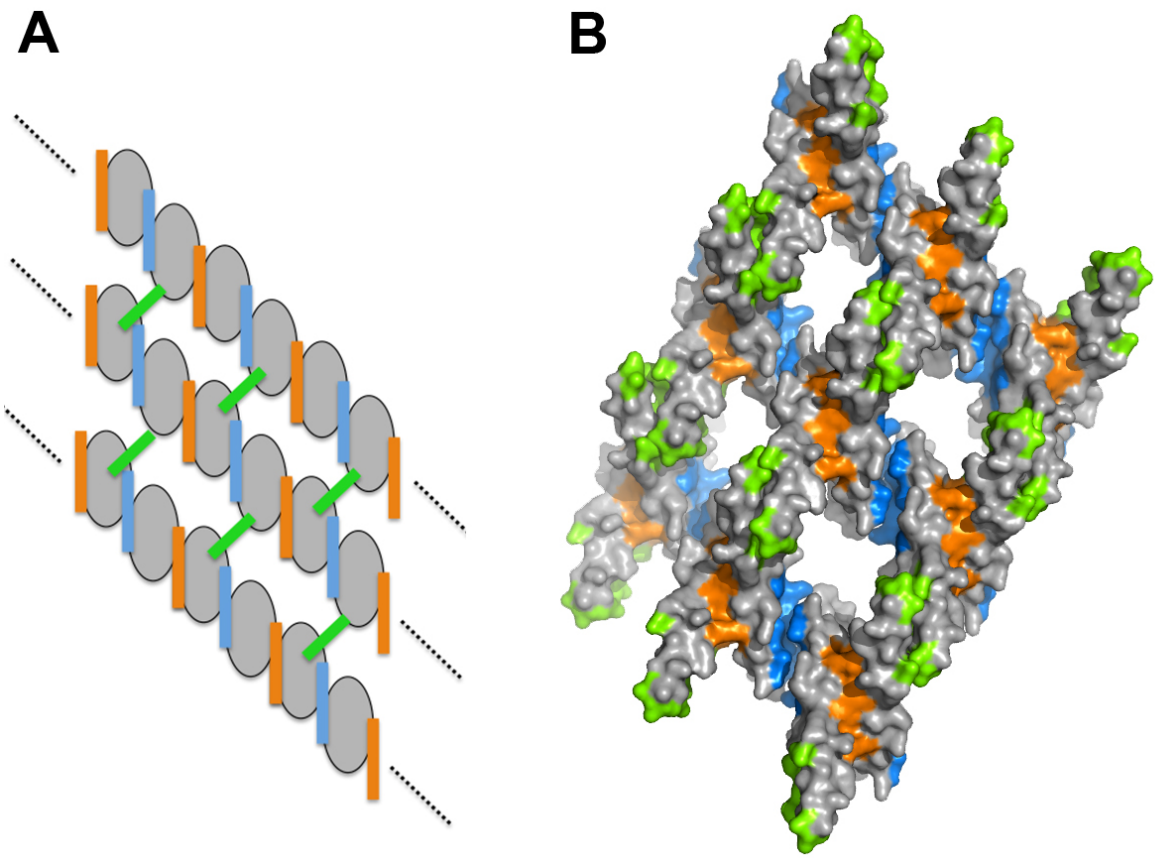


Figure 4.2: The *C-C* interface as a support brace. **(A)** Domain schematic showing modelling of how *C-C* interfaces can form in a way that would connect *A-A B-B A-A* strands. Modeling of this interaction network suggests that multiple *A-A B-B A-A* strands would be required to form a closed helical structure using the *C-C* interface. **(B)** Molecular model showing how the *A-A*, *B-B*, and *C-C* might jointly form to create a Rev assembled helical network.

filament assembly. It is possible that the CTD serves as a structural support that does not form part of the final filament architecture, akin to scaffolding proteins in virus capsid assembly. Another possibility is that the intrinsically-disordered CTD forms an ordered domain when Rev subunits are aligned in a specific way, e.g. in the context of filament assembly. Rev was also analysed through protein disorder prediction servers, such as RONN (Regional Order Neural Network) (Yang et al., 2005) and PONDR (Romero et al., 2001); however, the results did not indicate a higher probability of disorder for the CTD versus the NTD.

Raman spectroscopy on Rev filaments indicated the presence of ~50% α-helical and~20% β-sheet content (Watts et al., 1998). The Rev NTD comprises two α-helices (38% of full polypeptide), suggesting that the remaining secondary structure is attributable to the CTD. Altogether, these data may have ordered secondary structure (and likely tertiary structure) in Rev filament assemblies. Determining the structure of Rev filaments will answer these questions regarding environment-specific CTD ordering and the role it may have in larger-scale Rev multimerisation.

4.1.5 Rev filaments have heterogeneous width

Careful observation of micrographs of Rev filaments recorded under negatively-stained or vitrified ice conditions reveals that the filaments' diameter fluctuates, even within a particular tube. Preliminary visual analysis of cryo-EM micrographs indicated that the diameter can vary up to ~25% from one extreme to the other (**Fig 4.4**). This heterogeneity complicates structural investigation and probably contributes to the poor diffraction patterns observed for Rev filaments. However, it also provides a question of interest: what is the nature of the protein-protein interaction network that accommodates such conformational variability in Rev filaments?

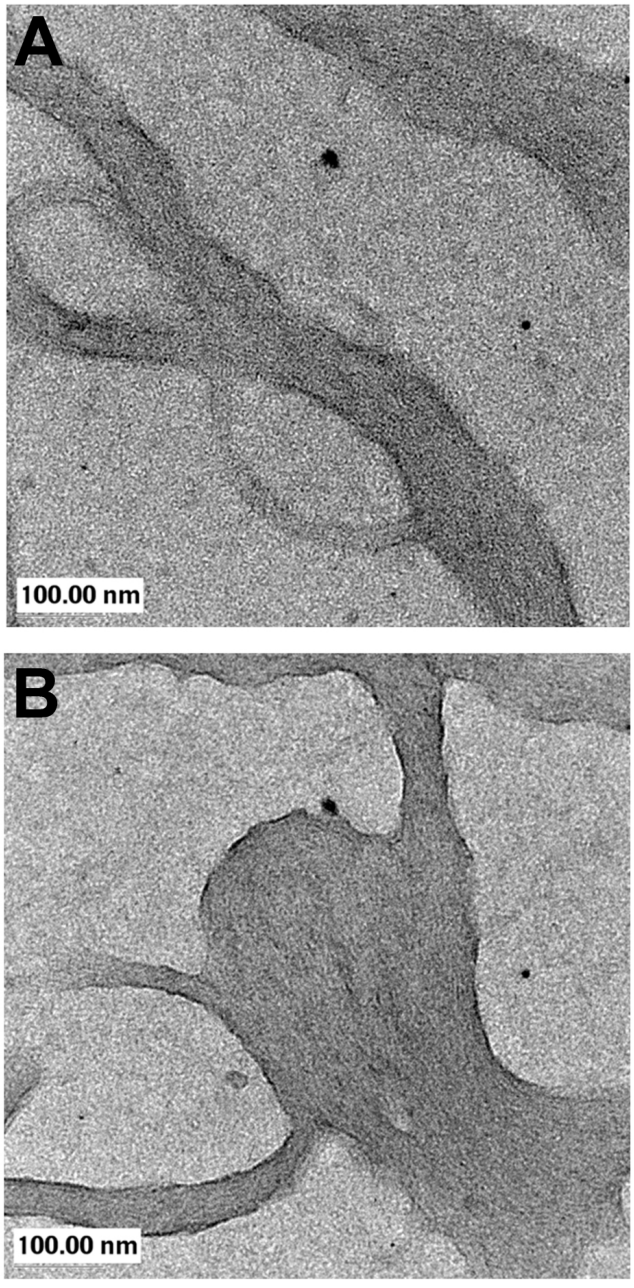


Figure 4.3: $Rev_{1\text{-}69}$ does not form filaments *in vitro*. (A and B) Two example images of negative-stain electron micrograph of Rev NTD (residues 1-69). No recognisable Rev filaments were observed, suggesting that the CTD is necessary for *in vitro* filament formation.

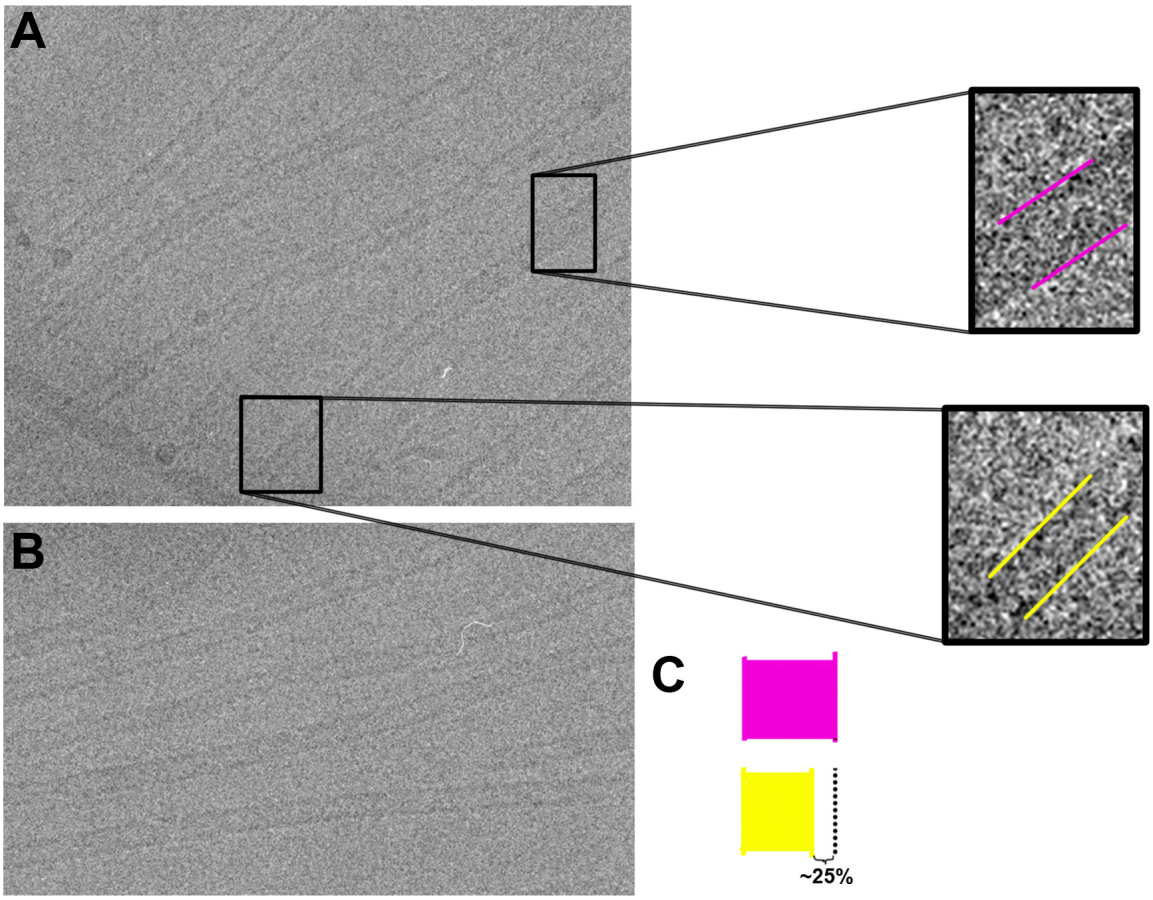


Figure 4.4: Heterogeneous diameter of Rev filaments. **(A and B)** Panels show cryo-electron micrographs of Rev filaments. Two insets reveal close-up view of filament sections with variable width. Magenta and yellow bars demarcate the filament section edges in each. **(C)** Transposition of the filament edges into the same frame shows that there is ~25% difference in diameter between the two sections. Notably, these sections are from the same filament and as such, it appears that filament width fluctuates within a given tube (with the assumption that the Rev filaments remain in-plane, that is, lie flat on the grid surface).

4.2 Data Collection and Particle Classification

4.2.1 Negative stain vs. cryo electron microscopy

Prior to visualising Rev filaments in the transmission electron microscope, the preparation of the Rev filament sample was optimized for data collection vis-a-vis the intended objective of the experiment, in this case a set of helical reconstructions at sub-nanometer resolution. The two principal specimen-preparation techniques for such an experiment are termed negative staining and plunge freezing. Negative staining involves applying a heavy metal salt solution that forms an electron-dense mould around individual macromolecules that, when imaged, gives rise to high contrast and a good signal-to noise ratio, particularly useful for smaller molecules. The resulting images are more radiation-resistant; however, the use of heavy metal stains induces non-native conditions on the specimen, often resulting in flattening and distortions of molecular structure. The resolution limit of image reconstructions from negatively-stained particles is ~20 Å. Conversely, plunge-freezing involves preserving the sample in its native conformation by rapidly freezing in liquid ethane such that crystalline ice does not form, preventing damage to the sample. The disadvantages to cryo-conditions are that the sample is much less resistant to radiation, the recorded images have a much lower signal-to-noise ratio, and also have lower contrast. For Rev filaments, it was decided that cryo-electron microscopy would be preferred in order to achieve higher resolution, permitting map features to be resolved that would discern the position and orientation of individual Rev dimers within the filament. Being that the helical tubes are long, their presence is visible even at low contrast levels that would otherwise occlude observation of smaller macromolecules.

4.2.2 Principal component analysis

From the cryo-electron micrographs, Rev filaments were cut into segments using a box size of 256 x 256 pixels (~470 x 470 Å). Segments were selected with 90% overlap between adjacent boxes in order to aid particle alignment. In total, 30,426 overlapping segments (the equivalent of ~3,000 unique particles). Principal component analysis (PCA) was used to separate the filament segments and classify them into homogeneous bins (see **Section 2.2.3** for methods). PCA uses inter-image variance to separate. Depending on the particular eigenimages (or principal components; PCs) and separation thresholds that are selected for hierarchical classification, distinct binning schemes can be produced using the same input data. An example separation scheme is shown in **Fig 4.5**. Common structural features were observed no matter which classification scheme was implemented. In particular, a distinctive "herring-bone" feature was noted, with a zigzag appearance along the exterior. Furthermore, the predominant visual distinction between the classes appeared to be filament width, with variation in diameter of up to 30 Å. Notably, no criterion specifying separation by diameter was incorporated into the classification algorithm: the inter-image variance used to calculate the eigenimages resolved the diameter flux itself. Therefore, it is suggested that the observed diameter differences are real and not an artifact of classification. This is consistent with visible diameter differences in the cryo-EM micrographs themselves (**Fig 4.4**). The herring-bone pattern seen in the class averages is ubiquitous: present in both the thinnest and widest, as well as in the intermediate-sized filament classes. This may be indicative of an underlying similarity in the molecular arrangement of the assembly.

A limitation of particle classification by PCA is that there is no objective way to determine which eigenimages to use for classification, and therefore how many bins are needed to represent the full breadth of actual variation in one's biological specimen. The issue is compounded by the fact that it

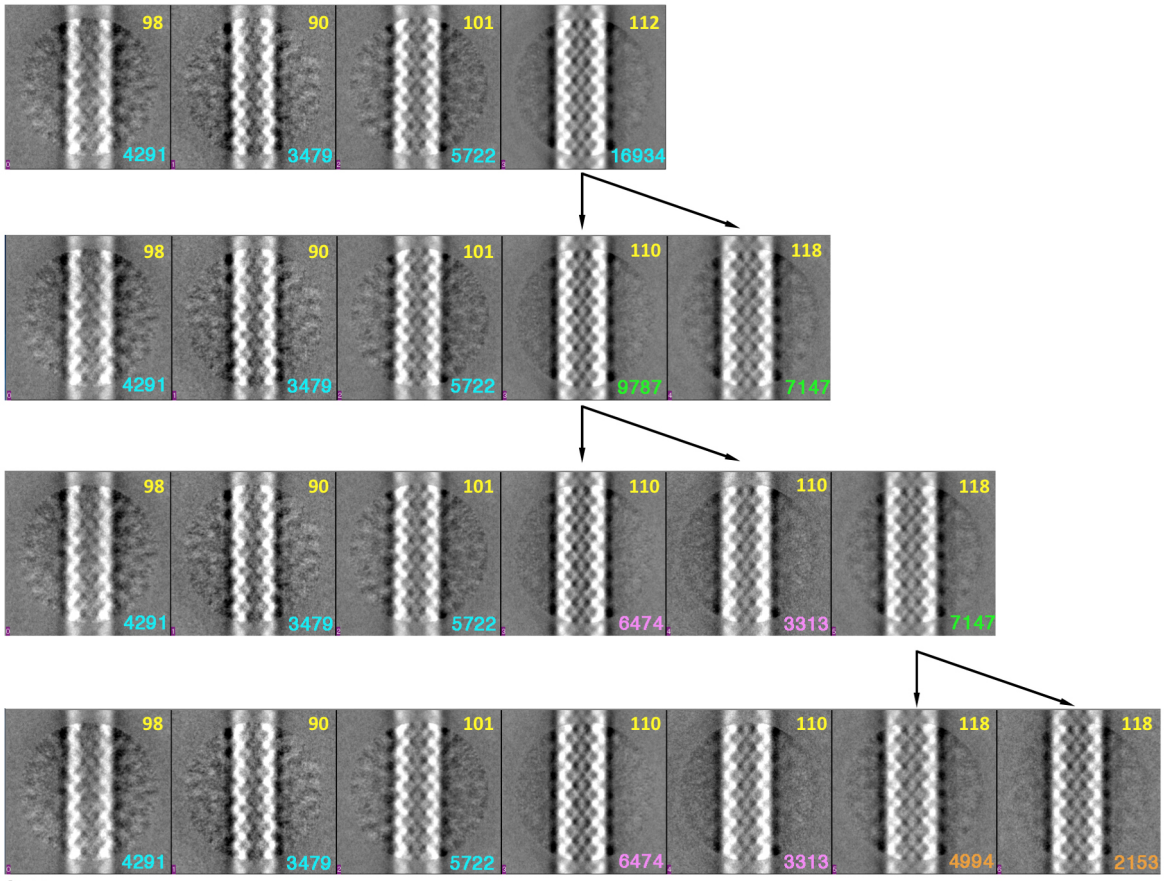


Figure 4.5: Example classification scheme by PCA. Example binning of Rev filaments using principal component analysis (PCA). Shown are two-dimensional averages for each bin. Yellow numbers are the peak-to-peak diameter for each class in Å. Colored numbers in lower right of each class average represent the number of particles in the class. See text for additional details.

is not known whether Rev filament heterogeneity is polymorphic (adopting a set of discrete conformations) or pleomorphic (adopting a continuum of conformations). In the case of Rev filaments, additional requested bins resulted in the observation of additional tube diameters (although this phenomenon appeared to diminish as bins were split into further sub-bins). Despite this complication, all size classes were not observed equally. For the scheme shown in **Fig 4.5**, the first three PCs representing the most inter-image variance were used for classification (see **Fig 2.7**). Four filament diameters were most frequently observed (measured peak-to-peak; see **Fig 2.8**): 90 Å, 100 Å, 110 Å, and 118 Å **Fig 4.5**.

4.2.3 Extensive binning of Rev filaments

In order to increase the likelihood that the particles were classified into homogeneous bins, a final PCA scheme was used in which the data were separated using eigenimages representing >99% of the inter-image variation (using the first nine PCs, coloured red in **Fig 2.7**). The resulting hierarchical classification gave rise to 26 bins, a number almost certainly greater than the observable variation at a sampling rate of 1.84 Å pixel^{-1} (**Fig 4.6**). This inference is supported by the observation that a subset of the resulting bins contain a small number of particles (< 300) relative to the full data set (~30,000), i.e. the number of bins into which the data were classified exceeded the inter-image variation within the data (**Fig 4.6**). The value in having homogeneous (or nearly homogeneous) bins is that each subset has a greater probability of converging to the correct three-dimensional structure during the subsequent image alignment and reconstruction process. An additional goal to performing the "over-classification" was to determine the full spread of conformational variability within the Rev tubes, as well as whether the variation is discrete or continuous.

4.2.4 Consolidation of bins into final Rev filament size classes

Two variables were used to identify which bins could be merged into larger classes while maintaining near-homogeneity (100% true homogeneity is close to impossible given the variability in the data):

1. peak-to-peak diameter measured using a transverse density plot and;
2. converged helical parameters upon reconstruction of the particles in each class

The results of these calculations are presented for each of the 26 classes in **Fig 4.7**. The top and bottom graphs show each class plotted versus subunit rise and diameter, respectively. The helical parameters for each class can be represented by subunit rise alone, as subunit rotation remains a constant 30° for all classes investigated. The spread of diameter values suggests that the conformational variability of Rev filaments is more of a continuum than one of discrete states. However, upon juxtaposing the plot of the helical rise for each class alongside the plot of diameter values, the classes seem to cluster into at least three, and possibly up to five discrete populations. More importantly, there is a direct correlation between helical symmetry and filament diameter: as the tube width increases, so does the subunit rise.

Given this analysis, the bins were recombined into five final Classes (termed I-V) based on similarity of helical symmetry and diameter (**Fig 4.7**). It is possible that Classes I-V do not fully sample the range of conformations that occur, particularly if Rev filaments are indeed pleomorphic and not polymorphic; however, the goal is to simply sample a representative range of conformations. The range of filament diameters and corresponding helical symmetries observed, as well as the particle spread among the bins, suggest that the final recombined Classes represent the range of conformational variability afforded by Rev filaments.

4.3 Image Reconstruction

4.3.1 Subunit rise increases with filament diameter

The iterative helical real space reconstruction (IHRSR) algorithm was applied to cryo-EM images of frozen-hydrated Rev filaments (for full procedures see **Chapter 2**). Five Rev filament cryo-EM reconstructions of varying diameter, termed Class I through V, were generated to an estimated resolution of ~12-14 Å (FSC, threshold 0.50) (**Fig 4.8**) (van Heel and Harauz, 1988). These assemblies have six-fold rotational symmetry (C6) as well as helical (screw) symmetry. The azimuthal rotation from one subunit to the next was consistently 30° for all filament sizes; however, the subunit rise increased from 26.6 Å to 34.0 Å from Class I to Class V filaments, respectively.

4.3.2 Structural features of Rev filaments

All of the Rev filament reconstructions are marked by a thin wall with a large hollow core. The filament reconstructions range in diameter from 115 Å to 145 Å, increasing in steps of 5 Å between adjacent Classes (except for the increase of 15 Å in diameter between Class I and II). However, color coding by radial distance from the filament axis (**Fig 4.9**) reveals the remarkably different surface features between the filaments of differing width. The exterior surface of Class I filaments are decorated with bi-lobed mounds (one mound per subunit), which are flanked by deep channels (**Fig 4.9**) that run in a criss-cross fashion around the tube. In Class II filaments, the external surfaces features are not mounds, but rather form thinner ribbed protrusions separated by a wider and more shallow channel. A similar progression is observed for Class III-V: the protrusions become farther apart and fade into the filament wall, concurrent with the interspersed channels becoming more shallow. Remarkably, these trends take

an extreme form in Class V filaments, having a relatively unfeatured surface with flat sheets that wrap around the cylindrical surface instead of any discernible protrusions or channels. Reconstructions of Class IV and V filaments are somewhat less definitive due to the lack of surface features that would otherwise aid in better determination of particle orientations. Nonetheless, Classes I to III are fully sufficient to demonstrate the structural changes that take place between different size Rev filaments.

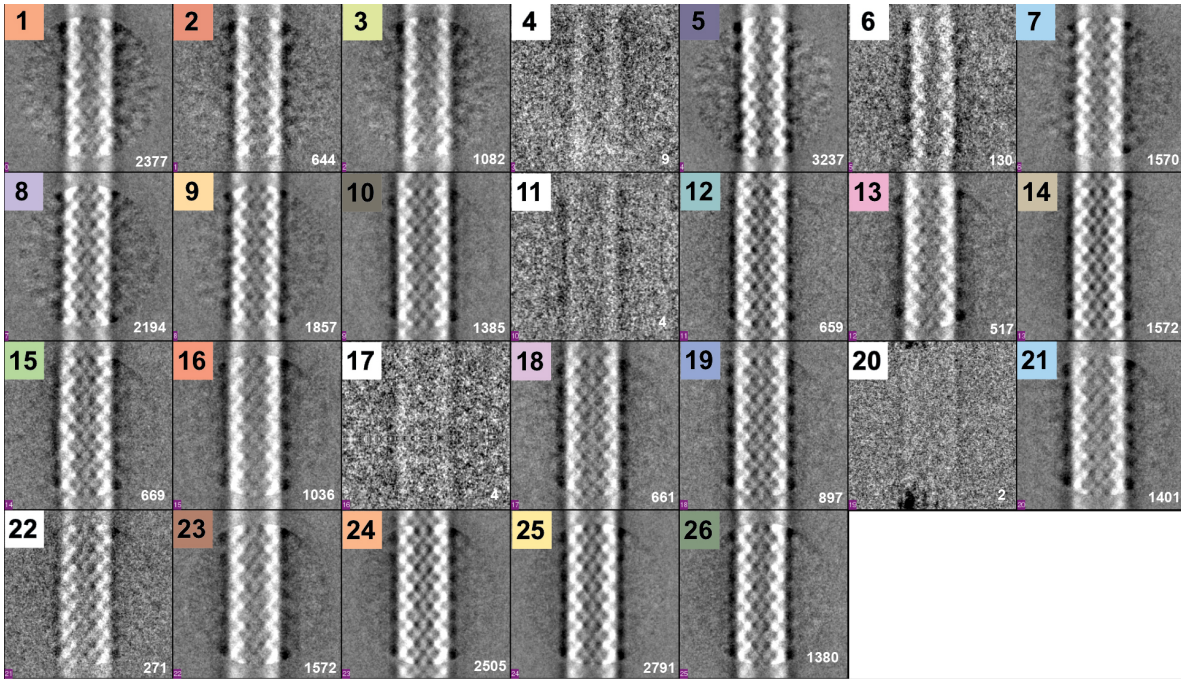


Figure 4.6: Final classification scheme. Shown are the final 26 bins into which the Rev filament segments were classified. Classification parameters were chosen to generate many classes, with the intent to use virtually all of the inter-image variation. The numbers (in white) correspond to the number of particles in each bin.

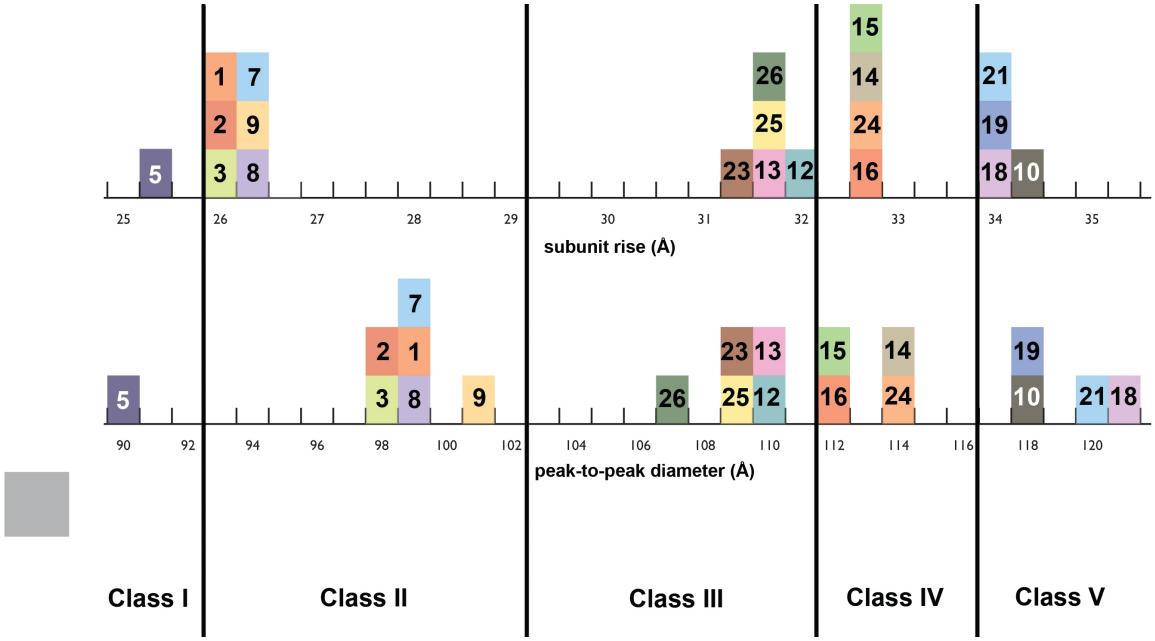


Figure 4.7: Characterisation and consolidation of bins. Each of the 26 bins (see **Fig 4.6** are plotted vs. helical subunit rise (top; in Å) and vs. diameter (bottom Å). There is a correlation between the helical symmetry of each bin and its calculated peak-to-peak diameter. This analysis was used to consolidate the bins into five classes, termed Class I-V. The vertical lines demarcate how the bins were consolated. The number of particles per class for I-V are 3237, 9723, 6919, 5782, and 4344, respectively.

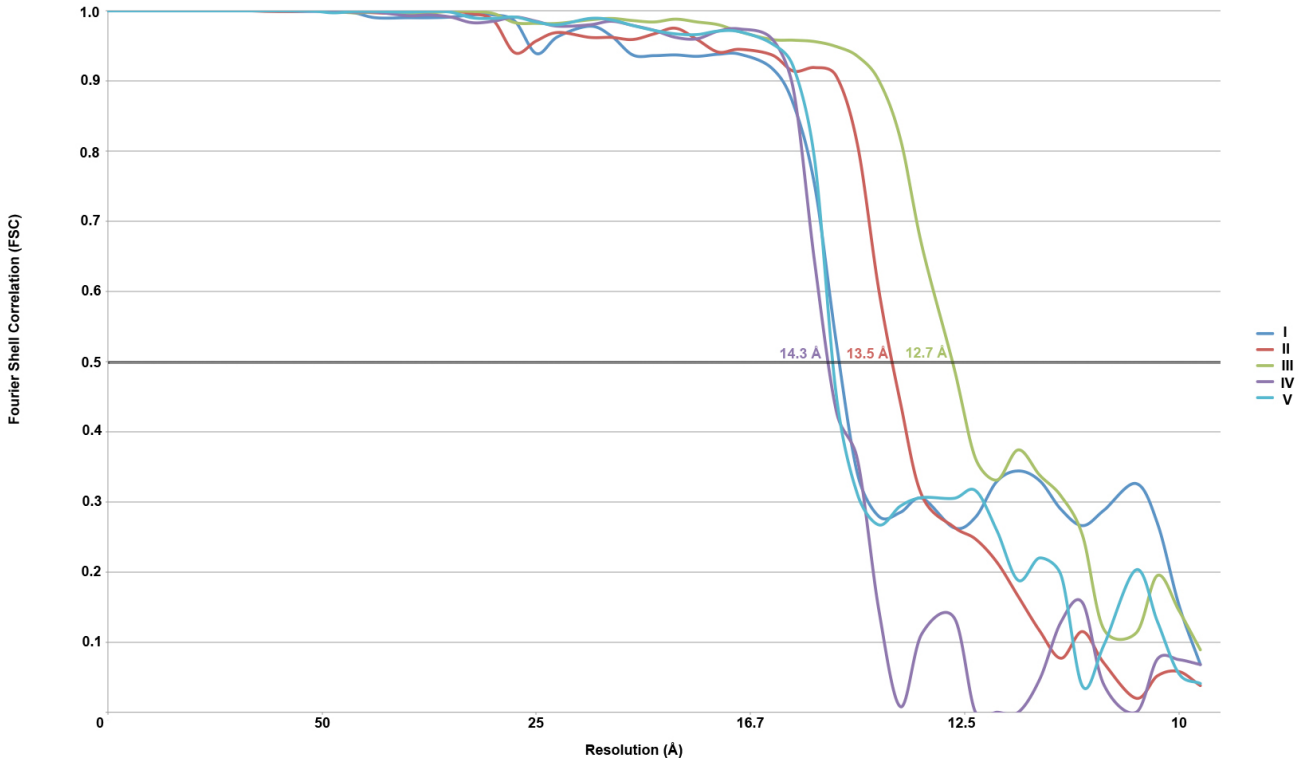


Figure 4.8: Fourier Shell Correlation Plot. Resolution estimation using Fourier shell correlation (FSC) coefficient values between two half-data sets plotted versus spatial frequency. The resolution was taken to be where the correlation coefficient dropped below 0.5, indicated by the horizontal line. The gold standard method of resolution estimation was not used in this analysis; therefore, the half set reconstructions are not fully independent. Each color represents a different Class (I through V).

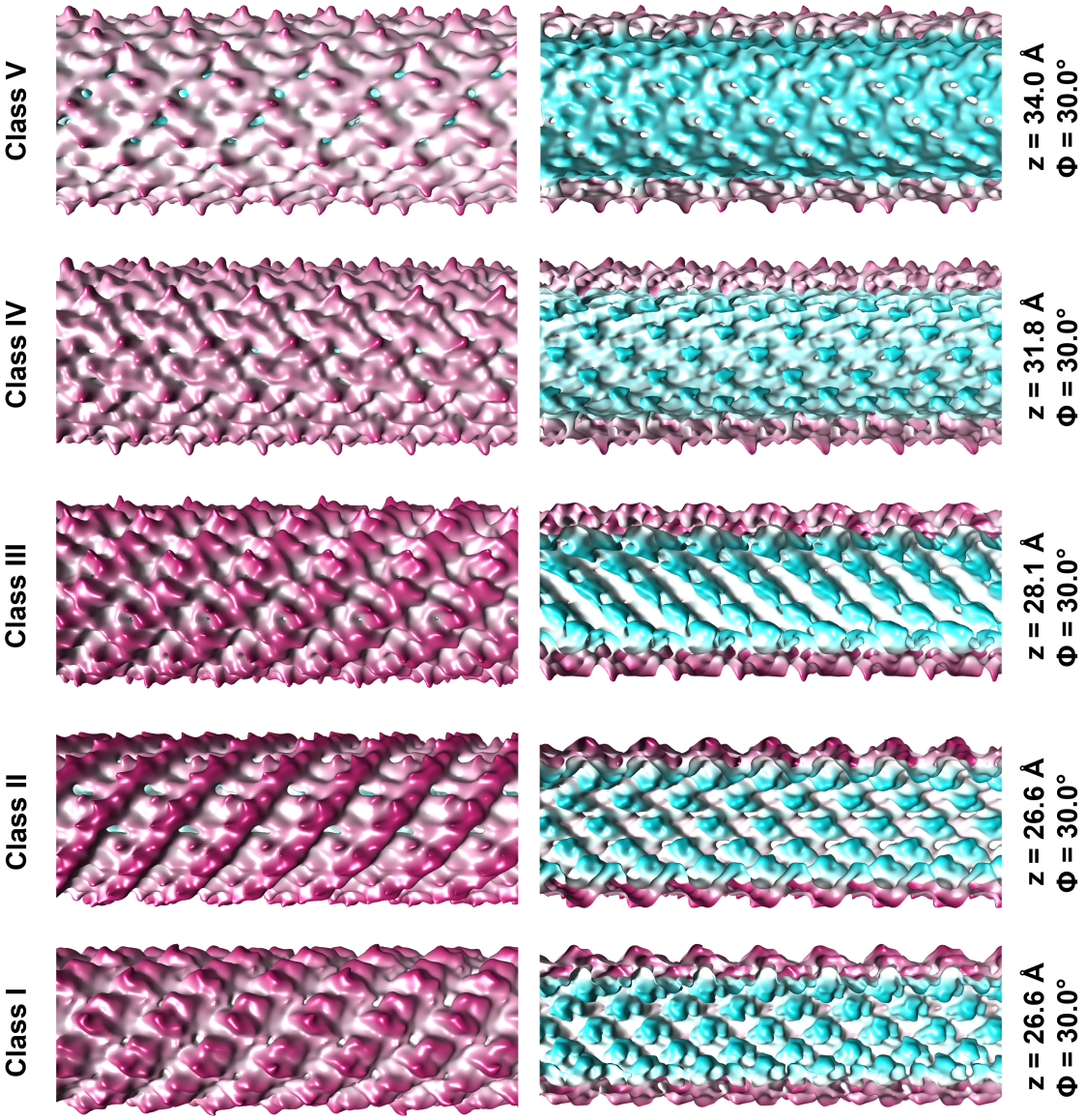


Figure 4.9: Reconstructions of HIV Rev filaments. Shown are radially-colored isosurface renderings of the Rev filament reconstructions, one for each size class (Class I-V). The top row depicts the outer filament surface and the bottom are cross-sections that reveal the filament interior. The helical parameters to which each Class converged during reconstruction are also shown (subunit rise (z); subunit rotation (Φ). The maps are filtered to 14 Å resolution.

Interestingly, the interior features follow the same trend as the exterior. Despite being very thin in certain regions, Class I filaments have bulbous protrusions on their interior wall that extend toward the hollow center. Comparison of these inner protrusions as the filament width increases reveals that the protrusions become significantly less prominent. Again, upon reaching Class V, the protrusions have receded into the interior wall, leaving a relatively smooth inner surface in its absence (**Fig 4.9**).

These concerted conformational changes as the Rev filament expands are evocative of an assembly expansion taking place, akin to the large-scale molecular re-arrangements that take place during capsid maturation of many bacteriophages (e.g. see Ionel et al. (2011)) and contraction of bacteriophage tail sheaths (e.g. see Aksyuk et al. (2011)). The central sections of the volumes, shown both parallel and perpendicular to the filament axis in **Fig 4.10**, provide another perspective on the putative expansion process. These views reveal the cross-sections of the filament wall for each class, and comparison of the conformational changes that take place as Rev filaments change diameter can clearly be observed. The angular, V-shaped pattern of the filament wall in the thin Class I filaments progressively becomes more smooth as the filament widens, The expansion of Rev filaments can also be seen in central sections that are perpendicular to the filament axis ("top down" view). The inner and outer protrusions and grooves become less pronounced as the filament tube expands, analogous to the expansion process that takes place during bacteriophage prohead expansion and maturation.

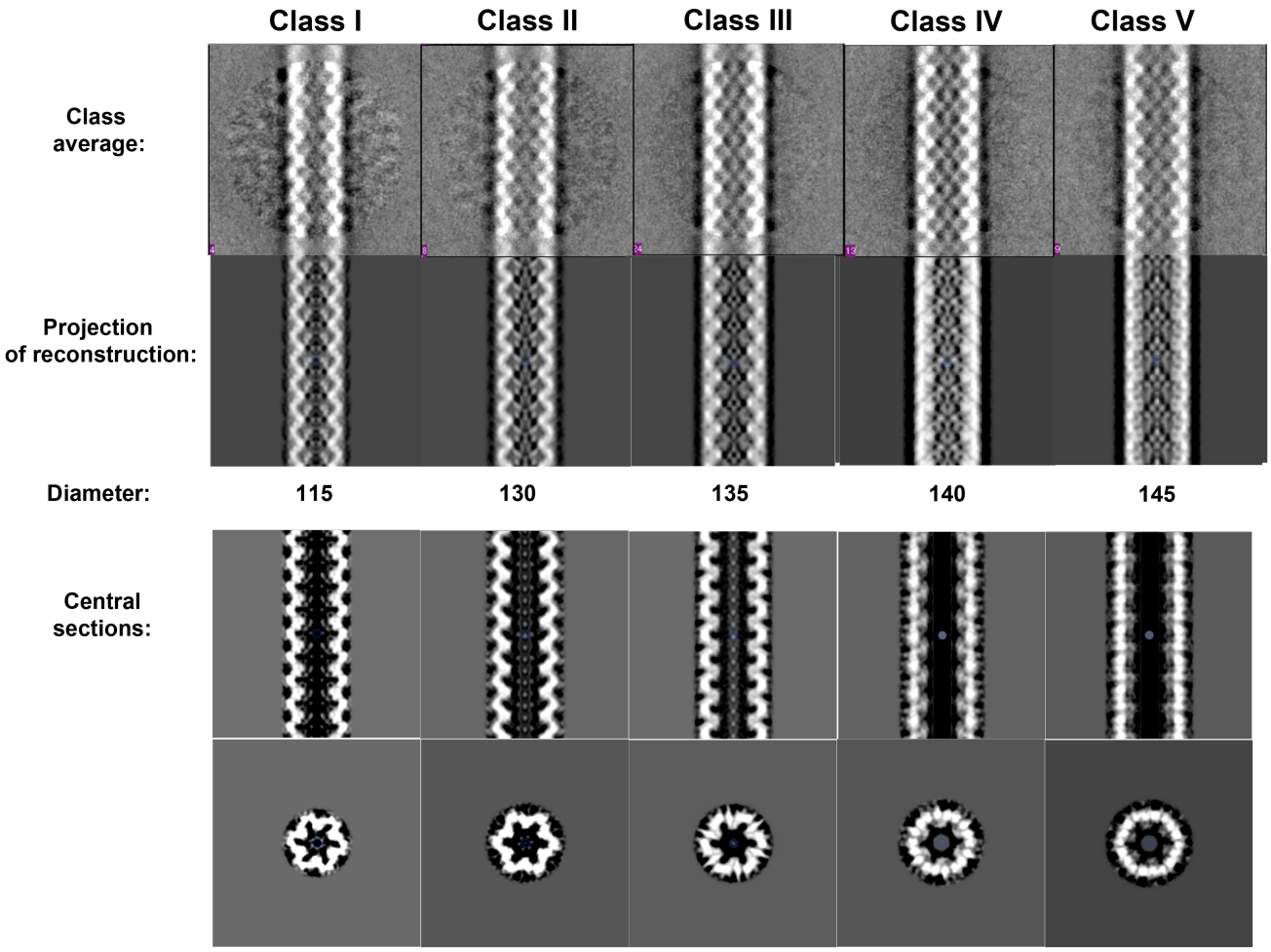


Figure 4.10: Rev Filament Re-Projections and Central Sections. Each column represents one size class, from Class I (left) to Class V (right). Top row: Class averages of the particles. Second row: re-projections of the reconstruction, resolution unfiltered. Filament diameters are in Å. Third and Fourth rows: Central sections of each reconstruction along z- and y-axes. Filament expansion can be seen as features become less pronounced with increasing filament diameter.

4.4 Pseudo-atomic model of Rev filaments

4.4.1 Molecular docking

The resolution of the reconstructions is sufficient to perform a preliminary docking of the Rev dimer crystal structures in order to obtain a pseudo-atomic model of Rev filaments. An example of the features present in the maps is shown with the Class II reconstruction in **Fig 4.11A** at a high σ threshold. Globular densities were observed connecting to strands of density that circumnavigate the filament. Prior to molecular docking, the repetitive, parallel fashion of the globular densities was immediately evocative of Rev subunits multimerised in an *A-A B-B A-A* fashion, as modelled in **Fig 4.1**. It must be noted that only approximately half of the Rev protein—the N-terminal domain for which we have a crystal structure—has been modelled into the filament reconstructions. The full details of how the Rev dimer subunits were docked into the reconstructions—including an objective approach free from any preconceived idea about how Rev subunits dock—are given in **Section 2.2.6**. This approach also eliminated the ambiguity of which half of the map to attribute to the NTD vs. the CTD (for which there is currently no model). The Rev NTD subunits fit into only one of the two reconstruction enantiomorphs, where they are related by a 26-34 Å rise (depending on the size class) and a 30.0° azimuthal rotation along a left-handed six-start helix.

In the docking arrangement shown in **Fig 4.11B**, the α-helices of Rev NTD subunits are in approximate alignment with the filament axis, each fitting into a successive globular density (**Fig 4.11C**). This fit allows Rev subunits to propagate via an *A-A B-B A-A* interaction network, continuing indefinitely around the filament axis. The *A* and *B* oligomerisation platforms are oriented toward the center of the filament, with *A-A* interfaces occupying the inner protrusions (**Fig 4.11D**). The electron density

is stronger at the *A-A* interface than at the *B-B* interface, consistent with molecular dynamics studies indicating that the former is a stronger protein-protein interaction than the latter **Fig 4.11C** (Venken et al., 2012). The *C-C* interfaces forms as well, connecting the *A-A B-B A-A* strands to complete the helical lattice.

4.4.2 Molecular rearrangement among the different Rev filament states

Using rotational and helical symmetry constraints, Rev subunits were docked into all five Rev filament sizes (Class I-V) (**Fig 4.12**). Notably, the *A*, *B*, and *C* interfaces are formed in each state. The Rev subunit crossing angles vary as needed so that the subunits can engage each other despite large-scale expansion and contraction of the filament. The channels on the exterior of the filament widen and become more shallow as the filament diameter increases (going from Class I to Class V). Upon docking of Rev subunits into the different reconstructions, it is apparent that the V-shape opening of *A-A B-B A-A* oligomerised Rev subunits corresponds to the V-shape contour of these exterior grooves (**Fig 4.12**). Furthermore, the subunit rise increases from 26.6 Å to 34.1 Å with increasing filament width, while the subunit rotation remains a consistent 30°, suggestive of expansion and contraction along the filament axis without alternation of the twist of the helix. Such helical changes are consistent with the relative opening and closing of Rev dimer subunits along the filament axis itself. Therefore, it was hypothesised that widening of the grooves may be consistent with progressively more obtuse *A-A* and *B-B* crossing angles.

Subsequent to molecular docking, the Rev dimer *A-A* crossing angles were calculated for the docked subunits in each Class. Filament width and subunit crossing angle are positively correlated (**Fig 4.13**). The most acute *A-A* interface crossing angle, 115°, corresponds to the thinnest Class I filaments, and is

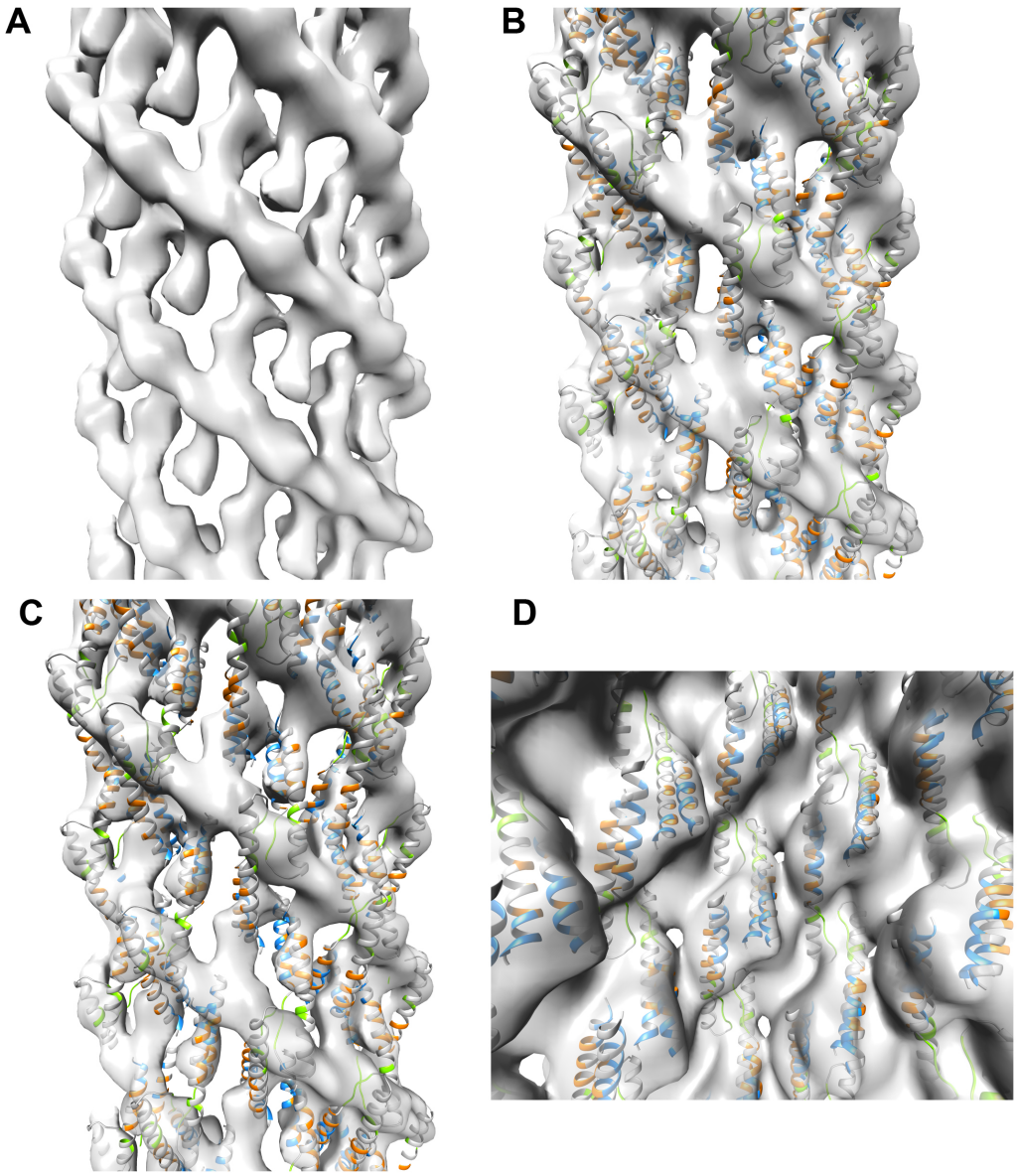


Figure 4.11: Molecular docking of Rev subunits. **(A)** Isosurface of Class II Rev filament, rendered at σ threshold of 3.0 to show contours of electron density. **(B)** Fit of Rev dimer subunits into Class II filament. *A*, *B*, and *C* faces are colored in blue, orange, and green. The *A*-*A* *B*-*B* *A*-*A* interactions are formed, as well as *C*-*C* interfaces that connect the *A*-*A* *B*-*B* *A*-*A* strands laterally. (σ threshold: 1.2) **(C)** Same molecular docking as in (**B**), but rendered at σ threshold of 3.0 (same as **A**) **(D)** Fit of filament interior surface. Inner protrusions are occupied by the oligomerisation motifs, centered around the *A*-*A* interface. (σ threshold: 1.2)

consistent with the most acute angle observed in the Rev dimer crystal structures. The crossing angle for the intermediate three Classes (II-IV) hovers near 140±5°. This is the most obtuse angle observed in the Rev crystal structures. Interestingly, in the most expanded Class V filaments, the crossing angle opens up to 165°. Consequently, the grooves seen in the thinner filaments have transformed into wide, flat platforms that wrap around the expanded filament.

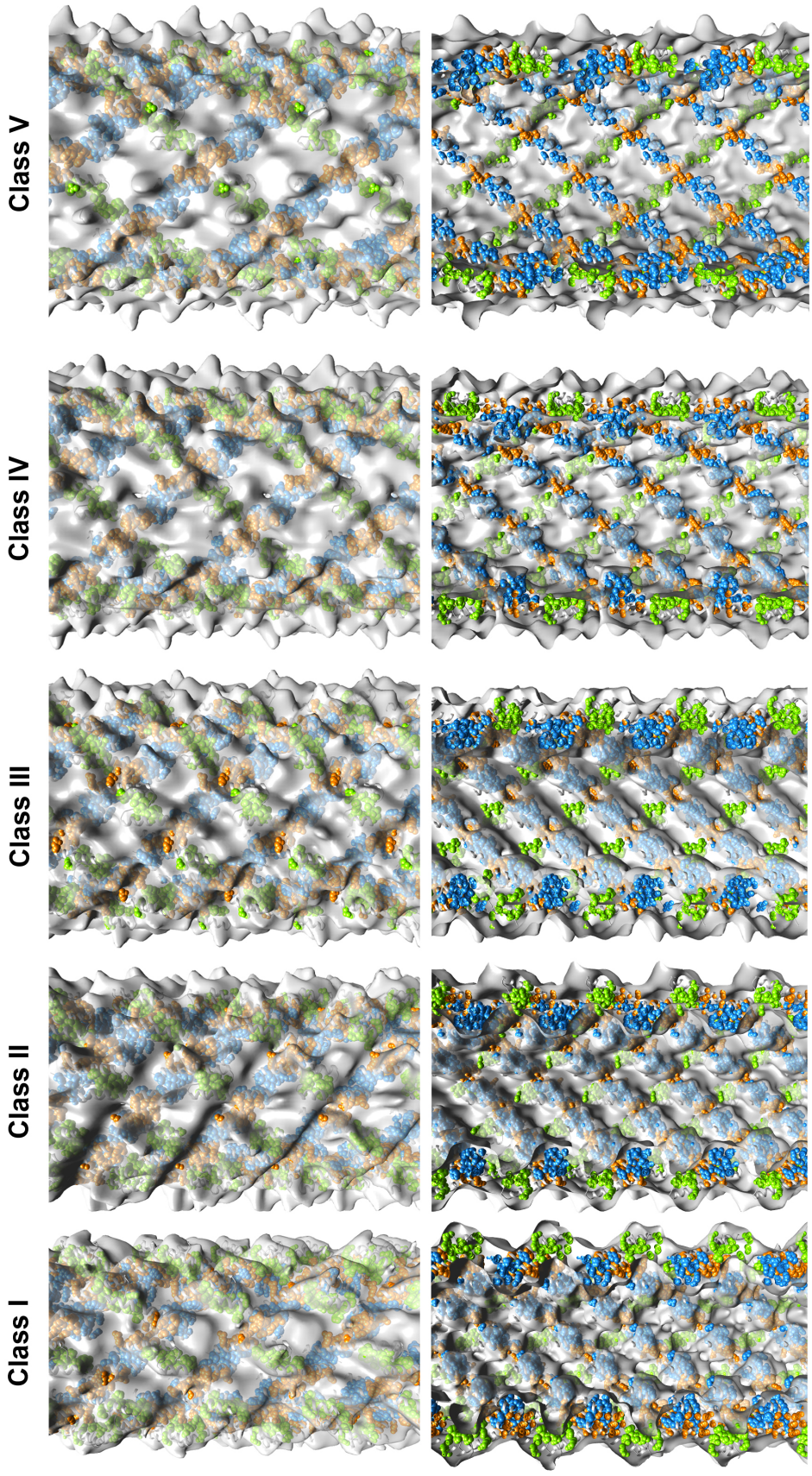


Figure 4.12: Rev filament expansion and contraction. Shown are translucent isosurface renderings of the different size Rev filament reconstructions (Class I-V) with Rev subunits docked into each reconstruction. Oligomerisation interface residues are shown in colored spheres (standard color scheme: *A*: blue, *B*: orange, *C*: green.

4.4.3 CTD-associated density is present in difference maps

The Rev crystal structure accounts for the NTD, which is approximately half of the Rev polypeptide. Difference maps were calculated for each Class by subtracting the calculated structure factors (corresponding to the docked Rev NTD assembly) from the experimental cryo-EM maps. These maps are shown in **Fig 4.14** with the positive difference density colored in gold. The extra density is attributed to the CTD, unobserved in the crystal structures due to intrinsic disorder. The CTD appears to occupy a gap in between successive *C*-*C* interfaces. The contour and volume of the CTD-associated density is different in each size class (**Fig 4.14**). This could be suggestive of the CTD forming different sub-structures in different molecular contexts, as has been previously characterised for intrinsically-disordered domains (Fong et al., 2009).

A close-up view of the CTD-associated density shows that its solvent-accessible surface is largely on the exterior of the Rev filament (**Fig 4.15**). Interestingly, the ordered C-termini of two Rev NTD subunits (residue 65; shown in Red in **Fig 4.15**) co-localise at the positive difference density. The arrangement is evocative of the two subunits, which are from different *A*-*A* *B*-*B* *A*-*A* strands, contributing their CTDs in a way that joint folding of a new domain takes place. Fong et al. describe how symmetrical arrangements of homodimers can be essential for "keeping functional disordered regions close together in space to form joint binding interfaces or to form near-interface regions to regulate the accessibility of the binding partner" (Fong et al., 2009). In other words, it is possible that the Rev CTD only forms an ordered domain when Rev NTD subunits are in proper oligomerised register via its *A*, *B*, and *C* interfaces. Relatedly, it has been shown that self-interactions between disordered regions in homodimers can be of evolutionary and functional importance (Simon et al., 2008; Andreeva and Murzin, 2006).

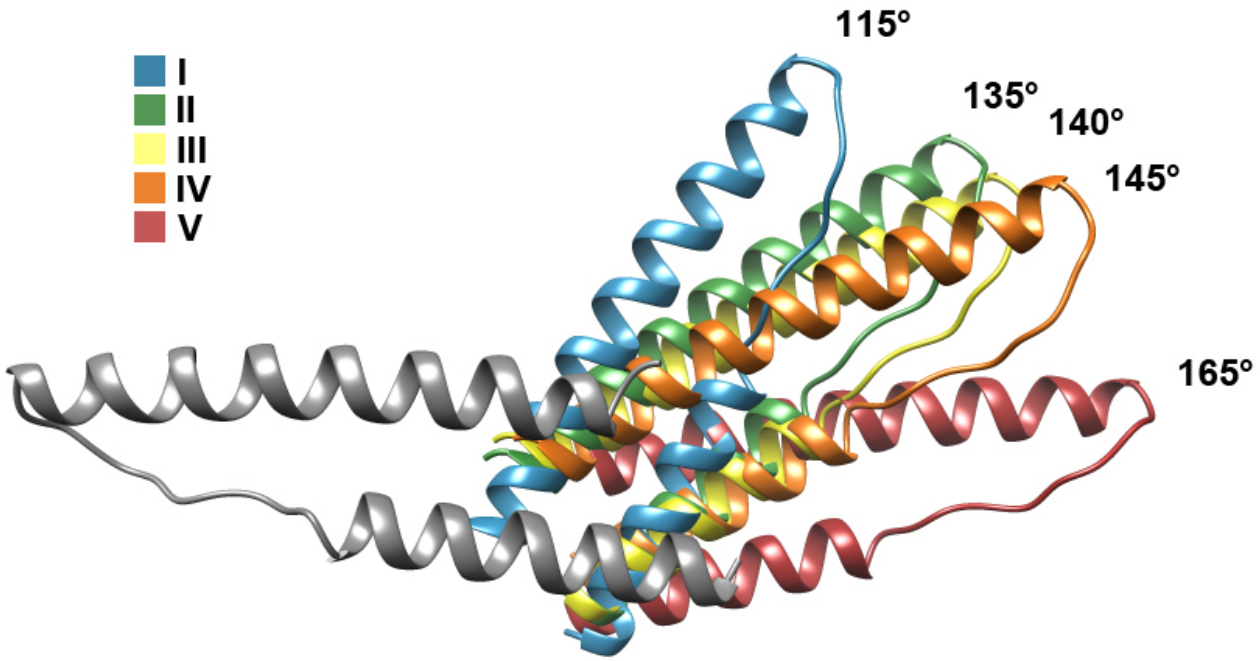


Figure 4.13: Conformational variability of the Rev dimer in filaments. Shown are the different Rev dimers (*A-A* interface) that fit best into the EM maps of the different size Rev filaments, superposed on the fixed Rev monomer shown in gray. The crossing angles are a result of a combination of real space refinement and manual docking of the monomers into EM electron density. The conformational variability seen is similar to that observed in the different Rev dimer crystal structures (see **Chapter 3**). The filament diameter correlates with Rev dimer crossing angle (increasing diameter is proportional to larger crossing angle.

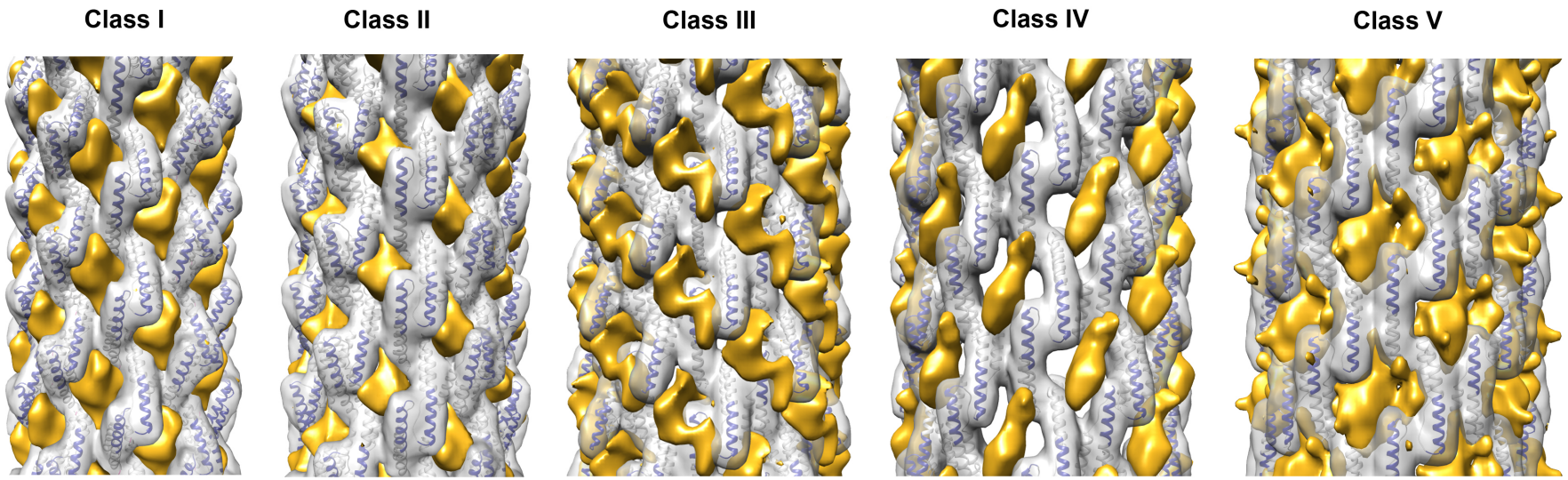


Figure 4.14: Difference maps. Shown are Fo-Fc difference maps for the reconstruction of each size class (Class I-V). Structure factors for the Rev NTD model were subtracted from the experimental cryo map, leaving the gold difference density. This density is attributed to the CTD, for which there is no model. The gold density appears to have different surface features between the different Classes, but its location in the filament is constant. By setting the σ threshold to the level used for docking of the NTD, the volume of the CTD-associated densities in each reconstruction can be used to estimate what percentage of the CTD is ordered. However, before such calculations can be meaningful, higher resolution data is needed to ensure the orientation for each NTD is optimal.

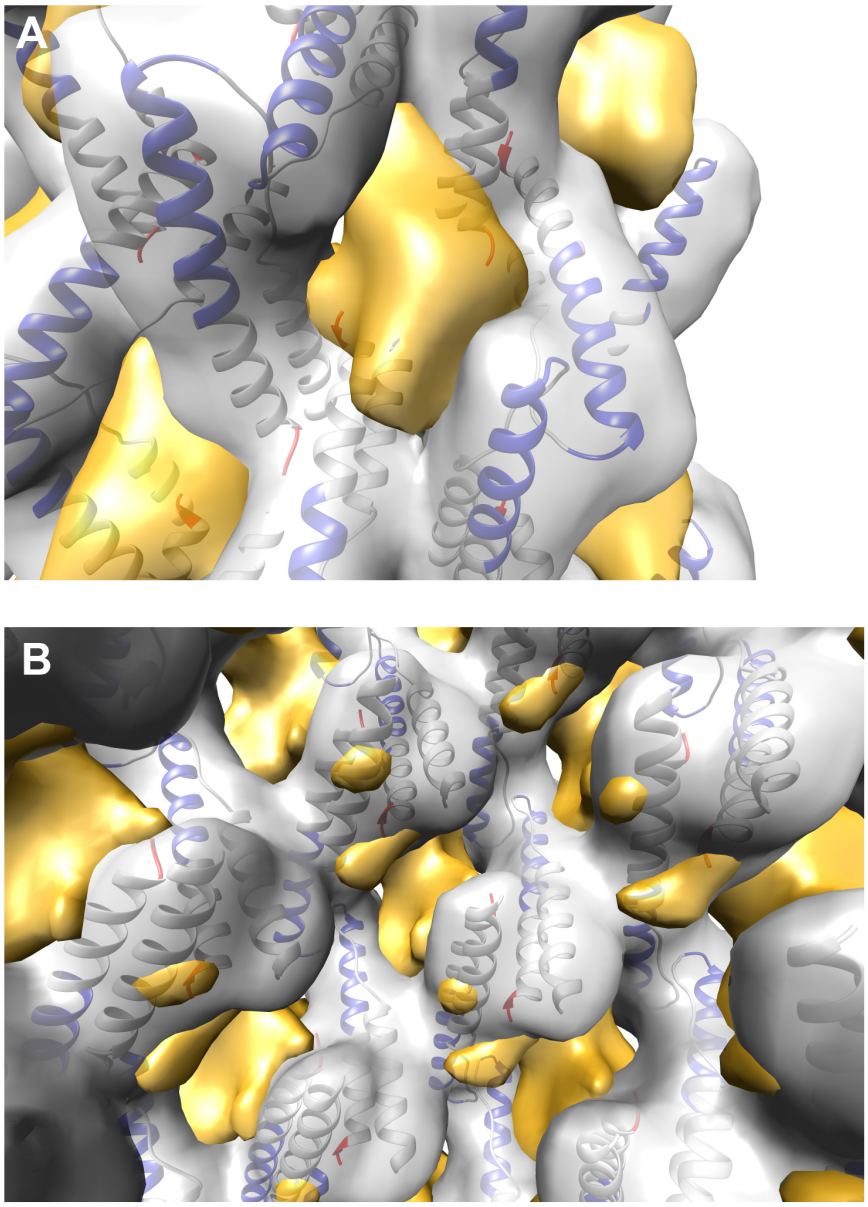


Figure 4.15: Putative density for C-terminal domain. **(A)** External filament surface. **(B)** Internal filament surface. The putative CTD density is on the exterior of the filament and very little density is seen when viewing the interior surface. Notably, the ordered C-termini of adjacent Rev NTD (residue 65) co-localise at the difference map density, consistent with the notion that CTDs of different Rev molecules are juxtaposed in Rev filaments and may jointly form a new domain.

4.5 Conclusions

In this chapter, the structure of *in vitro* Rev filament assemblies have been determined to ~13 Å resolution. Molecular docking of Rev subunits is limited at this resolution, given the dearth of surface features to guide the fitting of the molecules. This molecular docking would be facilitated by a reconstruction at 8-9 Å resolution. At this resolution, α-helices are generally distinguishable as tubes of density. The process is further complicated by the fact that the model for Rev only contains the NTD, leaving approximately half of the polypeptide unaccounted for. However, even with this limitation, random global fitting of a Rev dimer into more featured filaments (Class I-III) resulted in only one orientation with the *A* and *B* surfaces not facing the exterior or interior. This is an important constraint because the *A* and *B* oligomerisation surfaces are very hydrophobic, and their solvent-exposure in an assembled filament is unlikely to be energetically favored. Additionally, these surfaces are thought to form *A*-*A* and *B*-*B* interfaces in Rev oligomers, a reasonable claim given a wealth of biochemical and structural data that support it (Jain and Belasco, 2001; Edgcomb et al., 2008; DiMattia et al., 2010; Daugherty et al., 2010b). Taking advantage of the high degree of symmetry in the Rev filaments (both rotational (C6) and helical (screw)) greatly reduces the possible degrees of freedom for finding the correct orientation of Rev subunits in the filament. The final solution involves Rev NTD molecules whose α-helices are approximately aligned with the filament axis, consistent with the result of electron micro-diffraction studies on aligned Rev filaments (Watts et al., 1998).

Remarkably, the molecular docking solution reveals that, in addition to the well-characterised *A*-*A* and *B*-*B* oligomerisation faces, the *C*-*C* interface is also observed in the Rev filament assembly. The repeated presence of the *C*-*C* interface in different crystal forms (with varying buffer conditions and

protein concentrations) and in Rev filaments altogether suggests that the *C-C* interface is a favorable interaction for Rev multimerisation. Furthermore, the *C-C* interface connects adjacent "strands" of Rev dimers that are oligomerised in an *A-A B-B A-A* fashion, suggesting that it may be an auxiliary interaction that promotes additional Rev subunits to bind using the *A* and *B* surfaces. The participating residues of the *C-C* interface—the poly-proline stretch and Trp45—lend support to this imputed role, given that similar protein-protein interactions are observed in other contexts where weak, extended structures interface so that future protein-protein interactions can form, i.e. in certain cytoskeletal rearrangements and signalling interactions (Kay et al., 2000; Biedermannova et al., 2008).

We have observed significant crossing angle variation among the various crystal structures of the Rev dimer *A-A* interface (see **Chapter 3**). This conformational variability was also seen in the *C-C* interface, and is projected to occur in the *B-B* interface (there are relatively few crystal views of it). We now know that the crossing angle variation of dimer subunits underlies the Rev filament polymorphism that we observe by EM. The observed *A-A* crossing angles in the crystal structures were between 115° and 140°. The filaments comprise Rev dimers with crossing angles in a similar range, between 115° and 165°. The acute-angled Rev dimers produce filaments with narrower diameter and the more obtuse dimers produce wider filaments. The opening of the Rev dimer subunit as the tube grows less featured and wider suggests that the assemblies may be undergoing a stochastic expansion and contraction. As shown in **Fig 4.4**, the diameter of Rev tubes can fluctuate along the length of a given tube, suggesting that the expansion and contraction is highly dynamic and requiring of the concerted movement of many different Rev dimers at a given position along the filament. Speculation of the possible physiological role that this polymorphism may have during HIV replication is discussed in **Chapter 6**.

The CTD of Rev bears the NES needed to initiate engagement of Crm-1 and the other host cell proteins involved in the nuclear export of the Rev-RRE RNP. Therefore, whether it is ordered or not,

it must be surface accessible in whatever architecture the Rev-RRE complex ultimately adopts. Difference maps (Fo-Fc) were calculated by subtracting the pseudo-atomic model of the Rev NTD from the experimental cryo maps. Strong positive difference density (remaining beyond a threshold of 5-6σ) was present in the space adjacent to the *C-C* interfaces. Notably, the putative CTD-density is on the exterior of the filament and connects *A-A B-B A-A* strands, similar to the manner in which the *C-C* interface appears to be a supportive brace. The external location of the CTD is consistent with Fab-decoration of Rev filaments with an Fab whose epitope localises to the CTD (data not shown). Furthermore, the ordered C-termini of two adjacent NTDs point toward the same positive difference density blob, suggesting possible CTD dimerisation. This ordering of the CTD (and possible dimerisation) may be key to stabilising the Rev filament assembly, consistent with the observation that $Rev_{1\text{-}69}$ cannot form filaments. Lastly, each CTD has two cysteine residues which may form either intra- or inter-molecular disulfide bridges within the positive difference density attributed to the CTD. However, it should be noted that Rev filaments readily form in both oxidizing and reducing conditions.

How is it possible that the CTD is ordered in the context of Rev filament formation but not in the crystal structures? Firstly, there is ample evidence that folding of disordered proteins (or protein domains) into ordered structures may occur upon binding to specific partners (e.g. see Dunker et al. (2002); Gunasekaran et al. (2003); Fink (2005)). Furthermore, a recent structural analysis found that disorder content in homodimers, especially in symmetric homodimers, is significantly higher than in heterodimers (Fong et al., 2009). Such intrinsic disorder is often thought to have significant functional importance. In the case of Rev, we know the CTD contains the NES and therefore is essential for nuclear export. The authors of Fong et al. (2009) suggest that "symmetrical arrangements of homodimers may be crucial to keep functional disordered regions juxtaposed in space to form joint binding interfaces." A striking parallel is noted with the interaction network of Rev molecules in assembled

filaments. It is conceivable that the Rev CTDs are only ordered when the NTDs are spatially arranged in a particular way (e.g. through the use of the *A*-*A*, *B*-*B*, and *C*-*C* interfaces in the filament assembly) that promotes CTD ordering and possibly dimerisation, as well.

Chapter 5

The Structure of HBV e-antigen

5.1 Background

HBcAg and HBeAg have been viewed as serologically distinct (Imai et al., 1982; Salfeld et al., 1989; Conway et al., 1998). However, a recent analysis of a panel of six monoclonal antibodies found four to cross-react with both antigens, albeit with markedly differing affinities, and one each specific for HBeAg and HBcAg (Watts et al., 2010). The HBeAg-specific antibody, termed e6, does not bind to HBcAg (as assembled capsids), but can bind to *in vitro* prepared soluble HBcAg dimers. This chapter reports the crystallization and structure determination of the HBeAg-Fab e6 complex to a resolution of 3.3 Å, solved by molecular replacement, exploiting a 2.5 Å structure for the Fab e6 alone, which was determined separately. The Fab fragment facilitated crystallization (**Fig 2.13**). While the resolution is limited, the key features of HBeAg are firmly established (**Fig 2.12**). Here, the structural data are used to explore the long-standing question of how the 10-residue propeptide retained by HBeAg transforms the protein's propensity to assemble, its antigenic character, and its apparent ability to modulate the immune response to favor viral persistence.

5.2 Crystal structure of HBeAg-Fab e6 complex

5.2.1 HBeAg and HBcAg share the same tertiary structure

The structure shows HBeAg as a dimer with each subunit bound to an e6 Fab (**Fig 5.1G**), consistent with the finding that HBeAg is dimeric (Steven et al., 2005) and the reported stoichiometry of the immune complex (Watts et al., 2010). Overall, the tertiary structure is similar to that of HBcAg: helices α1 and α2 and the loop between them encircle an amphipathic hairpin of kinked helices (α3 and α4), followed by α5 and a proline-rich C terminal loop (**Fig 5.1B**) (Wynne et al., 1999; Packianathan et al., 2010). The respective assembly domains superpose with an RMSD of 1.6 Å for the 141 corresponding Cα s, with variability most affecting the spike apex, the loop between α2 and α3a, and the C-terminal loop. However, the greatest difference is the propeptide in HBeAg (**Fig 5.1E, F**), which adopts a loop structure that forms an intra-molecular disulfide between C(-7) and C61 (**Fig 5.1B**). These disulfides in HBeAg (two per dimer)—*versus* the single inter-molecular C61-C61 disulfide that stabilizes the HBcAg dimer—have been reported to be critical for the secretion of HBeAg (Nassal and Rieger, 1993; Schodel et al., 1993; Wasenauer et al., 1992).

5.2.2 HBeAg and HBcAg form radically different dimers

HBcAg dimerizes through the pairing, in parallel, of two helical hairpins to form a four-helix bundle (**Fig 5.2A, right**). The interface between the hairpins is largely hydrophobic and rather flat (Packianathan et al., 2010; Wynne et al., 1999). Flanking the four-helix bundle are salt bridges and hydrogen bonds between the charged, polar residues of the N-terminal strand and the α2-α3 loop. In contrast, in HBeAg, the propeptide loop makes stabilizing hydrophobic contacts with the central part of the α3-α4

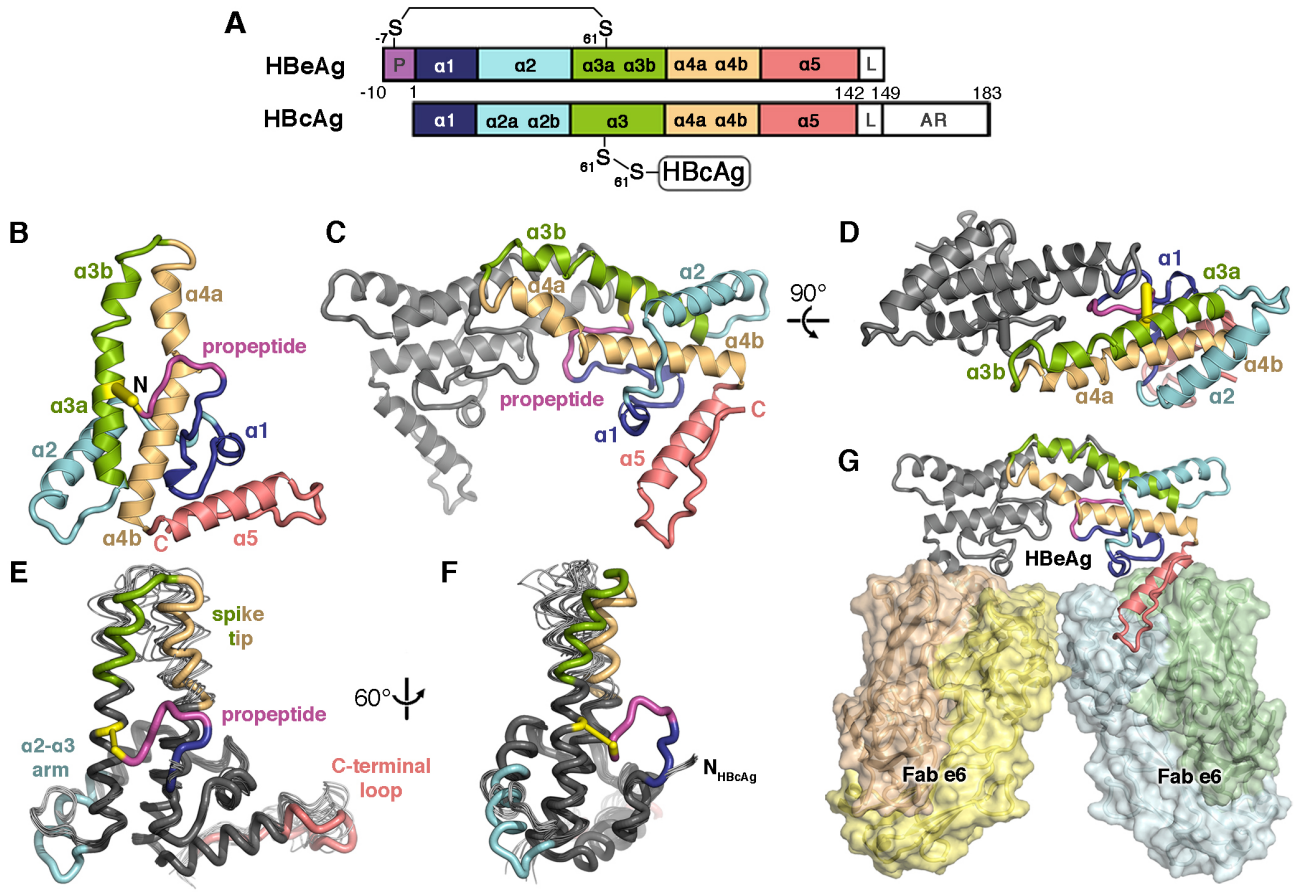


Figure 5.1: Structure of HBV e-antigen. **(A)** Domain schematic of HBeAg and HBcAg, which share the assembly domain but differ in that HBeAg retains the N-terminal propeptide (P; magenta) and lacks the RNA-binding arginine-rich domain (AR). The linker region (L) is present, but disordered in crystal structures of both antigens. **(B)** Ribbon diagram of HBeAg subunit, colored blue to red from N- to C-terminus, with the propeptide (magenta) shown forming an intra-molecular disulfide (yellow) between C(-7) and C61. Secondary structure elements are entirely α-helical (α1-α5); propeptide is irregular coil. C48 and C107, which do not form disulfide bridges (and have been mutated to Ala), are within α2 and α5, respectively, with their side chains oriented toward the interior of the subunit. **(C and D)** Ribbon diagram of the HBeAg dimer. Front subunit is colored according to the scheme in (A); rear subunit is colored gray. Hairpins of the α3b and α4a helices from each subunit form the dimer interface, supported by the propeptides intercalated between them. **(E and F)** Superposition of HBeAg and HBcAg subunits (thick and thin ribbon, respectively). **(G)** Crystal asymmetric unit depicts HBeAg dimer complexed with two Fab e6 molecules (shown as molecular surface) binding at α5 and the C-terminal loop (red).

surface of its own polypeptide chain, where it sterically blocks the formation of an HBcAg-like dimer (**Fig 5.1C**). Instead, HBeAg dimerization involves (part of) the same molecular surfaces but with a relative rotation of ~140° between the hairpins (**Fig 5.2A, left**). In HBeAg, the two α3b helices interact in an anti-parallel manner and make an array of hydrophobic contacts as they pair to form a ridge in the dimer interface. In contrast, the α4a helices do not interact with each other, but instead interact with the exterior surface of the partner's propeptide, also via hydrophobic contacts. In the view along the two-fold axis of the HBeAg dimer (**Fig 5.1D**), it is apparent that the α4a helices are too far apart to interact and that the intercalated propeptide loop completes the dimer interface.

5.2.3 Molecular mimicry of the dimer interfaces

In the HBeAg dimer is a constellation of hydrophobic residues that mimics many of the hydrophobic interactions at the HBcAg dimer interface. For example, propeptide residues W(-4), L(-3), W(-2) in HBeAg make dimer contacts analogous to those of L68, W71, and L76 in HBcAg (**Fig 5.2B**). These three propeptide residues form a hydrophobic cluster accounting for greater than 50% of the surface area buried at the interface. Along the α3b ridge, another hydrophobic patch involves L68, V72, L76, and W71 of both dimer subunits, whereas in HBcAg these residues interact with L84, Y88, M93, and V95 (**Fig 5.2B**). However, the HBeAg dimer interface is much smaller than that of HBcAg, burying 1640 vs. 3920 $Å^2$, has poorer shape complementarity (0.61 vs. 0.70) (Lawrence and Colman, 1993), and makes fewer predicted hydrogen bonds, consistent with biophysical data showing that the HBeAg dimer is stable but less so than the HBcAg dimer (melting temperatures 51 and 65 °C, respectively) (Watts et al., 2011). Despite this difference in stability, the propeptide introduces steric hindrance that blocks HBeAg from adopting the HBcAg conformation. Hydrophobic regions that are buried in

the HBcAg dimer–but would otherwise be exposed in the HBeAg conformation–are shielded by the propeptide (**Fig 5.2C**). This leaves many of the flanking polar residues that participate in dimer-dimer capsid interactions in HBcAg, exposed in HBeAg.

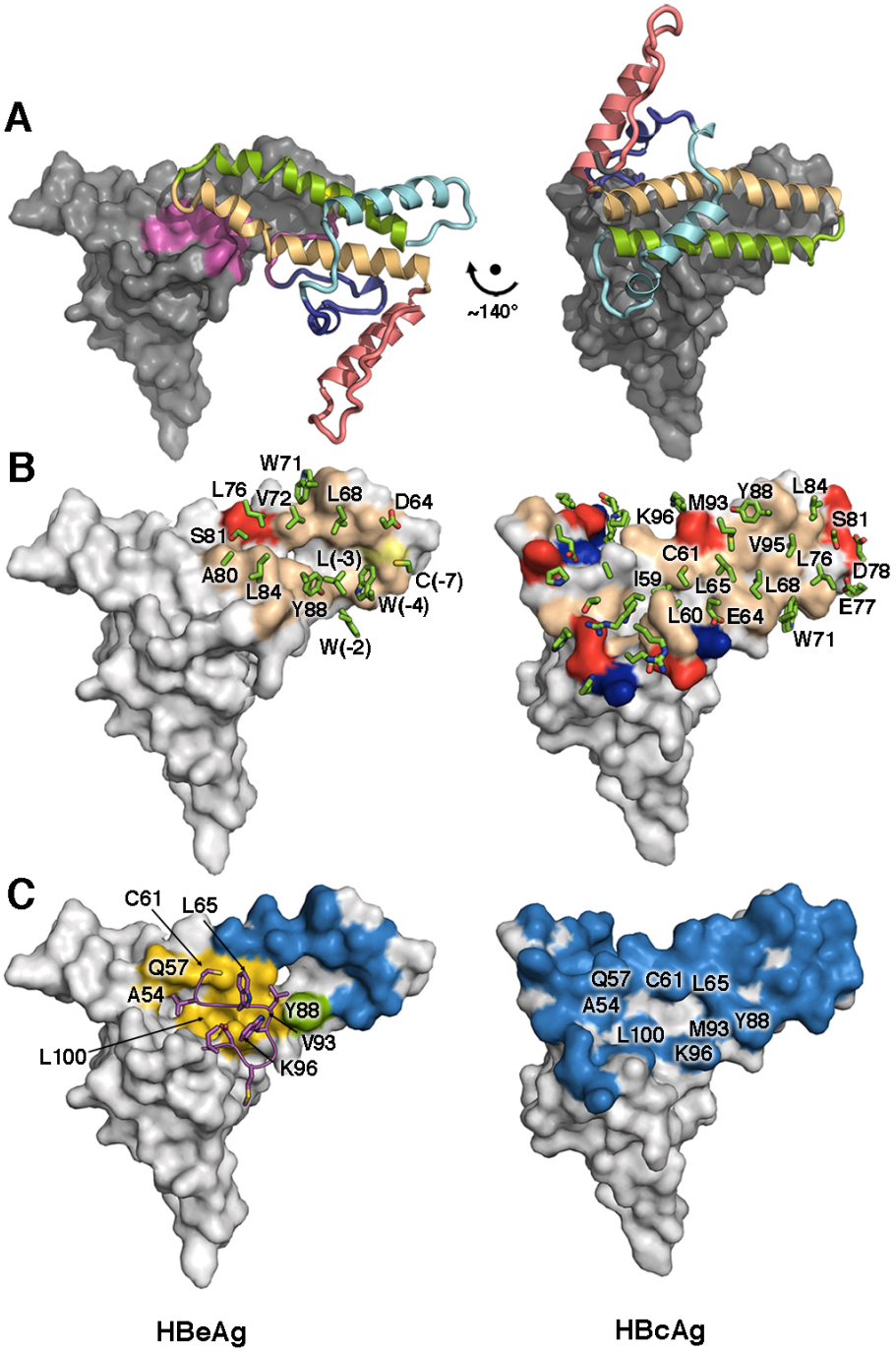


Figure 5.2: Comparison of HBeAg and HBcAg Structures. **(A)** The conformational switch. Propeptide density sterically interferes to block formation of the HBcAg dimer interface, allowing HBeAg to form a different dimer interface using the same surfaces, but ~140° rotated. The gray surface-rendered subunits of each dimer are shown in the same orientation, providing a frame of reference in which the relative rotation between the colored ribbon subunits is evident. **(B)** Dimer interface comparison. HBeAg employs molecular mimicry to form a similar, hydrophobic dimer interface with α3/α4 hairpins, in an inverted orientation relative to HBcAg. Arrays of hydrophobic groups, and even the positioning of aromatics such as Y88 and W71 in HBcAg are replicated by different residues, such as W71 and W(-4) in HBeAg. **(C)** The propeptide mediates hydrophobic contacts within elements that would otherwise already be buried in the context of the HBcAg dimer.

5.3 The Role of the Propeptide

5.3.1 Unlike oxidized HBeAg, the reduced protein can assemble into capsids

Previous work has shown that, under certain in vitro conditions, HBeAg can be induced to form capsids (Watts et al., 2011). On the other hand, the crystal structure of HBeAg reveals a conformation that is incompatible with capsid assembly (**Fig 5.3**). Given the presence of the HBeAg-specific disulfide bridge, we hypothesized that the protein's oxidation state may affect its state of assembly. To test this idea, HBeAg samples were analyzed by analytical ultracentrifugation sedimentation velocity (AUC) using material verified to be dimeric and completely oxidized, as well as the same material but pre-treated with reductant. Oxidized HBeAg remains dimeric (single homogenous sedimenting boundary ~2.5S), whereas reduced HBeAg contains about 30% of a fast-sedimenting component (>40S), consistent with high-molecular weight protein (**Fig 5.4A, B**). Negative stain-electron microscopy (EM) confirmed the presence of capsids in reductant-treated samples together with some other polymeric structures, of which there was no sign in the oxidized sample (**Fig 5.4C, D**). Consistent with these results, it has been shown that an HBeAg mutant in which C(-7) is substituted to A forms dimers with the same intermolecular C61-C61 disulfide bond seen in HBcAg dimers. This change increases the protein's melting temperature to 62 °C (similar to the value observed for HBcAg) (Watts et al., 2011), suggesting that in the absence of the C(-7)-C61 disulfide bridge, the propeptide is displaced, allowing HBeAg to adopt an HBcAg-like conformation. This inference is supported by previous studies that HBeAg could form capsid-like particles (Watts et al., 2011), and we have now demonstrated that this occurs only when the disulfide is disrupted by reduction (**Fig 5.4**).

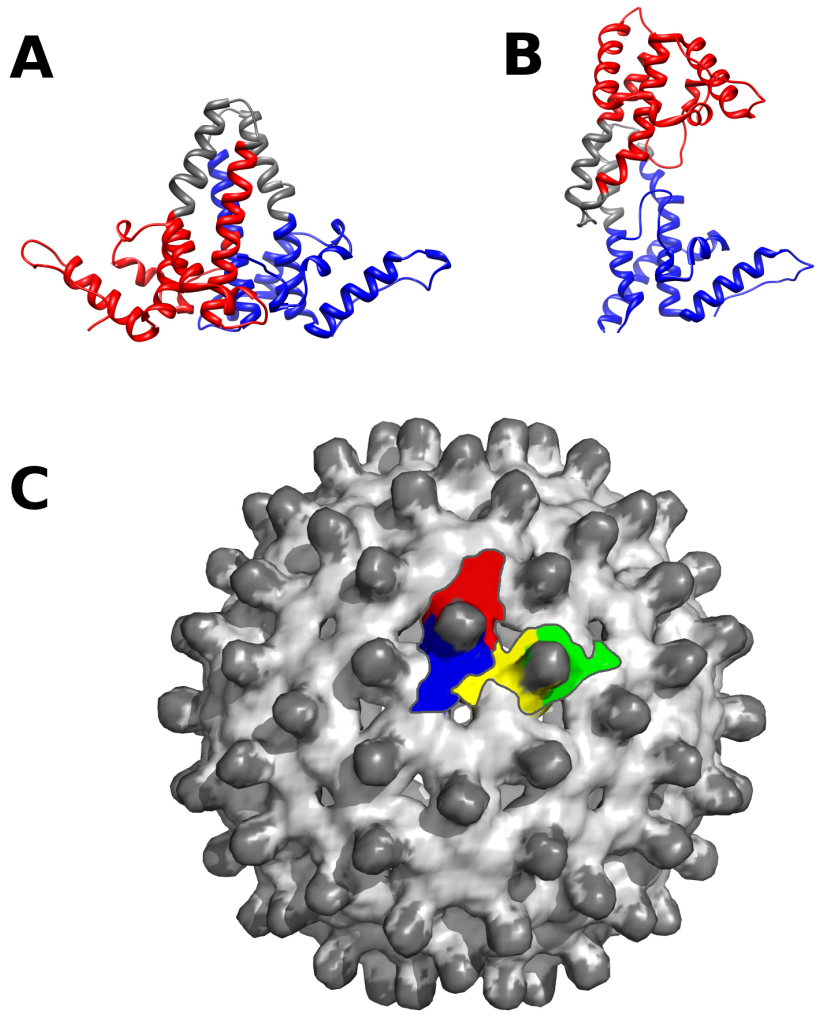


Figure 5.3: Quaternary Structure of HBeAg is incompatible with capsid assembly. **(A)** Ribbon diagram of HBcAg dimer, the fundamental assembly building block of HBV capsids. The stretches of each polypeptide that are involved in forming inter-dimer contacts in assembled capsids are shown in blue and red, respectively. **(B)** Ribbon diagram of HBeAg dimer, in same orientation as the HBcAg is displayed in (A). The radically different quaternary structure of HBeAg prevents the formation of the inter-dimer contacts necessary for capsid assembly. **(C)** Surface rendering of T=4 HBV capsid. Two HBcAg dimers are shown in color to clarify how the dimer building blocks are arranged in the context of the assembled capsids. The red and blue surfaces correspond to the polypeptide stretches of the same colors in (A).

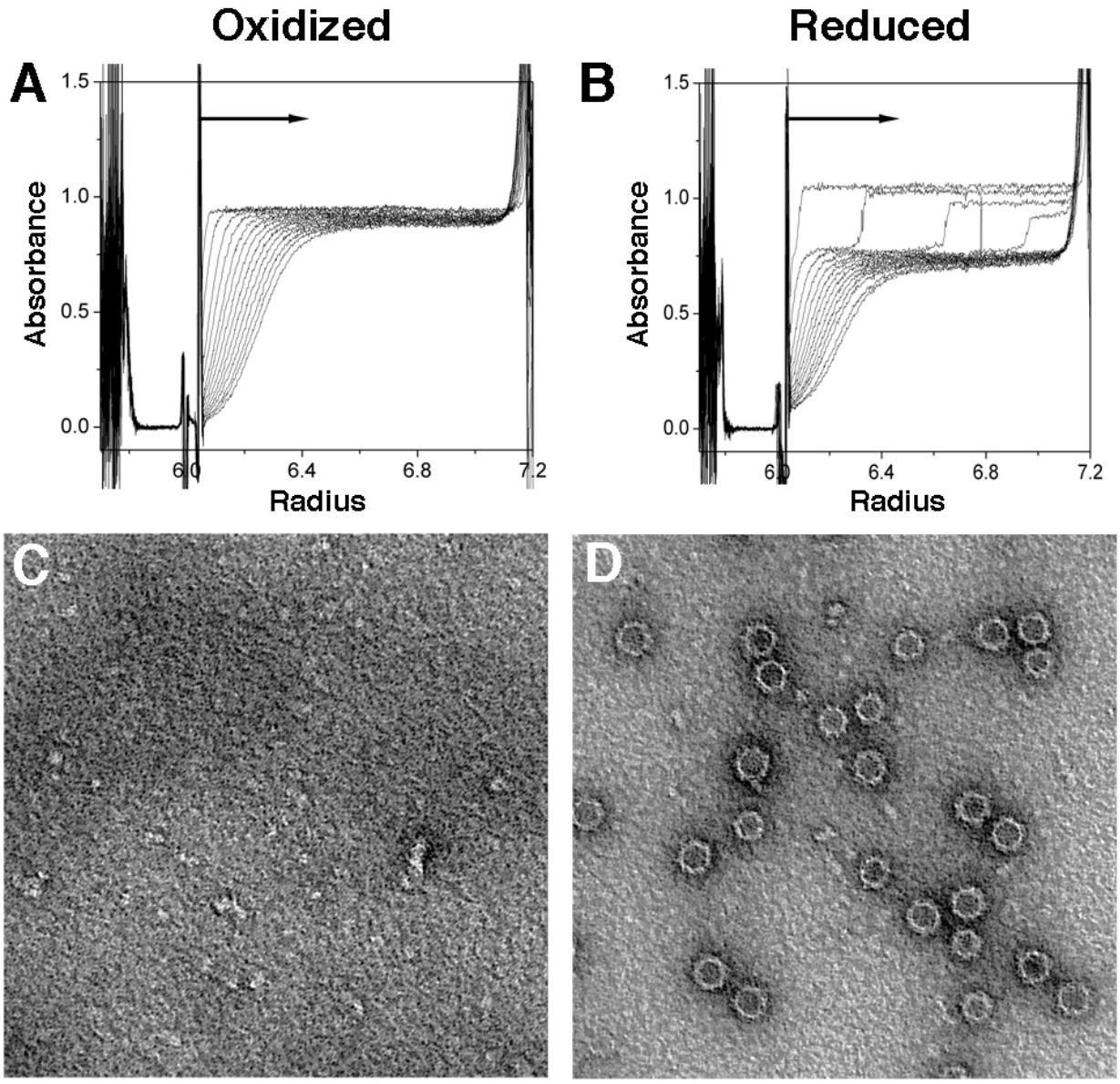


Figure 5.4: Centrifugation and EM experiments on HBeAg in reduced and oxidized forms. (A and B) Sedimentation velocity analyses performed on (A) oxidized and (B) reduced HBeAg using a Beckman Optima XL-1 analytical ultracentrifuge, absorption optics, an An-60 Ti rotor, and standard double-sector centerpiece cells. Measurements at 20 °C were taken at 45,000 rpm for 3 h with data collection at 10 min intervals. The profiles show protein absorbance at 280 nm as a function of radial distance. **(C and D)** Negative-staining EM of assembly products of (C) oxidized and (D) reduced HBeAg. Both samples, in PBS +- DTT, were buffer-exchanged into TBS to avoid precipitation. Both images are at the same magnification; capsids are ~32 nm in diameter.

5.3.2 Steric hindrance of the Propeptide

The HBeAg structure shows how the presence of the propeptide prevents formation of the HBcAg dimer interface by ordering against the inner α3-α4 hairpin surface and sterically blocking dimerization (**Fig 5.2C**). Remarkably, the HbeAg monomer circumvents this steric obstacle by using parts of the same molecular surfaces to form an alternative dimer, but with the subunits rotated ~140° relative to their orientation in HBcAg (**Fig 5.2A**). In this altered conformation, bulky hydrophobic residues W(-4), L(-3), W(-2) in the propeptide shield hydrophobic groups on the opposing subunit's surface that would otherwise be exposed (**Fig 5.2B**). Further, intercalation of the two propeptides between the α3-α4 hairpins of opposing subunits completes the HbeAg dimer interface, mimicking molecular interactions that take place in HBcAg. The importance of the propeptide is corroborated by the near complete conservation of its sequence among the mammalian *Hepadnaviridae* (Revill et al., 2010) (**Fig 5.5**).

5.3.3 Indispensability of the C(-7)-C61 disulfide bridge

The aforementioned data imply that the C(-7)-C61 intra-molecular disulfide bridge present in HBeAg is crucial for maintaining the observed propeptide conformation. Evidence for this inference includes the propeptide's lack of secondary structure and disorder of propeptide residues S(-10), K(-9), L(-8). We tested the influence of this disulfide on HBeAg samples by analytical ultracentrifugation and EM (**Fig 5.4**). When the disulfide bridge is disrupted, the HBeAg subunits revert to an HBcAg-like mode of association, and are able to form capsids. This reversion is consistent with calorimetry data indicating that the HBcAg is more stable than the (oxidized) HBeAg dimer (melting temperature of 65 °C versus 51 °C (Watts et al., 2011). This conclusion as to relative stability is further supported by the HBeAg dimer interface being markedly smaller than that of HBcAg (1640 versus 3970 $Å^2$). It

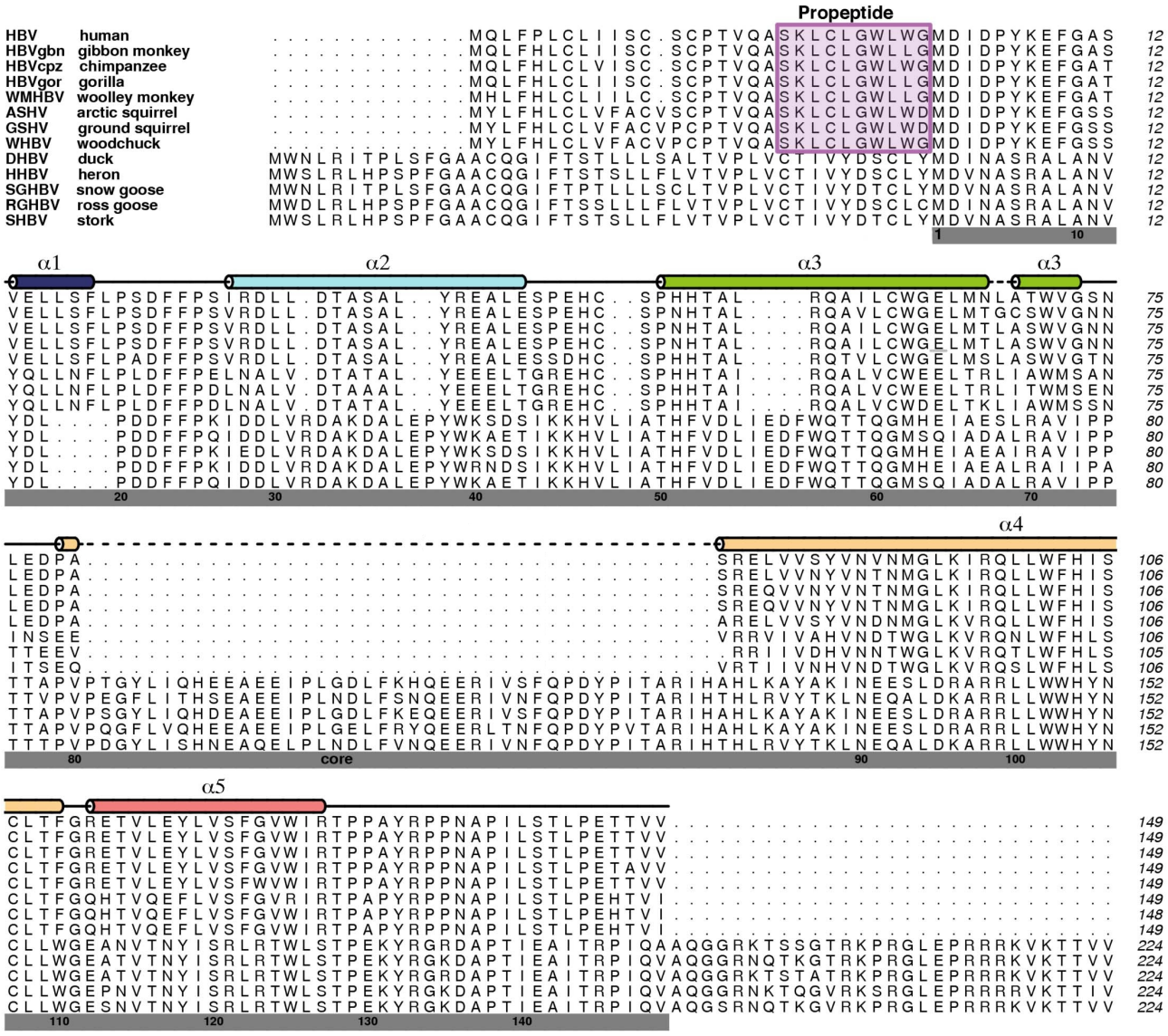


Figure 5.5: Sequence alignment of *Hepadnaviridae* HBeAg sequences. The top eight sequences represent *Orthoviridae* (mammalian) and the remaining five encompass members of *Avihepadnaviridae* (birds). Secondary structure elements are colored as in Fig 5.1 and propeptide sequences are boxed in magenta.

follows that once the disulfide is formed, it locks the propeptide into place. Alternatively, when it is disrupted, the subunits revert to the thermodynamically favored HBcAg conformation. The HBeAg dimer is assembly-incompetent because the subunits are arranged such that the inter-dimer contacts necessary for capsid assembly–between adjacent C-terminal loops and α2-α3 arms–cannot form (Ceres and Zlotnick, 2002).

5.4 Conclusions

5.4.1 Re-interpreting core- and e-Antigenicity

The lack of a high-resolution structure for HBeAg has hampered understanding of its antigenicity. Some antibodies are cross-reactive between HBcAg and HBeAg, while others recognize only one, and the antigens are not considered cross-reactive at the B cell level in natural infection (Baumeister et al., 2000; Imai et al., 1982; Milich and Liang, 2003a; Salfeld et al., 1989; Steven et al., 2005; Watts et al., 2010). The human antibody response to HBcAg is primarily conformational and directed against the outer part of the capsid spike (residues 74-89) as well as the floor of the capsid around the threefold axis (Ferns and Tedder, 1986; Kandiah et al., 2012; Salfeld et al., 1989).

Cryo-EM analyses of Fab-labeled capsids have characterized the epitopes of six murine anti-HBcAg monoclonal antibodies–all conformational–on the capsid surface (**Fig 5.6A**, reviewed in Steven et al. (2005)). These epitopes have been mapped onto the HBeAg structure (**Fig 5.6B**). HBcAg epitopes are typically juxtapositions of two or more loops from different subunits or discontinuous regions of the same subunit. In HBeAg, these loops are moved apart, leaving single loops with reduced affinity. For antibodies 3105 and F11A4, the reduction is by two and three orders of magnitude, respectively (Watts et al., 2010). Furthermore, the epitope of anti-HBcAg antibody 3120 maps to the capsid floor around the three-fold axis, bridging between two adjacent dimers (**Fig 5.6A**). Such floor-binding antibodies do not bind to HBeAg, because formation of their composite epitopes requires capsid assembly (**Fig 5.6B**). In addition to epitopes shared with HBcAg, the HBeAg structure reveals a large, accessible molecular surface that is likely to present antibody determinants unique to HBeAg. This surface exposes regions that are inaccessible in assembled capsids, including the epitope of the e6 antibody used in this study

(binding to α5 and the adjacent C-terminal loop) (**Fig 5.1G**), as well as new surfaces created from the rearrangement of the helical hairpins.

5.4.2 Insight into the Antigens' "split" immune tolerance

HBcAg and HBeAg appear to be regulated independently by the immune system, resulting in a significantly more immunogenic (HBcAg) or tolerogenic (HBeAg) T cell response (Chen et al., 2005; Milich et al., 1997a,b; Milich and McLachlan, 1986; Vanlandschoot et al., 2003). One aspect of this regulation is the observation that HBcAg, but not HBeAg, can directly bind to and activate B cells without the requirement for T cell support, leading to a robust humoral and cytotoxic T cell response (Lazdina et al., 2001; Milich et al., 1997a,b; Milich and McLachlan, 1986). How this takes place is unclear, although it has been proposed that structural differences between the antigens are responsible: the array of spike-resident epitopes on the capsid surface (**Fig 5.6A**) may crosslink B cell receptors, activating the B cells and initiating the cascade (Milich et al., 1997a,b; Vanlandschoot et al., 2003). Because HBeAg has a conformation that precludes capsid formation and the clustering of B cell receptors, this may explain its inability to activate B cells as well as its reduced immunogenicity compared to HBcAg. However, it is known that HBcAg and HBeAg remain cross-reactive at the T cell level, due to sequence identity (as T cell activation involves MHC-presentation of short antigenic peptides) (Milich and Liang, 2003b). This duality may engender an immune response in which T cell cross-reactivity is necessary for inducing clonal tolerance to HBcAg, while the antigenic switching allows HBeAg to avert the robust immune response that HBcAg elicits.

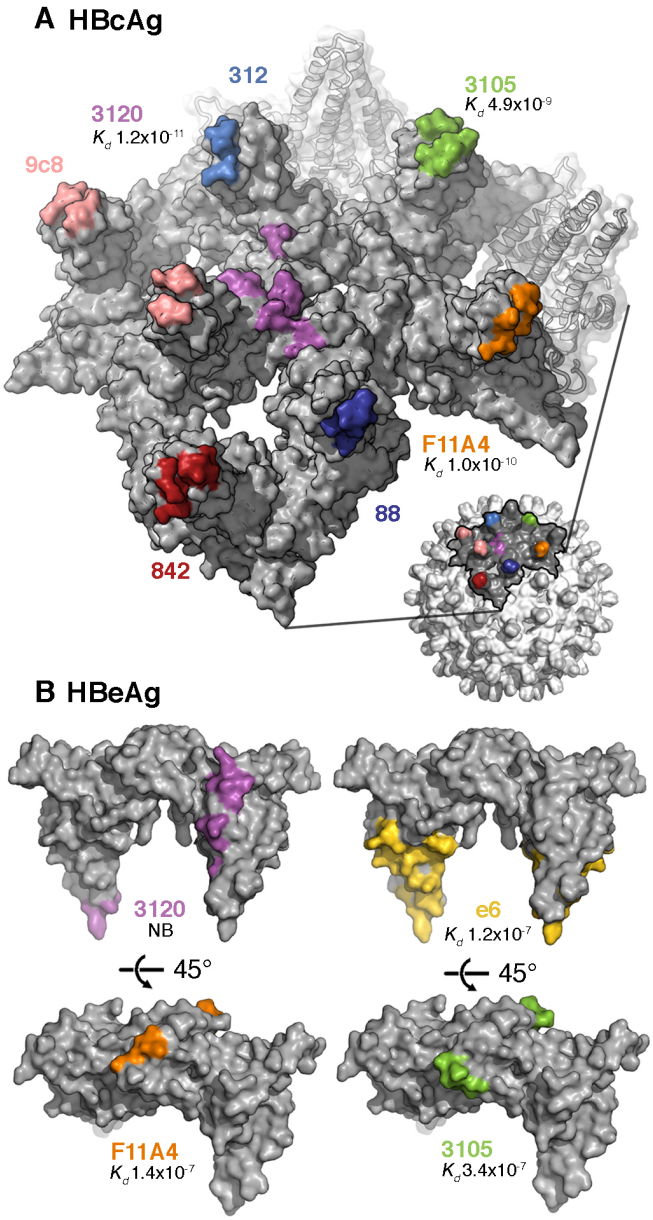


Figure 5.6: Antigenicity of HBeAg and HBcAg. **(A)** Partial surface of an HBcAg capsid with the epitopes for several anti-HBc/eAg antibodies mapped in colors. Most of these epitopes reside around the spike tips, either on one subunit of the dimeric spike or bridging both subunits. Mab 3120 binds to the floor around the threefold and fivefold symmetry axes. **(B)** The same epitopes (matched colors) mapped on the HBeAg dimer. Affinity constants determined by surface plasmon resonance (Kd values; (Watts et al., 2010) are given, where available. A 2-3 order of magnitude drop is observed for Mabs F11A4 and 3105 on HBeAg relative to HBcAg, consistent with the constituent loops, which are close together on HBcAg, being well separated on HBeAg.

Chapter 6

Concluding Remarks

HIV Rev and HBV e-antigen infiltrate entirely different spheres of human cellular metabolism during each of their respective virus' replication cycles. Rev regulates the nuclear export of intron-containing HIV mRNAs and e-antigen modulates the host immune system to favor chronic HBV persistence. However, there are some common features that the two proteins share. Both are key viral proteins which have been under biochemical investigation for several decades in the absence of a structure. Structural data have been lacking for both proteins due to their propensity to aggregate and/or form higher-order assemblies that are refractory to crystallisation.

This dissertation presents the structure determination of HIV Rev and HBV e-antigen by X-ray crystallography. Essential to the structure determination for both proteins was complexing of each with a monoclonal antibody fragment (Fab; antigen binding). Fab molecules are a common tool for forming stabilising and rigid complexes with molecules that are otherwise intransigent to structural study, particularly with soluble and integral membrane proteins (e.g. Newton et al. (2008); Ye et al. (2008); Rasmussen et al. (2011)) and viral proteins (e.g. Colman et al. (1989); Ding et al. (1998); Aoki et al. (2009)). In the present cases, the anti-Rev Fab maintained the protein in a non-polymerised state and the anti-e-antigen Fab facilitated crystal packing.

Cryo-electron microscopy and helical image reconstruction was also performed on *in vitro* assembled Rev filaments. This investigation allowed for synthesis of structural information acquired from the various scFv-Rev and Fab-Rev crystal structures to better understand Rev oligomerisation. Models

for how multimerised Rev may interact with RNA are presented, including a new model based on this hybrid methods approach.

6.1 HIV Rev

It has been nearly three decades since the first reports identifying HIV Rev as an essential component of HIV replication (Feinberg et al., 1986; Sodroski et al., 1986). We now know that Rev mediates the nuclear export of unspliced and partially-spliced viral mRNAs that encode the HIV structural proteins and enzymes. Some of the unspliced transcripts are also packaged into nascent virions where they serve as viral genomes. Rev specifically binds to and multimerises upon these viral mRNAs by recognition of a cognate *cis*-acting element in the HIV genome known as the Rev Response Element (RRE). This event is crucial for HIV replication to proceed; without Rev, these longer transcripts would be retained in the nucleus due to the introns they contain.

In addition to its important role in HIV replication, study of Rev has led to multiple discoveries in the field of nucleocytoplasmic transport, including the first identification of the "leucine rich" NES, as well as that of its binding partner Crm-1 and the export pathway it mediates (see **Fig 1.6** and **Fig 1.8**) (Fornerod et al., 1997; Neville et al., 1997). All of these components are now known to be key players in nuclear export of many host cell macromolecules beyond the context of HIV biology. Nevertheless, much of the structural information underlying Rev function—in particular Rev's ability to self-associate, bind the RRE, and engage other host cell proteins—has remained obscure. This is primarily due to the absence of a crystal structure for Rev.

6.1.1 The structural basis for Rev oligomerisation

Presented in **Chapter 3** is the crystal structure of HIV Rev, first determined to 3.2 Å resolution. Following optimisation of the Fab molecule (to which Rev was complexed) down to a single chain variable fragment (scFv), tigher crystal packing was afforded and the resolution was improved to 2.3 Å. The structure revealed that the N-terminal domain (NTD) of Rev adopts an antiparallel helix-loop-helix structure in which the two helices are in a nearly parallel orientation (**Fig 3.2**). On both sides of this "helical hairpin" are hydrophobic patches that comprise the oligomerisation motifs through which Rev subunits self-associate (**Fig 3.4**). The Fab-Rev crystal asymmetric unit included a dimer of Rev, revealing how the oligomerisation motifs join together in the dimer. For clarity and ease of discussion, the Rev oligomerisation faces have been termed the *A* and *B* faces (or platforms). Essentially, identical hydrophobic patches mate such that dimer subunits adopt a "V-shape" wherein they cross at an angle of ~140°.

6.1.2 Rev-RRE proposed model I

The Rev dimer structure reveals how Rev oligomerisation and RNA-association are spatially segregated, allowing both sets of interactions (one highly hydrophobic and the other highly charged) to proceed without interfering with each other. Dimerisation of Rev subunits situates the RNA-binding ARMs at the inner edge of the "V-shaped" dimer and with both ARMs facing the same side of the dimer (**Fig 3.2**). As additional Rev subunits oligomerise via the *A*-*A* and *B*-*B* interfaces, this same structural arrangement is propagated: ARMs of successive subunits localize to the inner edges of the "V-shaped" oligomer (**Fig 3.7**).

At first glance, such an arrangement appears visually ideal for RNA to run down the central canyon

of the "V-shaped" Rev oligomer, interacting with the ARMs of each successive subunit in an alternating fashion. However, the crossing angle of the dimer subunits in the Fab-Rev dimer structure was 140°, too obtuse for a strand of RNA (either single- or double-stranded) to make simultaneous contact with successive ARMs as described. This was confirmed upon modelling of stem loop IIB RNA (SLIIB, the putative nucleation site for Rev-RNA interactions) onto the Rev dimer structure, taking advantage of an NMR structure of the ARM-SLIIB complex (**Fig 3.5**) (Battiste et al., 1996). Accordingly, a two-track model was proposed whereby RNA runs alongside the outer edge of both ends of the Rev oligomer instead of a single strand through the central canyon (**Fig 3.6**).

This model does offer insight into the cooperative nature of Rev-RNA interactions, but not in a way that excludes the plausibility of other possible models. In theory, any model in which the alignment of RNA along successive ARMs in a manner that is structurally compatible with Rev oligomerisation (i.e. via the formation of successive *A-A B-B A-A* interfaces) is a plausible explanation for cooperative binding of Rev subunits on RNA.

6.1.3 Rev-RRE 'jellyfish' model

A second model for Rev-RRE interaction has been proposed based on the same Rev *A-A B-B A-A* oligomeric arrangement, but with subunit crossing angles of ~120°, as observed in the $Rev_{1\text{-}70}$ crystal structure (Daugherty et al., 2010b). In this model, a hexamer of Rev subunits, oligomerised through the *A-A* and *B-B* interfaces, makes contact with two disparate strands of RNA (ostensibly connected via the various stem-loops and stem-junctions of the RRE). The RNA strands stretch diagonally across the Rev hexamer, interfacing with ARMs along the way. The same NMR structure of the SLIIB-ARM complex used above for Rev-RRE model I was used for modelling of the 'jellyfish' model.

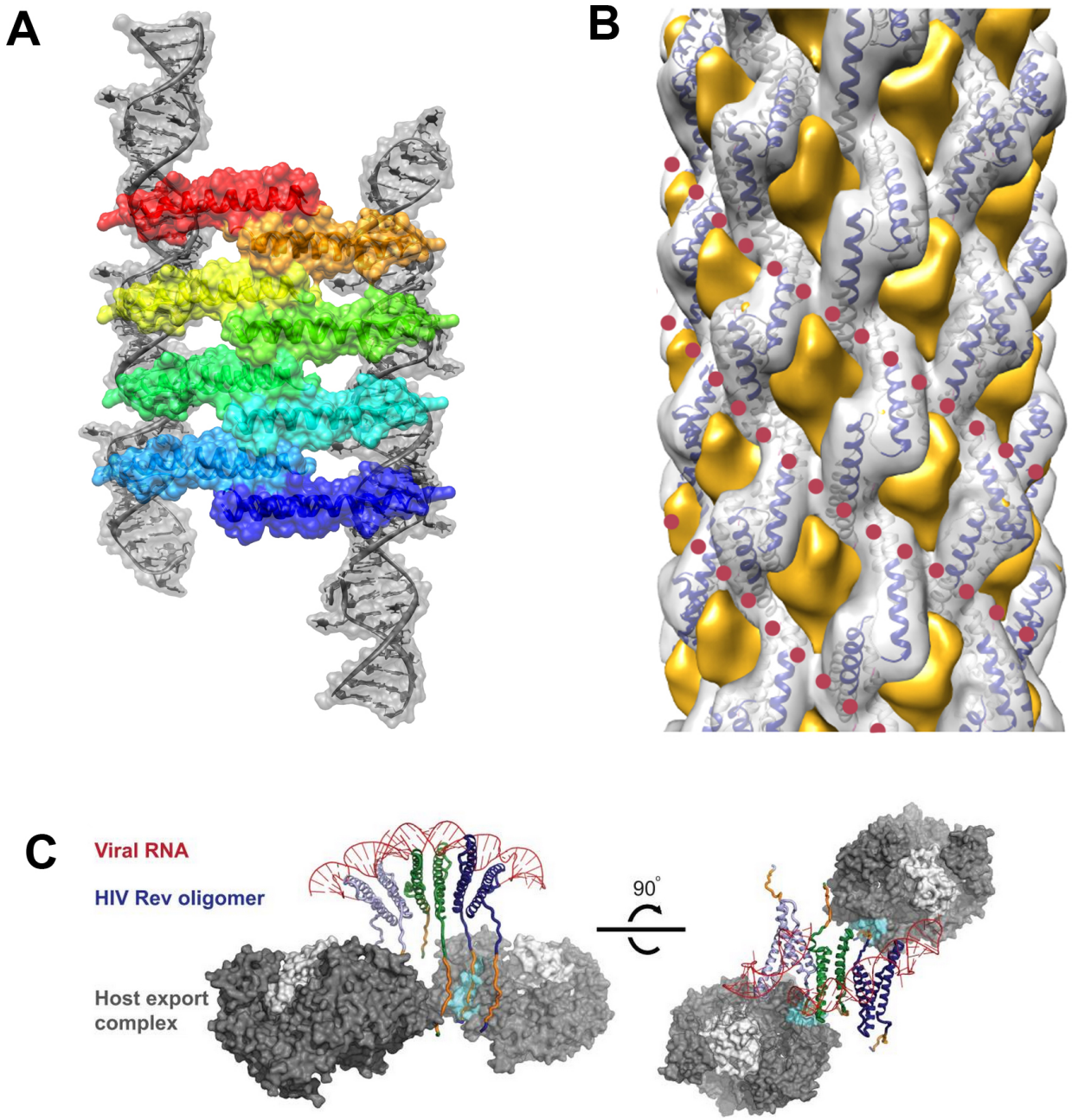


Figure 6.1: Proposed models of Rev-RRE assembly. **(A)** Rev-RRE proposed model I (two-track model). This model was proposed in **Chapter 3** whereby RNA runs alongside the outer edge an oligomer of Rev dimers. The origin of this model stems from modeling of the Rev dimer crystal structure (PDB ID: 2X7L) (DiMattia et al., 2010) onto the NMR structure of SLIIB (Battiste et al., 1996). **(B)** 'Jellyfish' model. This model was proposed by Daugherty et al. and is based on Rev dimer crystal structure that has two point mutations at the *A-A* interface (Daugherty et al., 2010b). In this model, RNA strands stretch diagonally across a Rev hexamer, interfacing with ARMs on one side. The disorderd CTDs are presumed to be accessible on the oppsing side of the RNP, where they can interact with host cell export protein, Crm-1. **(C)** Rev-RRE proposed model II (filament model). This model takes into account the cryo-EM reconstructions of Rev filaments, showing how the alignment of ARMs on the exterior of the filament along helical valleys is suggestive of RNA-binding channels. Alteration of the Rev dimer dihedral angle creates filaments in which these valleys are variable in depth. This may permit the ARMs of the Rev subunits within the assembly to interact with either RNA during export (acute dihedral; narrow channel) or importin-β during import (obtuse dihedral; wide channel).

Evidence supporting this model include the characterization of discrete, asymmetric Rev-RRE complexes containing six Rev subunits and a SAXS envelope corresponding to a Rev dimer and SLIIB RNA (Daugherty et al., 2010a). The CTD was removed from the construct used in crystallization of $Rev_{1\text{-}70}$; however, the disordered domains are projected to extend away from the base of the hexamer, on the opposite side of the Rev subunits as the ARMs (giving the impression of a jellyfish).

6.1.4 The 'hybrid' approach

A combined analysis of the cryo-EM reconstructions of *in vitro* assembled Rev filaments and the scFv-Rev crystal structures affords a clearer picture of Rev oligomerisation than what was previously known:

- In addition to the *A*-*A* and *B*-*B* interfaces, there is a third Rev-Rev interaction, here termed the *C*-*C* interface.
- Organisation of Rev subunits using all three oligomerisation interfaces (*A*, *B*, and *C*) engenders ordering of the CTD (or at least a subset of it) into what appears to be a new domain.
- Variation in Rev dimer crossing angles was observed among the different scFv-Rev crystal structures, ranging from 115°-140°, providing structural evidence that Rev oligomerisation is a highly dynamic process.
- The variation in crossing angles correlates with Rev filament polymorphism. More acute dimers give thinner filaments with larger canyons that wrap around the exterior; more obtuse dimers give wider filaments without canyons, instead having a smooth exterior.

Previously, the Rev filaments observed by EM have been assumed to be an artifact of *in vitro* preparation. However, it is plausible that the filament structure reveals an artifically long structure that

is representative of actual Rev-Rev interactions that occur during viral replication. Taken together, these results convey Rev oligomerisation to be a highly dynamic system of protein-protein interactions, two of which are predominantly flat and hydrophobic, (*A* and *B*) and one that involves a specific interaction between poly-proline loops and a tryptophan. Conformational variability in all of the interfaces has been observed, revealing that Rev can alter its supramolecular structure by concerted adjustment of the crossing angles between subunits. Notably, only in the context of Rev filaments (thus far), has any ordering of the Rev CTD been observed.

Biochemical characterisation of the residues participating at the *A* and *B* interfaces has confirmed their formation during Rev multimerisation onto the RRE. These interactions are indispensable for Rev-mediated nuclear export. Therefore, it is probable that they form in a similar fashion in the context of RNA-association. Furthermore, Havlin et al. (2007) 2D solid-state ^{13}C-^{13}C NMR spectroscopy on Rev and Rev-RNA filaments, in which all Ile, Val, and Ala residues were uniformly labeled with ^{13}C, indicated that the secondary structure of Rev was essentially the same in both assemblies. It is notable that their results also indicated the presence of additional helix content (beyond what was attributed to the NTD) at Ile and Val residues in the Rev CTD.

6.1.5 Rev filament polymorphism may enable binding to different interaction partners

6.1.5.1 Rev-RRE proposed model II

Conformational variability within all of the Rev-Rev protein interfaces allows for Rev helical filament assemblies to exist with drastically different surface features. When the crossing angles between Rev subunits are acute (115°), thin tubes (Class I - 11.5 nm in diameter) are observed with prominent

grooves that run along the exterior surface. The positively-charged ARMs coat these grooves in the same manner that was observed along Rev *A-A B-B A-A* oligomers in the previous two Rev-RRE models. These ARM-coated grooves are evocative of RNA-binding channels that might coat the exterior of the tube, which is consistent with the suggestion of the Rev-RNA filament surface being negatively-charged (Havlin et al., 2007). However, two problems immediately emerge with this model:

1. Rev-RNA filaments observed to date are ~8 nm in diameter, not in the range of 11.5-14.5 nm for the Rev filament structures (see **Fig 1.10** and **Fig 4.10**).
2. A crossing angle of 115° between Rev subunits is too large for successive ARMs to contact RNA running down the central "V-shaped" track of the Rev molecules. Jain and Belasco, in their modeling of the Rev-RRE interaction more than a decade ago, suggested that a Rev NTD crossing angle of ~85° would be necessary.

However, before dispensing with this model, it must be noted that Rev subunit crossing angle and filament diameter are intrinsically linked (see **Chapter 4**). If RRE-containing RNA were to wrap around the outside of Rev subunits, themselves bonded through the *A*, *B*, and *C* surfaces, it is plausible the RNA could induce a thermodynamic state that is only favorable in the presence of RNA, i.e with a smaller crossing angle. Accordingly, it is plausible that the diameter of such a Rev-RNA filament could be 2-3 nm smaller in diameter than Class I filaments given a 30° smaller crossing angle. Approximately the same degree of variation is seen, both in terms of filament diameter and subunit crossing angle, between Class I and Class V filaments.

It is also plausible that the presence of RNA induces a Rev assembly that deviates from the six-fold rotational symmetry. Nucleation of a Rev-RRE assembly that is guided by the presence of RNA may instead form a one- or two-start helix that still uses the same *A*, *B*, and *C* interfaces. Such a helical

filament would probably be more flexible, consistent with what is observed by negative-stain EM.

This Rev-RRE model, in which RNA wraps around the exterior of a thin Rev filament is compelling for multiple reasons. First, it explains how RNA could guide Rev into a qualitatively different structure than Rev filaments while still using the same underlying Rev-Rev *A*, *B*, and *C* oligomerisation faces. Second, it exposes the CTD to the exterior of the filament, unobstructued by RNA association, which allows for interaction of Crm-1, Ran, and other host cell factors needed to mediate translocation through the NPC. The natural curvature of karyopherin solenoids is highly flexible (Forwood et al., 2010), such that Crm-1 could engage the exterior of the Rev-RNA assembly. The filament framework, in addition to providing the right environment for ordering of the CTD and exposure of the NES, would provide a structural platform for engagement of the plethora of host cell proteins that are involved in Rev-mediated nuclear export of RNA.

Implicit in this model is the assumption that Rev subunits continue to associate with RRE-containing RNA outside of the RRE region itself. Consistent with this idea is the observation by negative-stain EM of Rev-RNA filaments that are proportional to the size of RNA. This observation even holds true for long RNAs, such as a 2.4-kb stretch of *env* RNA that contains the RRE (**Fig 1.10**). Indeed, the phenomenon whereby viral RNA influences the oligomeric state of viral RNA-binding proteins is one that has precedent, as noted in the reports of crystal structures of rabies virus and respiratory syncytial virus nucleoprotein-RNA complexes (Albertini et al., 2006; Tawar et al., 2009). The Rev-RRE assemblies that have been observed *in vitro* are all approximately 8-10 nm in diameter.

The supramolecular filament assembly provides a binding platform for the myriad transport factors and other host cell proteins that are necessary for Rev-mediated nuclear export. As translocation of macromolecular assemblies of this diameter is feasible without the need for large-scale quaternary remodelling at the NPC interface, the filament structure would itself facilitate cargo transport. Multiple

Crm-1 molecules can accumulate along the length of the RNP, synergistically facilitating transport as the length of Rev-RRE filament (with bound host cell proteins) traverses the NPC. (**Fig 1.5D**).

6.1.5.2 Proposed Rev-importin-β model

When finished with a round of export, Rev must be able to bind importin-β to be re-imported into the nucleus, presumably to initiate another round of RNA export. There is evidence that Rev undergoes nuclear import in a multimerised state (Kubota et al., 1992; Szilvay et al., 1997). In the Class V filaments, the Rev subunit crossing angles are so obtuse (165°) that the ARM grooves are non-existent, forming instead into essentially flat ARM sheets. It is known that the Rev NLS (co-incident with the ARM) directly engages importin-β in order to effect nuclear import. Intriguingly, importin-β is too large to fit into the narrow grooves of Class I filaments (**Fig 1.9**). Therefore, it is possible that the Class V filaments, where the NLS motifs align in extended flat arrays on the filament surface, may be used for nuclear import. Even with a diameter of ~15 nm plus additional width for bound importin-β molecules, such a hypothetical structure is not too large for translocation through the NPC (see **Section 1.1.2.4**). Engagement of importin-β on the exterior of a Class V-like filament would allow for the concave side of the importin-β to match the curvature of the filament, while permitting the convex side to interact with the FG Nups of the NPC during nuclear import (**Fig 1.9**).

The polymorphic nature of Rev filaments may therefore be to accommodate different functional states. *In vitro*, we see Rev filaments (without RNA, importin-β, or any other host cell factors) existing in a dynamic range of conformational states. These proposed models are testable and would predict that RNA-Rev filaments would exist as a nominally "Class 0" filament, thinner in diameter and with fewer helical strands, while maintaining the same inter-molecular interactions between Rev subunits. The Rev filament-importin model may also be tested by complexing the two components and seeing if

decorated filaments are formed with a preferentially large diameter (i.e. similar to Class V).

6.2 HBV e-antigen

It has been four decades since the first description of HBeAg (Magnius and Espmark, 1972), and three since its isolation from serum and the discovery that it is closely related to HBcAg (Ferns and Tedder, 1984; MacKay et al., 1981). Despite this close relation, the two antigens have proved to have profoundly different biophysical and functional properties (Chen et al., 2004, 2005; Milich and Liang, 2003b; Nassal and Rieger, 1993; Steven et al., 2005; Wasenauer et al., 1992). The crystal structure of HBeAg reveals that its subunit has much the same fold as that of HBcAg (Wynne et al., 1999) but a radically different mode of dimerization, which explains the biophysical and antigenic differences between the two proteins. While there are many examples of alternative dimer interfaces in protein crystal packing, the phenomenon of alternative physiologically relevant modes of dimerization has few precedents (Silvan et al., 2012).

The molecular switch that relates the pairing of assembly domains in HBeAg and HBcAg explains many of the fundamental biophysical and antigenic differences between them (**Fig 6.2**) and has implications for their respective immunological properties in the context of chronic HBV infection. There are probably other levels of regulation in which HBeAg also takes part to establish chronicity, including, for example, modulation of the innate and adaptive immune responses via direct interaction with host immune proteins (Lang et al., 2011; Purvina et al., 2012; Visvanathan et al., 2007; Yang et al., 2006). A long-standing perception of HBeAg has been that it is a monomeric protein with an N-terminal propeptide that happens to be requisite for HBeAg functionality. Our data have shown that HBeAg is not a monomer, but a dimer, and that the crucial feature of its propeptide is its strategically positioned Cys

residue that dictates the structure of the HBeAg dimer. This structure may, we conjecture, confer an ability to interact directly with host immune proteins in ways that HBcAg cannot. The crystal structure now provides a framework upon which further study can fully elucidate the role of HBeAg in HBV persistence.

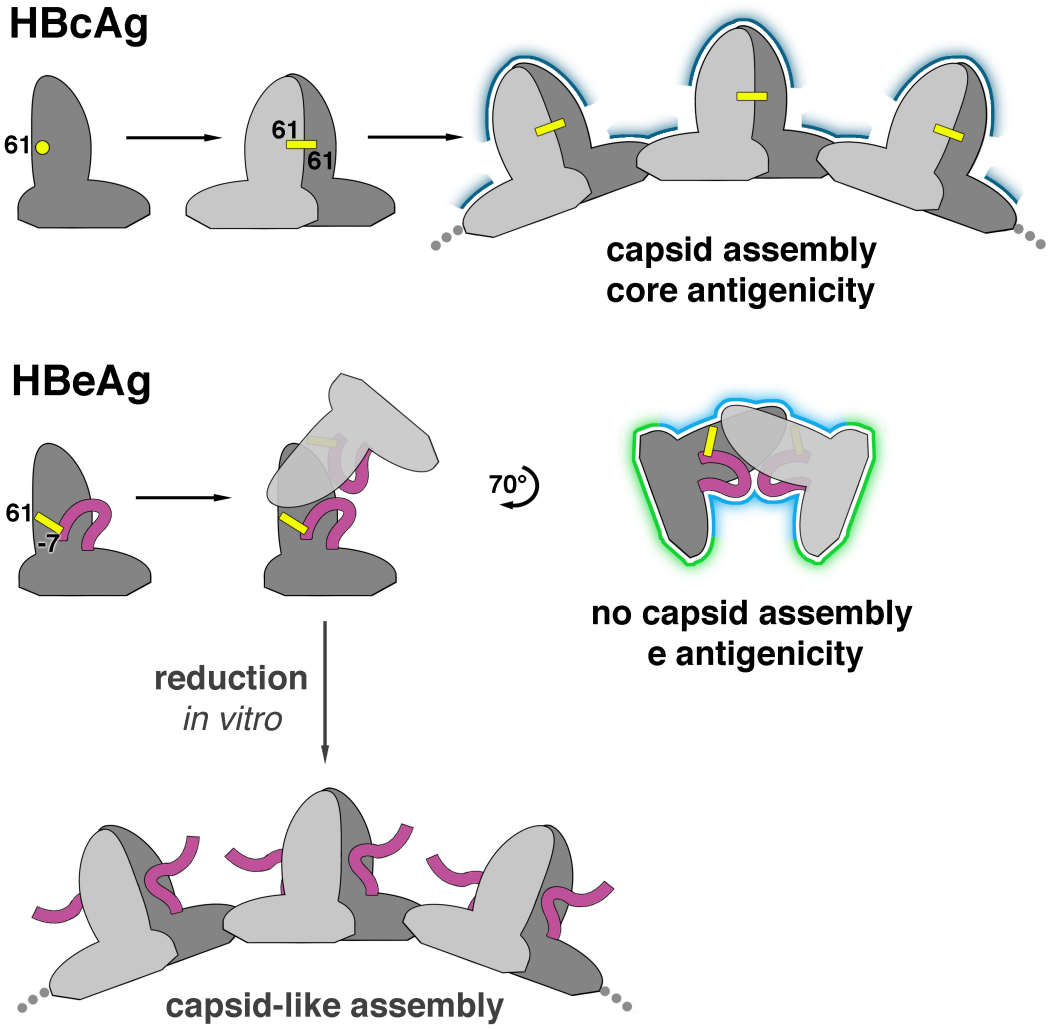


Figure 6.2: Antigenic switching of HBV capsid protein. HBcAg and HBeAg polypeptides share the same tertiary structure. When HBcAg dimerizes, an intermolecular disulfide bridge forms between C61 of each subunit. In the HBeAg dimer, however, two intra-molecular disulfides form between C61 and propeptide residue C(-7). Once locked into place, the propeptides block HBcAg dimer formation. Instead, the HBeAg subunits dimerize in an entirely different quaternary arrangement. As a result, HBeAg dimers cannot form the dimer-dimer contacts employed in capsid assembly. The molecular switch also explains how HBeAg and HBcAg are antigenically distinct: surfaces presenting conformational epitopes on HBcAg (dark blue) are altered in HBeAg (light blue) and antibody inaccessible surfaces on the interior of capsids are exposed in HBeAg (green). Antibodies that bind to only one subunit within the shared surfaces (light and dark blue) may be cross-reactive for both antigens. If HBeAg is subjected to reducing conditions *in vitro*, these disulfides are disrupted and capsid-like assemblies can form.

6.3 Future Work

6.3.1 Cryo-EM of Rev-RRE complexes

Structural determination of Rev-RRE complexes are a clear next step for understanding Rev oligomerisation onto the RRE, as well as how the architecture of the RNP underlies interaction with Crm-1 and other host cell factors. While the Rev-RRE assemblies observed in negative stain EM are far less regular than Rev filaments (**Fig 1.10**), the single-particle IHRSR method (as done in this dissertation for Rev-only filaments) may be a plausible route to structural determination of the Rev-RRE assembly. Very recently, a similar approach was implemented to determine the structure of influenza RNPs, which incidentally are also helical and heterogeneous in structure (Arranz et al., 2012; Moeller et al., 2012).

6.3.2 Structure determination of Rev- and HBeAg-host cell protein complexes

Rev and HBeAg, like many viral proteins, interact with many host cell proteins during their respective viral replication cycles. Many of these interactions are probably transient and would be difficult to capture in an *in vitro* setting. However others may engage Rev or HBeAg with high enough affinity for structural characterisation to be possible.

For Rev, an interaction with human nucleosome assembly protein 1 (Nap1) has been characterised that is stable, stoichiometry and functionally-relevant (Cochrane et al., 2009). While the function is not certain, the complex may form in the nucleus to prevent Rev self-association prior to engagement of Rev with RRE-containing RNAs. Crystallisation trials of the Rev-Nap1 complex are underway. Rev has also been shown to interact with HIV Integrase and it is thought that this interaction downregulates further HIV genome integration events (which could lead to apoptosis or other programmed cell death

outcomes) (Levin et al., 2010). Structure determination of the Rev-IN complex, should it be possible to isolate, would be most illuminating for understanding this aspect of the HIV replication cycle (Benyamini et al., 2011).

Additionally, HBeAg engages various host cell proteins during HBV replication that are involved in modulation of the immune response and inflammatory signalling pathways (e.g. IL-1) (Lang et al., 2011; Purvina et al., 2012; Visvanathan et al., 2007; Yang et al., 2006). Notably, the HBeAg structure that is radically different from that of HBcAg, and therefore may be able to interact with the aforementioned proteins in a way that HBcAg cannot. Structure determination of any such HBeAg-host cell complex would advance our understanding of how HBeAg functions in viral persistence.

6.3.3 Development of anti-Rev/HBeAg scFvs as therapeutic intrabodies

The anti-Rev and anti-HBeAg scFvs were primarily developed for use as crystallization chaperones. However, the fact that anti-Rev scFv binds to one of Rev's oligomerisation faces (*B*), effectively depolymerising Rev *in vitro*, signifies that translation of this interaction to an *in vivo* context may similarly inhibit Rev oligomerisation onto the RRE. As HIV-1 replication has been shown to be dependent on the ability of Rev to assemble as a multi-protein complex on the RRE, there is value in exploring the possible anti-HIV-1 therapeutic potential of the scFv. The high-affinity between scFv and Rev further improves the prospect of utilizing scFv (or a derivative thereof) as a novel Rev inhibitor.

Similarly, the anti-HBeAg Fab specifically binds to HBeAg and not HBcAg and the crystal structure presented in **Chapter 5** reveals the structure of the epitope-paratope interface. The development of novel HBeAg immunoglobulin-based therapeutics may be useful for artificially inducing an anti-HBe response in infected patients. In fact, immunoglobulin therapy is already a clinically accepted ap-

proach in HBV treatment, with HBsAg targeted as such in the prophylactic vaccine. Such anti-HBeAg treatment may lead to the "mopping up" of serum HBeAg, or alternatively, functional interference of intracellular HBeAg. Reduction of HBeAg load could effectively interrupt the tolerogenic effect on the innate immune response, and engender a more robust host immune response to HBV infection. These outcomes would enhance the success rate of current interferon-α treatment (Walsh and Locarnini, 2012) and may also reduce the likelihood of developing chronic HBV infection.

One way to translate the scFv into a clinical context would be via anti-Rev or anti-HBeAg intracellular immunization with scFv targeted to the cytoplasm. However, intracellular scFv therapy has two major obstacles: (1) scFv instability in the cellular milieu; and (2) finding an efficient method of delivery. The lack of disulfide bond formation in the reducing environment of the cytosol severely reduces the stability, solubility, and degree of correct folding of clinical intrabodies. Additionally, the process of delivering the intrabody across the plasma membrane to the appropriate cell type is non-trivial. Nonetheless, scFv scaffolds have been developed that maintain stability in the cytosol by not requiring disulfide bond formation for proper folding. The scFv presented herein could have increased intracellular stability by introducing the necessary mutations based on previously developed frameworks. Another option would be to perform CDR grafting of the scFv onto a stable intracellular scFv.

6.3.4 Design of peptidic or small-molecule inhibitors of Rev oligomerisation

Three Rev oligomerisation interfaces, termed *A*, *B*, *C* have been structurally characterised. No HIV anti-viral drugs have been designed that target Rev. Rev-mediated nuclear export of viral mRNAs is a primary event in HIV replication, without which there would be no generation of nascent viral progeny. Therefore, it constitutes an important target for therapeutic intervention. The 2.3 Å Rev dimer crystal

structure reported in **Chapter 3** is the highest resolution determined to date for HIV Rev. It may aid in structure-based drug design of Rev oligomerisation inhibitors.

Structural analysis of the *A-A* interface revealed a key hydrophobic residue, Leu64, that mediates formation of the interface. This residue has not been biochemically characterised in terms of its role in Rev multimerisation, but it is likely to be essential. It may therefore serve as a "hot spot" for structure-based drug design (Kuttner and Engel, 2012). Additionally, the improved resolution of the scFv-Rev structure (2.3 Å affords a more detailed analysis of the Fab-Rev paratope-epitope interface. Structural analysis of this interface may lead to design of peptide mimics of the interaction, which may retain the ability to bind Rev and inhibit Rev oligomerisation.

References

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart (2010), "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallogr. D Biol. Crystallogr.*, 66, 213–221.

Aksyuk, A. A., L. P. Kurochkina, A. Fokine, F. Forouhar, V. V. Mesyanzhinov, L. Tong, and M. G. Rossmann (2011), "Structural conservation of the myoviridae phage tail sheath protein fold." *Structure*, 19, 1885–1894.

Albertini, A. A., A. K. Wernimont, T. Muziol, R. B. Ravelli, C. R. Clapier, G. Schoehn, W. Weissenhorn, and R. W. Ruigrok (2006), "Crystal structure of the rabies virus nucleoprotein-RNA complex." *Science*, 313, 360–363.

Andreeva, A. and A. G. Murzin (2006), "Evolution of protein fold in the presence of functional constraints." *Curr. Opin. Struct. Biol.*, 16, 399–408.

Aoki, S. T., E. C. Settembre, S. D. Trask, H. B. Greenberg, S. C. Harrison, and P. R. Dormitzer (2009), "Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab." *Science*, 324, 1444–1447.

Arnold, M., A. Nath, J. Hauber, and R. H. Kehlenbach (2006), "Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1." *J. Biol. Chem.*, 281, 20883–20890.

Arranz, R., R. Coloma, F. J. Chichon, J. J. Conesa, J. L. Carrascosa, J. M. Valpuesta, J. Ortin, and J. Martin-Benito (2012), "The structure of native influenza virion ribonucleoproteins." *Science*, 338, 1634–1637.

Auer, M., H. U. Gremlich, J. M. Seifert, T. J. Daly, T. G. Parslow, G. Casari, and H. Gstach (1994), "Helix-loop-helix motif in HIV-1 Rev." *Biochemistry*, 33, 2988–2996.

Battiste, J. L., H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, A. D. Frankel, and J. R. Williamson (1996), "Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex." *Science*, 273, 1547–1551.

Baumeister, M. A., A. Medina-Selby, D. Coit, S. Nguyen, C. George-Nascimento, A. Gyenes, P. Valenzuela, G. Kuo, and D. Y. Chien (2000), "Hepatitis B virus e antigen specific epitopes and limitations of commercial anti-HBe immunoassays." *J. Med. Virol.*, 60, 256–263.

Behrmann, E., G. Tao, D. L. Stokes, E. H. Egelman, S. Raunser, and P. A. Penczek (2012), "Real-space processing of helical filaments in SPARX." *J. Struct. Biol.*, 177, 302–313.

Bell, N. M. and A. M. Lever (2013), "HIV Gag polyprotein: processing and early viral particle assembly." *Trends Microbiol.*, 21, 136–144.

Benyamini, H., A. Loyter, and A. Friedler (2011), "A structural model of the HIV-1 Rev-integrase complex: the molecular basis of integrase regulation by Rev." *Biochem. Biophys. Res. Commun.*, 416, 252–257.

Biedermannova, L., K. E Riley, K. Berka, P. Hobza, and J. Vondrasek (2008), "Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane." *Phys Chem Chem Phys*, 10, 6350–6359.

Blanc, E., P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea, and G. Bricogne (2004), "Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT." *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2210–2221.

Blanco, F. J., S. Hess, L. K. Pannell, N. W. Rizzo, and R. Tycko (2001), "Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form." *J. Mol. Biol.*, 313, 845–859.

Bogerd, H. and W. C. Greene (1993), "Dominant negative mutants of human T-cell leukemia virus type I Rex and human immunodeficiency virus type 1 Rev fail to multimerize in vivo." *J. Virol.*, 67, 2496–2502.

Bottcher, B., S. A. Wynne, and R. A. Crowther (1997), "Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy." *Nature*, 386, 88–91.

Brown, P. H. and P. Schuck (2006), "Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation." *Biophys. J.*, 90, 4651–4661.

Carpenter, C. C., M. A. Fischl, S. M. Hammer, M. S. Hirsch, D. M. Jacobsen, D. A. Katzenstein, J. S. Montaner, D. D. Richman, M. S. Saag, R. T. Schooley, M. A. Thompson, S. Vella, P. G. Yeni, and P. A. Volberding (1996), "Antiretroviral therapy for HIV infection in 1996. Recommendations of an international panel. International AIDS Society-USA." *JAMA*, 276, 146–154.

Ceres, P. and A. Zlotnick (2002), "Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids." *Biochemistry*, 41, 11525–11531.

Chang, C., G. Enders, R. Sprengel, N. Peters, H. E. Varmus, and D. Ganem (1987), "Expression of the precore region of an avian hepatitis B virus is not required for viral replication." *J. Virol.*, 61, 3322–3325.

Chen, H. S., M. C. Kew, W. E. Hornbuckle, B. C. Tennant, P. J. Cote, J. L. Gerin, R. H. Purcell, and R. H. Miller (1992), "The precore gene of the woodchuck hepatitis virus genome is not essential for viral replication in the natural host." *J. Virol.*, 66, 5682–5684.

Chen, M., M. Sallberg, J. Hughes, J. Jones, L. G. Guidotti, F. V. Chisari, J. N. Billaud, and D. R. Milich (2005), "Immune tolerance split between hepatitis B virus precore and core proteins." *J. Virol.*, 79, 3016–3027.

Chen, M. T., J. N. Billaud, M. Sallberg, L. G. Guidotti, F. V. Chisari, J. Jones, J. Hughes, and D. R. Milich (2004), "A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen." *Proc. Natl. Acad. Sci. U.S.A.*, 101, 14913–14918.

Chook, Y. M. and K. E. Suel (2011), "Nuclear import by karyopherin-ÃŐÂšs: recognition and inhibition." *Biochim. Biophys. Acta*, 1813, 1593–1606.

Clackson, T. and J. A. Wells (1995), "A hot spot of binding energy in a hormone-receptor interface." *Science*, 267, 383–386.

Cochrane, A., L. L. Murley, M. Gao, R. Wong, K. Clayton, N. Brufatto, V. Canadien, D. Mamelak, T. Chen, D. Richards, M. Zeghouf, J. Greenblatt, C. Burks, and L. Frappier (2009), "Stable complex formation between HIV Rev and the nucleosome assembly protein, NAP1, affects Rev function." *Virology*, 388, 103–111.

Cochrane, A. W., A. Perkins, and C. A. Rosen (1990), "Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function." *J. Virol.*, 64, 881–885.

Cole, J. L., J. D. Gehman, J. A. Shafer, and L. C. Kuo (1993), "Solution oligomerization of the rev protein of HIV-1: implications for function." *Biochemistry*, 32, 11769–11775.

Colman, P. M., W. R. Tulip, J. N. Varghese, P. A. Tulloch, A. T. Baker, W. G. Laver, G. M. Air, and R. G. Webster (1989), "Three-dimensional structures of influenza virus neuraminidase-antibody complexes." *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 323, 511–518.

Conway, J. F., N. Cheng, A. Zlotnick, S. J. Stahl, P. T. Wingfield, D. M. Belnap, U. Kanngiesser, M. Noah, and A. C. Steven (1998), "Hepatitis B virus capsid: localization of the putative immunodominant loop (residues 78 to 83) on the capsid surface, and implications for the distinction between c and e-antigens." *J. Mol. Biol.*, 279, 1111–1121.

Conway, J. F., N. Cheng, A. Zlotnick, P. T. Wingfield, S. J. Stahl, and A. C. Steven (1997), "Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy." *Nature*, 386, 91–94.

Cook, K. S., G. J. Fisk, J. Hauber, N. Usman, T. J. Daly, and J. R. Rusche (1991), "Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate." *Nucleic Acids Res.*, 19, 1577–1583.

Cronshaw, J. M., A. N. Krutchinsky, W. Zhang, B. T. Chait, and M. J. Matunis (2002), "Proteomic analysis of the mammalian nuclear pore complex." *J. Cell Biol.*, 158, 915–927.

Cullen, B. R. (2003), "Nuclear mRNA export: insights from virology." *Trends Biochem. Sci.*, 28, 419–424.

Daly, T. J., K. S. Cook, G. S. Gray, T. E. Maione, and J. R. Rusche (1989), "Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro." *Nature*, 342, 816–819.

Daly, T. J., R. C. Doten, P. Rennert, M. Auer, H. Jaksche, A. Donner, G. Fisk, and J. R. Rusche (1993), "Biochemical characterization of binding of multiple HIV-1 Rev monomeric proteins to the Rev responsive element." *Biochemistry*, 32, 10497–10505.

Daugherty, M. D., D. S. Booth, B. Jayaraman, Y. Cheng, and A. D. Frankel (2010a), "HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes." *Proc. Natl. Acad. Sci. U.S.A.*, 107, 12481–12486.

Daugherty, M. D., I. D'Orso, and A. D. Frankel (2008), "A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA." *Mol. Cell*, 31, 824–834.

Daugherty, M. D., B. Liu, and A. D. Frankel (2010b), "Structural basis for cooperative RNA binding and export complex assembly by HIV Rev." *Nat. Struct. Mol. Biol.*, 17, 1337–1342.

Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, and D. C. Richardson (2007), "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." *Nucleic Acids Res.*, 35, W375–383.

Davis, L. I. and G. Blobel (1986), "Identification and characterization of a nuclear pore complex protein." *Cell*, 45, 699–709.

Dayton, A. I. (2004), "Within you, without you: HIV-1 Rev and RNA export." *Retrovirology*, 1, 35.

DeGrasse, J. A., K. N. DuBois, D. Devos, T. N. Siegel, A. Sali, M. C. Field, M. P. Rout, and B. T. Chait (2009), "Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor." *Mol. Cell Proteomics*, 8, 2119–2130.

Derewenda, Z. S. (2011), "It's all in the crystals." *Acta Crystallogr. D Biol. Crystallogr.*, 67, 243–248.

Dey, S., A. Pal, P. Chakrabarti, and J. Janin (2010), "The subunit interfaces of weakly associated homodimeric proteins." *J. Mol. Biol.*, 398, 146–160.

DiMattia, M. A., N. R. Watts, S. J. Stahl, C. Rader, P. T. Wingfield, D. I. Stuart, A. C. Steven, and J. M. Grimes (2010), "Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element." *Proc. Natl. Acad. Sci. U.S.A.*, 107, 5810–5814.

Ding, J., K. Das, Y. Hsiou, S. G. Sarafianos, A. D. Clark, A. Jacobo-Molina, C. Tantillo, S. H. Hughes, and E. Arnold (1998), "Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 A resolution." *J. Mol. Biol.*, 284, 1095–1111.

Dunker, A. K., C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovi? (2002), "Intrinsic disorder and protein function." *Biochemistry*, 41, 6573–6582.

Edgcomb, S. P., A. Aschrafi, E. Kompfner, J. R. Williamson, L. Gerace, and M. Hennig (2008), "Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes." *Protein Sci.*, 17, 420–430.

Egelman, E. H. (2000), "A robust algorithm for the reconstruction of helical filaments using single-particle methods." *Ultramicroscopy*, 85, 225–234.

Egelman, E. H. (2007), "The iterative helical real space reconstruction method: surmounting the problems posed by real polymers." *J. Struct. Biol.*, 157, 83–94.

Elfgang, C., O. Rosorius, L. Hofer, H. Jaksche, J. Hauber, and D. Bevec (1999), "Evidence for specific nucleocytoplasmic transport pathways used by leucine-rich nuclear export signals." *Proc. Natl. Acad. Sci. U.S.A.*, 96, 6229–6234.

Elgouhari, H. M., T. I. Abu-Rajab Tamimi, and W. D. Carey (2008), "Hepatitis B virus infection: understanding its epidemiology, course, and diagnosis." *Cleve Clin J Med*, 75, 881–889.

Emerman, M., R. Vazeux, and K. Peden (1989), "The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization." *Cell*, 57, 1155–1165.

Emsley, P. and K. Cowtan (2004), "Coot: model-building tools for molecular graphics." *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2126–2132.

Fagan, E. A., P. M. Smith, F. Davison, and R. Williams (1986), "Fulminant hepatitis B in successive female sexual partners of two anti-HBe-positive males." *Lancet*, 2, 538–540.

Fahrenkrog, B. and U. Aebi (2003), "The nuclear pore complex: nucleocytoplasmic transport and beyond." *Nat. Rev. Mol. Cell Biol.*, 4, 757–766.

Fang, J., E. Acheampong, R. Dave, F. Wang, M. Mukhtar, and R. J. Pomerantz (2005), "The RNA helicase DDX1 is involved in restricted HIV-1 Rev function in human astrocytes." *Virology*, 336, 299–307.

Fankhauser, C., E. Izaurralde, Y. Adachi, P. Wingfield, and U. K. Laemmli (1991), "Specific complex of human immunodeficiency virus type 1 rev and nucleolar B23 proteins: dissociation by the Rev response element." *Mol. Cell. Biol.*, 11, 2567–2575.

Feinberg, M. B. and W. C. Greene (1992), "Molecular insights into human immunodeficiency virus type 1 pathogenesis." *Curr. Opin. Immunol.*, 4, 466–474.

Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong-Staal (1986), "HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA." *Cell*, 46, 807–817.

Ferns, R. B. and R. S. Tedder (1984), "Monoclonal antibodies to hepatitis Be antigen (HBeAg) derived from hepatitis B core antigen (HBcAg): their use in characterization and detection of HBeAg." *J. Gen. Virol.*, 65 (Pt 5), 899–908.

Ferns, R. B. and R. S. Tedder (1986), "Human and monoclonal antibodies to hepatitis B core antigen recognise a single immunodominant epitope." *J. Med. Virol.*, 19, 193–203.

Fink, A. L. (2005), "Natively unfolded proteins." *Curr. Opin. Struct. Biol.*, 15, 35–41.

Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Luhrmann (1995), "The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs." *Cell*, 82, 475–483.

Fong, J. H., B. A. Shoemaker, S. O. Garbuzynskiy, M. Y. Lobanov, O. V. Galzitskaya, and A. R. Panchenko (2009), "Intrinsic disorder in protein interactions: insights from a comprehensive structural analysis." *PLoS Comput. Biol.*, 5, e1000316.

Fornerod, M., M. Ohno, M. Yoshida, and I. W. Mattaj (1997), "CRM1 is an export receptor for leucine-rich nuclear export signals." *Cell*, 90, 1051–1060.

Forwood, J. K., A. Lange, U. Zachariae, M. Marfori, C. Preast, H. Grubmuller, M. Stewart, A. H. Corbett, and B. Kobe (2010), "Quantitative structural analysis of importin-ÃŐÂš flexibility: paradigm for solenoid protein structures." *Structure*, 18, 1171–1183.

Freed, E. O. (2001), "HIV-1 replication." *Somat. Cell Mol. Genet.*, 26, 13–33.

Frishman, D. and P. Argos (1995), "Knowledge-based protein secondary structure assignment." *Proteins*, 23, 566–579.

Fritz, C. C. and M. R. Green (1996), "HIV Rev uses a conserved cellular protein export pathway for the nucleocytoplasmic transport of viral RNAs." *Curr. Biol.*, 6, 848–854.

Fukuda, M., S. Asano, T. Nakamura, M. Adachi, M. Yoshida, M. Yanagida, and E. Nishida (1997), "CRM1 is responsible for intracellular transport mediated by the nuclear export signal." *Nature*, 390, 308–311.

Gentile, M., T. Adrian, A. Scheidler, M. Ewald, F. Dianzani, G. Pauli, and H. R. Gelderblom (1994), "Determination of the size of HIV using adenovirus type 2 as an internal length marker." *J. Virol. Methods*, 48, 43–52.

Gerace, L. (1995), "Nuclear export signals and the fast track to the cytoplasm." *Cell*, 82, 341–344.

Gilks, C. F., S. Crowley, R. Ekpini, S. Gove, J. Perriens, Y. Souteyrand, D. Sutherland, M. Vitoria, T. Guerma, and K. De Cock (2006), "The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings." *Lancet*, 368, 505–510.

Gorlich, D. and I. W. Mattaj (1996), "Nucleocytoplasmic transport." *Science*, 271, 1513–1518.

Grabar, S., V. Le Moing, C. Goujard, C. Leport, M. D. Kazatchkine, D. Costagliola, and L. Weiss (2000), "Clinical outcome of patients with HIV-1 infection according to immunologic and virologic response after 6 months of highly active antiretroviral therapy." *Ann. Intern. Med.*, 133, 401–410.

Grunwald, D., R. H. Singer, and M. Rout (2011), "Nuclear export dynamics of RNA-protein complexes." *Nature*, 475, 333–341.

Gunasekaran, K., C. J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov (2003), "Extended disordered proteins: targeting function with less scaffold." *Trends Biochem. Sci.*, 28, 81–85.

Havlin, R. H., F. J. Blanco, and R. Tycko (2007), "Constraints on protein structure in HIV-1 Rev and Rev-RNA supramolecular assemblies from two-dimensional solid state nuclear magnetic resonance." *Biochemistry*, 46, 3586–3593.

Hazenberg, M. D., S. A. Otto, B. H. van Benthem, M. T. Roos, R. A. Coutinho, J. M. Lange, D. Hamann, M. Prins, and F. Miedema (2003), "Persistent immune activation in HIV-1 infection is associated with progression to AIDS." *AIDS*, 17, 1881–1888.

Heaphy, S., J. T. Finch, M. J. Gait, J. Karn, and M. Singh (1991), "Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs." *Proc. Natl. Acad. Sci. U.S.A.*, 88, 7366–7370.

Heinig, M. and D. Frishman (2004), "STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins." *Nucleic Acids Res.*, 32, W500–502.

Henderson, B. R. and P. Percipalle (1997), "Interactions between HIV Rev and nuclear import and export factors: the Rev nuclear localisation signal mediates specific binding to human importin-beta." *J. Mol. Biol.*, 274, 693–707.

Heymann, J. B. and D. M. Belnap (2007), "Bsoft: image processing and molecular modeling for electron microscopy." *J. Struct. Biol.*, 157, 3–18.

Hoelz, A., E. W. Debler, and G. Blobel (2011), "The structure of the nuclear pore complex." *Annu. Rev. Biochem.*, 80, 613–643.

Hofmann, W., B. Reichart, A. Ewald, E. Muller, I. Schmitt, R. H. Stauber, F. Lottspeich, B. M. Jockusch, U. Scheer, J. Hauber, and M. C. Dabauvalle (2001), "Cofactor requirements for nuclear export of Rev response element (RRE)- and constitutive transport element (CTE)-containing retroviral RNAs. An unexpected role for actin." *J. Cell Biol.*, 152, 895–910.

Hope, T. J., N. P. Klein, M. E. Elder, and T. G. Parslow (1992), "trans-dominant inhibition of human immunodeficiency virus type 1 Rev occurs through formation of inactive protein complexes." *J. Virol.*, 66, 1849–1855.

Imai, M., M. Nomura, T. Gotanda, T. Sano, K. Tachibana, H. Miyamoto, K. Takahashi, S. Toyama, Y. Miyakawa, and M. Mayumi (1982), "Demonstration of two distinct antigenic determinants on hepatitis B e antigen by monoclonal antibodies." *J. Immunol.*, 128, 69–72.

Ionel, A., J. A. Velazquez-Muriel, D. Luque, A. Cuervo, J. R. Caston, J. M. Valpuesta, J. Martin-Benito, and J. L. Carrascosa (2011), "Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7." *J. Biol. Chem.*, 286, 234–242.

Iwai, S., C. Pritchard, D. A. Mann, J. Karn, and M. J. Gait (1992), "Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein." *Nucleic Acids Res.*, 20, 6465–6472.

Jain, C. and J. G. Belasco (2001), "Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants." *Mol. Cell*, 7, 603–614.

Jaroszewski, L., L. Rychlewski, Z. Li, W. Li, and A. Godzik (2005), "FFAS03: a server for profile–profile sequence alignments." *Nucleic Acids Res.*, 33, W284–288.

Jones, S. (2012), "Computational and structural characterisation of protein associations." *Adv. Exp. Med. Biol.*, 747, 42–54.

Jovanovic-Talisman, T., J. Tetenbaum-Novatt, A. S. McKenney, A. Zilman, R. Peters, M. P. Rout, and B. T. Chait (2009), "Artificial nanopores that mimic the transport selectivity of the nuclear pore complex." *Nature*, 457, 1023–1027.

Kabsch, W. (2010), "XDS." *Acta Crystallogr. D Biol. Crystallogr.*, 66, 125–132.

Kandiah, E., N. R. Watts, N. Cheng, G. Cardone, S. J. Stahl, T. Heller, T. J. Liang, P. T. Wingfield, and A. C. Steven (2012), "Cryo-EM study of Hepatitis B virus core antigen capsids decorated with antibodies from a human patient." *J. Struct. Biol.*, 177, 145–151.

Kay, B. K., M. P. Williamson, and M. Sudol (2000), "The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains." *FASEB J.*, 14, 231–241.

Kim, S. Y., R. Byrn, J. Groopman, and D. Baltimore (1989), "Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression." *J. Virol.*, 63, 3708–3713.

Kjems, J., B. J. Calnan, A. D. Frankel, and P. A. Sharp (1992), "Specific binding of a basic peptide from HIV-1 Rev." *EMBO J.*, 11, 1119–1129.

Klasse, P. J. (2012), "The molecular basis of HIV entry." *Cell. Microbiol.*, 14, 1183–1192.

Kobiler, O., N. Drayman, V. Butin-Israeli, and A. Oppenheim (2012), "Virus strategies for passing the nuclear envelope barrier." *Nucleus*, 3, 526–539.

Kohler, A. and E. Hurt (2007), "Exporting RNA from the nucleus to the cytoplasm." *Nat. Rev. Mol. Cell Biol.*, 8, 761–773.

Krishnan, L. and A. Engelman (2012), "Retroviral integrase proteins and HIV-1 DNA integration." *J. Biol. Chem.*, 287, 40858–40866.

Krissinel, E. and K. Henrick (2007), "Inference of macromolecular assemblies from crystalline state." *J. Mol. Biol.*, 372, 774–797.

Kubota, S., R. Furuta, M. Maki, and M. Hatanaka (1992), "Inhibition of human immunodeficiency virus type 1 Rev function by a Rev mutant which interferes with nuclear/nucleolar localization of Rev." *J. Virol.*, 66, 2510–2513.

Kuttner, Y. Y. and S. Engel (2012), "Protein hot spots: the islands of stability." *J. Mol. Biol.*, 415, 419–428.

Lang, T., C. Lo, N. Skinner, S. Locarnini, K. Visvanathan, and A. Mansell (2011), "The hepatitis B e antigen (HBeAg) targets and suppresses activation of the toll-like receptor signaling pathway." *J. Hepatol.*, 55, 762–769.

Lawrence, M. C. and P. M. Colman (1993), "Shape complementarity at protein/protein interfaces." *J. Mol. Biol.*, 234, 946–950.

Lazdina, U., T. Cao, J. Steinbergs, M. Alheim, P. Pumpens, D. L. Peterson, D. R. Milich, G. Leroux-Roels, and M. Sallberg (2001), "Molecular basis for the interaction of the hepatitis B virus core antigen with the surface immunoglobulin receptor on naive B cells." *J. Virol.*, 75, 6367–6374.

Lee, S. K., M. Potempa, and R. Swanstrom (2012), "The choreography of HIV-1 proteolytic processing and virion assembly." *J. Biol. Chem.*, 287, 40867–40874.

Levin, A., Z. Hayouka, A. Friedler, and A. Loyter (2010), "Nucleocytoplasmic shuttling of HIV-1 integrase is controlled by the viral Rev protein." *Nucleus*, 1, 190–201.

Li, L., H. S. Li, C. D. Pauza, M. Bukrinsky, and R. Y. Zhao (2005), "Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions." *Cell Res.*, 15, 923–934.

Liang, T. J., K. Hasegawa, N. Rimon, J. R. Wands, and E. Ben-Porath (1991), "A hepatitis B virus mutant associated with an epidemic of fulminant hepatitis." *N. Engl. J. Med.*, 324, 1705–1709.

Liu, J., J. Henao-Mejia, H. Liu, Y. Zhao, and J. J. He (2011), "Translational regulation of HIV-1 replication by HIV-1 Rev cellular cofactors Sam68, eIF5A, hRIP, and DDX3." *J Neuroimmune Pharmacol*, 6, 308–321.

Lu, K., X. Heng, and M. F. Summers (2011), "Structural determinants and mechanism of HIV-1 genome packaging." *J. Mol. Biol.*, 410, 609–633.

MacKay, P., J. Lees, and K. Murray (1981), "The conversion of hepatitis B core antigen synthesized in E coli into e antigen." *J. Med. Virol.*, 8, 237–243.

Maddrey, W. C. (2000), "Hepatitis B: an important public health issue." *J. Med. Virol.*, 61, 362–366.

Magnius, L. O. and J. A. Espmark (1972), "New specificities in Australia antigen positive sera distinct from the Le Bouvier determinants." *J. Immunol.*, 109, 1017–1021.

Malim, M. H., S. Bohnlein, J. Hauber, and B. R. Cullen (1989a), "Functional dissection of the HIV-1 Rev trans-activator–derivation of a trans-dominant repressor of Rev function." *Cell*, 58, 205–214.

Malim, M. H. and B. R. Cullen (1991), "HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency." *Cell*, 65, 241–248.

Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen (1989b), "The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA." *Nature*, 338, 254–257.

Malim, M. H., L. S. Tiley, D. F. McCarn, J. R. Rusche, J. Hauber, and B. R. Cullen (1990), "HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence." *Cell*, 60, 675–683.

Mann, D. A., I. Mikaelian, R. W. Zemmel, S. M. Green, A. D. Lowe, T. Kimura, M. Singh, P. J. Butler, M. J. Gait, and J. Karn (1994), "A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression." *J. Mol. Biol.*, 241, 193–207.

Marfori, M., A. Mynott, J. J. Ellis, A. M. Mehdi, N. F. Saunders, P. M. Curmi, J. K. Forwood, M. Boden, and B. Kobe (2011), "Molecular basis for specificity of nuclear import and prediction of nuclear localization." *Biochim. Biophys. Acta*, 1813, 1562–1577.

Mattaj, I. W. and L. Englmeier (1998), "Nucleocytoplasmic transport: the soluble phase." *Annu. Rev. Biochem.*, 67, 265–306.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read (2007), "Phaser crystallographic software." *J Appl Crystallogr*, 40, 658–674.

Mehlin, H., B. Daneholt, and U. Skoglund (1992), "Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography." *Cell*, 69, 605–613.

Miedema, F., M. Tersmette, and R. A. van Lier (1990), "AIDS pathogenesis: a dynamic interaction between HIV and the immune system." *Immunol. Today*, 11, 293–297.

Milich, D. and T. J. Liang (2003a), "Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection." *Hepatology*, 38, 1075–1086.

Milich, D. and T. J. Liang (2003b), "Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection." *Hepatology*, 38, 1075–1086.

Milich, D. R., M. Chen, F. Schodel, D. L. Peterson, J. E. Jones, and J. L. Hughes (1997a), "Role of B cells in antigen presentation of the hepatitis B core." *Proc. Natl. Acad. Sci. U.S.A.*, 94, 14648–14653.

Milich, D. R., J. E. Jones, J. L. Hughes, J. Price, A. K. Raney, and A. McLachlan (1990), "Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance in utero?" *Proc. Natl. Acad. Sci. U.S.A.*, 87, 6599–6603.

Milich, D. R. and A. McLachlan (1986), "The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen." *Science*, 234, 1398–1401.

Milich, D. R., F. Schodel, J. L. Hughes, J. E. Jones, and D. L. Peterson (1997b), "The hepatitis B virus core and e antigens elicit different Th cell subsets: antigen structure can affect Th cell phenotype." *J. Virol.*, 71, 2192–2201.

Moeller, A., R. N. Kirchdoerfer, C. S. Potter, B. Carragher, and I. A. Wilson (2012), "Organization of the influenza virus replication machinery." *Science*, 338, 1631–1634.

Moore, M. S. and G. Blobel (1993), "The GTP-binding protein Ran/TC4 is required for protein import into the nucleus." *Nature*, 365, 661–663.

Moroianu, J., G. Blobel, and A. Radu (1996), "Nuclear protein import: Ran-GTP dissociates the karyopherin alphabeta heterodimer by displacing alpha from an overlapping binding site on beta." *Proc. Natl. Acad. Sci. U.S.A.*, 93, 7059–7062.

Morrison, C., P. Porcu, M. A. Caliguiri, and G. J. Nuovo (2001), "In situ determination of B-cell heavy chain and kappa/lambda light chain expression patterns: methodology and clinical utility." *Diagn. Mol. Pathol.*, 10, 171–178.

Mosca, R., B. Brannetti, and T. R. Schneider (2008), "Alignment of protein structures in the presence of domain motions." *BMC Bioinformatics*, 9, 352.

Nakielny, S. and G. Dreyfuss (1999), "Transport of proteins and RNAs in and out of the nucleus." *Cell*, 99, 677–690.

Nassal, M. and A. Rieger (1993), "An intramolecular disulfide bridge between Cys-7 and Cys61 determines the structure of the secretory core gene product (e antigen) of hepatitis B virus." *J. Virol.*, 67, 4307–4315.

Nekhai, S. and K. T. Jeang (2006), "Transcriptional and post-transcriptional regulation of HIV-1 gene expression: role of cellular factors for Tat and Rev." *Future Microbiol*, 1, 417–426.

Neville, M., F. Stutz, L. Lee, L. I. Davis, and M. Rosbash (1997), "The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export." *Curr. Biol.*, 7, 767–775.

Newton, K., M. L. Matsumoto, I. E. Wertz, D. S. Kirkpatrick, J. R. Lill, J. Tan, D. Dugger, N. Gordon, S. S. Sidhu, F. A. Fellouse, L. Komuves, D. M. French, R. E. Ferrando, C. Lam, D. Compaan, C. Yu, I. Bosanac, S. G. Hymowitz, R. F. Kelley, and V. M. Dixit (2008), "Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies." *Cell*, 134, 668–678.

Otwinowski, Z. and W. Minor (1997), *Processing of X-ray diffraction data collected in oscillation mode*, volume 276: Macromolecular Crystallography, Part A of *Methods in Enzymology*, 307–326. Academic Press, New York.

Ou, J. H. (1997), "Molecular biology of hepatitis B virus e antigen." *J. Gastroenterol. Hepatol.*, 12, S178–187.

Ou, J. H., O. Laub, and W. J. Rutter (1986), "Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen." *Proc. Natl. Acad. Sci. U.S.A.*, 83, 1578–1582.

Packianathan, C., S. P. Katen, C. E. Dann, and A. Zlotnick (2010), "Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly." *J. Virol.*, 84, 1607–1615.

Padian, N. S., S. I. McCoy, S. S. Karim, N. Hasen, J. Kim, M. Bartos, E. Katabira, S. M. Bertozzi, B. Schwartlander, and M. S. Cohen (2011), "HIV prevention transformed: the new prevention research agenda." *Lancet*, 378, 269–278.

Pante, N. and M. Kann (2002), "Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm." *Mol. Biol. Cell*, 13, 425–434.

Pasquinelli, A. E., R. K. Ernst, E. Lund, C. Grimm, M. L. Zapp, D. Rekosh, M. L. Hammarskjold, and J. E. Dahlberg (1997), "The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway." *EMBO J.*, 16, 7500–7510.

Pemberton, L. F. and B. M. Paschal (2005), "Mechanisms of receptor-mediated nuclear import and nuclear export." *Traffic*, 6, 187–198.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin (2004), "UCSF Chimera–a visualization system for exploratory research and analysis." *J Comput Chem*, 25, 1605–1612.

Piller, S. C., L. Caly, and D. A. Jans (2003), "Nuclear import of the pre-integration complex (PIC): the Achilles heel of HIV?" *Curr Drug Targets*, 4, 409–429.

Pollard, V. W. and M. H. Malim (1998), "The HIV-1 Rev protein." *Annu. Rev. Microbiol.*, 52, 491–532.

Pomerantz, R. J. and D. L. Horn (2003), "Twenty years of therapy for HIV-1 infection." *Nat. Med.*, 9, 867–873.

Pond, S. J., W. K. Ridgeway, R. Robertson, J. Wang, and D. P. Millar (2009), "HIV-1 Rev protein assembles on viral RNA one molecule at a time." *Proc. Natl. Acad. Sci. U.S.A.*, 106, 1404–1408.

Purvina, M., A. Hoste, J. M. Rossignol, and C. Lagaudriere-Gesbert (2012), "Human hepatitis B viral e antigen and its precursor P20 inhibit T lymphocyte proliferation." *Biochem. Biophys. Res. Commun.*, 417, 1310–1315.

Rasmussen, S. G., H. J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. Devree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Konetzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis, and B. K. Kobilka (2011), "Structure of a nanobody-stabilized active state of the ÃŐÂš(2) adrenoceptor." *Nature*, 469, 175–180.

Revill, P., L. Yuen, R. Walsh, M. Perrault, S. Locarnini, and A. Kramvis (2010), "Bioinformatic analysis of the hepadnavirus e-antigen and its precursor identifies remarkable sequence conservation in all orthohepadnaviruses." *J. Med. Virol.*, 82, 104–115.

Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker (2001), "Sequence complexity of disordered protein." *Proteins*, 42, 38–48.

Rout, M. P., J. D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, and B. T. Chait (2000), "The yeast nuclear pore complex: composition, architecture, and transport mechanism." *J. Cell Biol.*, 148, 635–651.

Saavedra, C., B. Felber, and E. Izaurralde (1997), "The simian retrovirus-1 constitutive transport element, unlike the HIV-1 RRE, uses factors required for cellular mRNA export." *Curr. Biol.*, 7, 619–628.

Salfeld, J., E. Pfaff, M. Noah, and H. Schaller (1989), "Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus." *J. Virol.*, 63, 798–808.

Schlicht, H. J., J. Salfeld, and H. Schaller (1987), "The duck hepatitis B virus pre-C region encodes a signal sequence which is essential for synthesis and secretion of processed core proteins but not for virus formation." *J. Virol.*, 61, 3701–3709.

Schodel, F., D. Peterson, J. Zheng, J. E. Jones, J. L. Hughes, and D. R. Milich (1993), "Structure of hepatitis B virus core and e-antigen. A single precore amino acid prevents nucleocapsid assembly." *J. Biol. Chem.*, 268, 1332–1337.

Seto, W. K., J. Fung, M. F. Yuen, and C. L. Lai (2012), "Future prevention and treatment of chronic hepatitis B infection." *J. Clin. Gastroenterol.*, 46, 725–734.

Shattock, R. J., M. Warren, S. McCormack, and C. A. Hankins (2011), "AIDS. Turning the tide against HIV." *Science*, 333, 42–43.

Silvan, U., C. Boiteux, R. Sutterlin, U. Schroeder, H. G. Mannherz, B. M. Jockusch, S. Berneche, U. Aebi, and C. A. Schoenenberger (2012), "An antiparallel actin dimer is associated with the endocytic pathway in mammalian cells." *J. Struct. Biol.*, 177, 70–80.

Simon, S. M., F. J. Sousa, R. Mohana-Borges, and G. C. Walker (2008), "Regulation of Escherichia coli SOS mutagenesis by dimeric intrinsically disordered umuD gene products." *Proc. Natl. Acad. Sci. U.S.A.*, 105, 1152–1157.

Smart, O. S., T. O. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vonrhein, and G. Bricogne (2012), "Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER." *Acta Crystallogr. D Biol. Crystallogr.*, 68, 368–380.

Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine (1986), "A second post-transcriptional trans-activator gene required for HTLV-III replication." *Nature*, 321, 412–417.

Stade, K., C. S. Ford, C. Guthrie, and K. Weis (1997), "Exportin 1 (Crm1p) is an essential nuclear export factor." *Cell*, 90, 1041–1050.

Stahl, S. J., N. R. Watts, C. Rader, M. A. DiMattia, R. G. Mage, I. Palmer, J. D. Kaufman, J. M. Grimes, D. I. Stuart, A. C. Steven, and P. T. Wingfield (2010), "Generation and characterization of a chimeric rabbit/human Fab for co-crystallization of HIV-1 Rev." *J. Mol. Biol.*, 397, 697–708.

Standring, D. N., J. H. Ou, F. R. Masiarz, and W. J. Rutter (1988), "A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in Xenopus oocytes." *Proc. Natl. Acad. Sci. U.S.A.*, 85, 8405–8409.

Stanfield, R. L., A. Zemla, I. A. Wilson, and B. Rupp (2006), "Antibody elbow angles are influenced by their light chain class." *J. Mol. Biol.*, 357, 1566–1574.

Steven, A. C. and W. Baumeister (2008), "The future is hybrid." *J. Struct. Biol.*, 163, 186–195.

Steven, A. C., J. F. Conway, N. Cheng, N. R. Watts, D. M. Belnap, A. Harris, S. J. Stahl, and P. T. Wingfield (2005), "Structure, assembly, and antigenicity of hepatitis B virus capsid proteins." *Adv. Virus Res.*, 64, 125–164.

Strawn, L. A., T. Shen, N. Shulga, D. S. Goldfarb, and S. R. Wente (2004), "Minimal nuclear pore complexes define FG repeat domains essential for transport." *Nat. Cell Biol.*, 6, 197–206.

Strebel, K. (2003), "Virus-host interactions: role of HIV proteins Vif, Tat, and Rev." *AIDS*, 17 Suppl 4, 25–34.

Suhasini, M. and T. R. Reddy (2009), "Cellular proteins and HIV-1 Rev function." *Curr. HIV Res.*, 7, 91–100.

Szebeni, A., B. Mehrotra, A. Baumann, S. A. Adam, P. T. Wingfield, and M. O. Olson (1997), "Nucleolar protein B23 stimulates nuclear import of the HIV-1 Rev protein and NLS-conjugated albumin." *Biochemistry*, 36, 3941–3949.

Szilvay, A. M., K. A. Brokstad, S. O. B?e, G. Haukenes, and K. H. Kalland (1997), "Oligomerization of HIV-1 Rev mutants in the cytoplasm and during nuclear import." *Virology*, 235, 73–81.

Takahashi, K., A. Machida, G. Funatsu, M. Nomura, S. Usuda, S. Aoyagi, K. Tachibana, H. Miyamoto, M. Imai, T. Nakamura, Y. Miyakawa, and M. Mayumi (1983), "Immunochemical structure of hepatitis B e antigen in the serum." *J. Immunol.*, 130, 2903–2907.

Tawar, R. G., S. Duquerroy, C. Vonrhein, P. F. Varela, L. Damier-Piolle, N. Castagne, K. MacLellan, H. Bedouelle, G. Bricogne, D. Bhella, J. F. Eleouet, and F. A. Rey (2009), "Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus." *Science*, 326, 1279–1283.

Terazawa, S., M. Kojima, T. Yamanaka, S. Yotsumoto, H. Okamoto, F. Tsuda, Y. Miyakawa, and M. Mayumi (1991), "Hepatitis B virus mutants with precore-region defects in two babies with fulminant hepatitis and their mothers positive for antibody to hepatitis B e antigen." *Pediatr. Res.*, 29, 5–9.

Thanos, C. D., W. L. DeLano, and J. A. Wells (2006), "Hot-spot mimicry of a cytokine receptor by a small molecule." *Proc. Natl. Acad. Sci. U.S.A.*, 103, 15422–15427.

Thomas, S. L., M. Oft, H. Jaksche, G. Casari, P. Heger, M. Dobrovnik, D. Bevec, and J. Hauber (1998), "Functional analysis of the human immunodeficiency virus type 1 Rev protein oligomerization interface." *J. Virol.*, 72, 2935–2944.

Tiley, L. S., M. H. Malim, H. K. Tewary, P. G. Stockley, and B. R. Cullen (1992), "Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein." *Proc. Natl. Acad. Sci. U.S.A.*, 89, 758–762.

Trikha, R. and D. W. Brighty (2005), "Phenotypic analysis of human immunodeficiency virus type 1 Rev trimerization-interface mutants in human cells." *J. Gen. Virol.*, 86, 1509–1513.

Truant, R. and B. R. Cullen (1999), "The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals." *Mol. Cell. Biol.*, 19, 1210–1217.

Uysal, S., V. Vasquez, V. Tereshko, K. Esaki, F. A. Fellouse, S. S. Sidhu, S. Koide, E. Perozo, and A. Kossiakoff (2009), "Crystal structure of full-length KcsA in its closed conformation." *Proc. Natl. Acad. Sci. U.S.A.*, 106, 6644–6649.

van Heel, M. and G. Harauz (1988), "Biological macromolecules explored by pattern recognition." *Scanning Microsc. Suppl.*, 2, 295–301.

Vanlandschoot, P., T. Cao, and G. Leroux-Roels (2003), "The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure." *Antiviral Res.*, 60, 67–74.

Venken, T., D. Daelemans, M. De Maeyer, and A. Voet (2012), "Computational investigation of the HIV-1 Rev multimerization using molecular dynamics simulations and binding free energy calculations." *Proteins*, 80, 1633–1646.

Visvanathan, K., N. A. Skinner, A. J. Thompson, S. M. Riordan, V. Sozzi, R. Edwards, S. Rodgers, J. Kurtovic, J. Chang, S. Lewin, P. Desmond, and S. Locarnini (2007), "Regulation of Toll-like receptor-2 expression in chronic hepatitis B by the precore protein." *Hepatology*, 45, 102–110.

Votteler, J. and U. Schubert (2008), *Human Immunodeficiency Viruses: Molecular Biology*, 515–525. Encyclopedia of Virology (3rd Ed).

Walsh, R. and S. Locarnini (2012), "Hepatitis B precore protein: pathogenic potential and therapeutic promise." *Yonsei Med. J.*, 53, 875–885.

Walter, T. S., J. M. Diprose, C. J. Mayo, C. Siebold, M. G. Pickford, L. Carter, G. C. Sutton, N. S. Berrow, J. Brown, I. M. Berry, G. B. Stewart-Jones, J. M. Grimes, D. K. Stammers, R. M. Esnouf, E. Y. Jones, R. J. Owens, D. I. Stuart, and K. Harlos (2005), "A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization." *Acta Crystallogr. D Biol. Crystallogr.*, 61, 651–657.

Wasenauer, G., J. Kock, and H. J. Schlicht (1992), "A cysteine and a hydrophobic sequence in the noncleaved portion of the pre-C leader peptide determine the biophysical properties of the secretory core protein (HBe protein) of human hepatitis B virus." *J. Virol.*, 66, 5338–5346.

Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess, R. Swanstrom, C. L. Burch, and K. M. Weeks (2009), "Architecture and secondary structure of an entire HIV-1 RNA genome." *Nature*, 460, 711–716.

Watts, N. R., J. F. Conway, N. Cheng, S. J. Stahl, A. C. Steven, and P. T. Wingfield (2011), "Role of the propeptide in controlling conformation and assembly state of hepatitis B virus e-antigen." *J. Mol. Biol.*, 409, 202–213.

Watts, N. R., M. Misra, P. T. Wingfield, S. J. Stahl, N. Cheng, B. L. Trus, A. C. Steven, and R. W. Williams (1998), "Three-dimensional structure of HIV-1 Rev protein filaments." *J. Struct. Biol.*, 121, 41–52.

Watts, N. R., J. G. Vethanayagam, R. B. Ferns, R. S. Tedder, A. Harris, S. J. Stahl, A. C. Steven, and P. T. Wingfield (2010), "Molecular basis for the high degree of antigenic cross-reactivity between hepatitis B virus capsids (HBcAg) and dimeric capsid-related protein (HBeAg): insights into the enigmatic nature of the e-antigen." *J. Mol. Biol.*, 398, 530–541.

Wingfield, P. T., S. J. Stahl, M. A. Payton, S. Venkatesan, M. Misra, and A. C. Steven (1991), "HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties." *Biochemistry*, 30, 7527–7534.

Wingfield, P. T., S. J. Stahl, R. W. Williams, and A. C. Steven (1995), "Hepatitis core antigen produced in Escherichia coli: subunit composition, conformational analysis, and in vitro capsid assembly." *Biochemistry*, 34, 4919–4932.

Winter, G. (2010), "xia2: an expert system for macromolecular crystallography data reduction." *J. Appl. Cryst.*, 43, 186–190.

World Health Organization (2000), *Hepatitis B: World Health Organization Fact Sheet 204*. World Health Organization.

Wright, P. E. and H. J. Dyson (1999), "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." *J. Mol. Biol.*, 293, 321–331.

Wynne, S. A., R. A. Crowther, and A. G. Leslie (1999), "The crystal structure of the human hepatitis B virus capsid." *Mol. Cell*, 3, 771–780.

Yang, C. Y., T. H. Kuo, and L. P. Ting (2006), "Human hepatitis B viral e antigen interacts with cellular interleukin-1 receptor accessory protein and triggers interleukin-1 response." *J. Biol. Chem.*, 281, 34525–34536.

Yang, Z. R., R. Thomson, P. McNeil, and R. M. Esnouf (2005), "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins." *Bioinformatics*, 21, 3369–3376.

Ye, J. D., V. Tereshko, J. K. Frederiksen, A. Koide, F. A. Fellouse, S. S. Sidhu, S. Koide, A. A. Kossiakoff, and J. A. Piccirilli (2008), "Synthetic antibodies for specific recognition and crystallization of structured RNA." *Proc. Natl. Acad. Sci. U.S.A.*, 105, 82–87.

Yedavalli, V. S., C. Neuveut, Y. H. Chi, L. Kleiman, and K. T. Jeang (2004), "Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function." *Cell*, 119, 381–392.

Zapp, M. L., T. J. Hope, T. G. Parslow, and M. R. Green (1991), "Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif." *Proc. Natl. Acad. Sci. U.S.A.*, 88, 7734–7738.

Zemmel, R. W., A. C. Kelley, J. Karn, and P. J. Butler (1996), "Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element." *J. Mol. Biol.*, 258, 763–777.

Zhou, Y., J. H. Morais-Cabral, A. Kaufman, and R. MacKinnon (2001), "Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution." *Nature*, 414, 43–48.

Publications

Stahl, S.J., Watts, N.R., Rader, C., **DiMattia, M.A.**, Mage, R.G., Palmer, I., Kaufman, J.D., Grimes, J.M., Stuart, D.I., Steven, A.C., Wingfield, P.T. (2010). Generation and characterization of a chimeric rabbit/human fab for co-crystallization of HIV-1 Rev. J Mol Biol. 397(3): 697-708.

DiMattia, M.A., Watts, N.R., Stahl, S.J., Rader, C., Wingfield, P.T., Stuart, D.I., Steven, A.C., Grimes, J.M. (2010). Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. Proc Natl Acad Sci USA. 107(13):5810-4.

DiMattia, M.A., Watts, N.R., Stahl, S.J., Grimes, J.M., Steven, A.C., Stuart, D.I., Wingfield, P.T. (2013). Antigenic Switching of Hepatitis B Virus by Alternative Dimerization of the Capsid Protein. Structure. 21(1): 133-142.