Structural and Biophysical Studies of RNA-dependent RNA polymerases

Sam Wright

Nuffield Department of Clinical Medicine
Pembroke College
Oxford University

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2010

Structural and Biophysical Studies of RNA-dependent RNA polymerases

Sam Wright, Pembroke College, University of Oxford
D.Phil, Nuffield Department of Clinical Medicine
Trinity Term 2010

RNA-dependent RNA polymerases (RdRps) play a vital role in the life cycle of RNA viruses, being responsible for genome replication and mRNA transcription. In this thesis viral RdRps (vRdRps) of dsRNA bacteriophage ϕ6 (ϕ6 RdRp) and Severe Acute Respiratory Syndrome (SARS) coronavirus [non structural protein 12 (NSP-12)] are studied.

For SARS polymerase NSP-12, a library-based screening method known as ESPRIT (Expression of Soluble Protein by Random Incremental Truncation) was employed in an attempt to isolate domains of NSP-12 that express solubly in *Escherichia coli* (*E. coli*) and are thereby suitable for structural studies. This experiment identified for the first time in a systematic fashion, conditions under which the SARS polymerase could be solubly expressed at small scale and allowed mapping of domain boundaries. Further experiments explored different approaches for increasing expression levels of tractable fragments at large scale.

Bacteriophage ϕ6 RdRp is one of the best studied vRdRps. It initiates RNA synthesis using a *de novo* mechanism without the need for a primer. Although formation of the *de novo* initiation complex has been well studied, little is known about the mechanism for the transition from initiation to elongation (i.e. extension of an initiated dinucleotide daughter strand). In the ϕ6 RdRp initiation complex the C-terminal domain (CTD) blocks the exit path of the newly synthesised dsRNA which must be displaced for the addition of the third nucleotide. The crystal structure of a C-terminally truncated ϕ6 RdRp ($P2^{T1}$) reveals the strong non-covalent interactions between the CTD and the main body of the polymerase that must be overcome for the elongation reaction to proceed. Comparing new crystal structures of complexes of both wild-type (WT) and a mutant RdRp (E634 to Q, which removes a salt-bridge between the CTD and main body of the polymerase) with various oligonucleotides (linear and hairpin), nucleoside triphosphates (NTPs) and divalent cations, alongside their biophysical and biochemical properties, provides an insight into the precise molecular details of the transition reaction.

Thermal denaturation experiments reveal that Mn^{2+} acquired from the cell and bound at the ϕ6 RdRp non-catalytic ion site sufficiently weakens the polymerase structure to facilitate the displacement of the CTD. Our crystallographic and biochemical data also indicate that Mn^{2+} is released during this displacement and must be replaced for the elongation to proceed. Our data explain the role of the non-catalytic divalent cation in vRdRps and pinpoint the Mn^{2+}-dependent step in viral replication. In addition, by inserting a dysfunctional Mg^{2+} at the non-catalytic ion site for both WT and E634Q RdRps we captured structures with two NTPs bound within the active site in the absence of Watson-Crick base pairing with template and could map movements of divalent cations during preinitiation through to initiation. Oligonucleotides present on the surface of ϕ6 RdRp allowed mapping of key residues involved in template entry and unwinding of dsRNA; these preinitiation stages have not been observed previously. Considering the high structural homology of ϕ6 RdRp with other vRdRps, particularly from (+)ssRNA hepatitis C virus (HCV), insights into the mechanistic and structural details of ϕ6 RdRp are thought to be relevant to the general understanding of vRdRps.

I would like to dedicate this thesis to my parents

Mark and Antonietta Wright

Acknowledgements

Firstly I would like to thank Dr. Jonathan Grimes for giving me the opportunity to carry out this exciting piece of research under his supervision. Thank you for your friendship, guidance and support. A special thanks to Prof. David Stuart for his intellectual rigour and useful discussions throughout my studies. Thank you to Prof. Dennis Bamford and Dr. Darren Hart for great collaborations. I thank specifically Dr. Minna Poranen and Peter Sarin from the Bamford Laboratory at the University of Helsinki and Philippe Mas from the Hart Laboratory at the EMBL in Grenoble for all their input and insightful discussions.

Many people at the Division of Structural Biology and Oxford Protein Production Facility have contributed to this work. Thank you to all my friends in the department who have inspired me throughout my studies and for creating a wonderful environment to work in. Special thanks to Dr. Joanne Nettleship, Dr. Louise Bird, Nahid Rahman, Dr. Sarah Sainsbury, Dr. Erika Mancini and Margaret Jones for their friendship and support. This work would have taken much longer and would have been much harder without your friendship and guidance around the lab. I am also specifically grateful to Dr. Stephen Graham, Dr. Karl Harlos and Jun Dong for their help with data collection, crystal handling and computing.

I would like to say a special thank you to my friends Dr. Maria Carroll, Emily Martinez-Perez, Nicola Blackford, Jane Orton, Cat Ball and especially Stephen Beebee for putting up with me during the hard times and for being there for me when I needed them the most. A huge thank you to both my parents and my grandfather for their love and support; I wouldn't have been able to get through this without you. Finally, I would like to say a special thank you to my husband, Mark McBride-Wright. I love you with all my heart.

Declaration of Work

The work described herein is my own except as described below or where stated in the text.

Chapter 2: Protein expression and purification of WT RdRp, E634Q and E491Q were jointly carried out by myself in Oxford and by Minna Poranen (senior scientist) as part of a collaborative project with Prof. Dennis Bamford at the University of Helsinki, Finland. Serine protease digestion of WT ϕ6 RdRp to form a C-terminally truncated protein ($P2^{T1}$) for structural analysis was carried out by Peter Sarin (graduate student) at the University of Helsinki as part of this collaboration.

Chapter 3: In order to confirm the inferred role of Mn^{2+} in the transition to elongation during ϕ6 RdRp-catalysed polymerisation from our structural analyses, a series of supporting biochemical experiments were carried out by Minna Poranen (senior scientist) at the University of Helsinki. **Fig. 3.22** summarises the results of these studies which are included in a publication alongside crystal structures of WT and E634Q complexes and thermal denaturation experiments reported herein (Wright *et al*., 2010; submitted).

Chapter 4: The ESPRIT library screening experiments and preliminary small scale expression studies were carried out by myself under the guidance of Philippe Mas (biochemistry expression technician) as part of a collaborative project with Dr. Darren Hart at the European Molecular Biology Laboratory (EMBL) in Grenoble, France. Liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS) experiments on NSP-12 protein truncations were carried out by Dr. Joanne Nettleship using the quality assurance service provided by the Oxford Protein Production Facility (OPPF) for the University of Oxford (Nettleship *et al*., 2008).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

αα	Amino acid
ϕ6	Bacteriophage ϕ6
ACT	2′-amino-2′-deoxycytidine-5′-triphosphate
ADRP	Adenosine diphosphate-ribose-1-phosphatase domain (SARS-CoV)
AMPPCP	β,γ-methyleneadenosine 5'-triphosphate
ASU	Asymmetric unit
AMP	Adenosine monophosphate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BAP	Biotin acceptor peptide
BTV	Blue Tongue Virus
BVDV	Bovine Viral Diarrhoea Virus
CTP	Cytidine triphosphate
CV	Coxsackievirus
GMP	Guanosine monophosphate
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
DdDP	DNA-dependent DNA polymerase
DdRP	DNA-dependent RNA polymerase
DENV	Dengue Virus
D1 / D2	NTP that base-pairs with T1 / T2 nucleotide of ϕ6 RNA
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECL	Enzyme-linked chemiluminescence
EDTA	Ethylenediamine tetraacetic acid
EMBL	European Molecular Biology Laborartory
ESRF	European Synchrotron Radiation Facility
ExoIII	Exonuclease III
FMDV	Foot and Mouth Disease Virus
FUMP	5-fluorouridine monophosphate
FUTP	5-fluorouridine triphosphate
GFL	Growth factor-like domain (SARS-CoV)
GST	Glutathione-*S*-transferase
GnHCl	Guanidine hydrochloride
HCV	Hepatitis C Virus
HEL1	Helicase protein (SARS-CoV)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
His_6	Hexa-histidine tag
HIV	Human Immunodeficiency Virus
HRP	Horse radish peroxidase
HRV	Human Rhinovirus
HTP	High-throughput
IBDV	Infectious Bursal Disease Virus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kav	Gel phase distribution coefficient

LB	Luria-Bertani media
ESI-MS	Electrospray ionisation-mass spectrometry
MES	Morpholinoethanesulfonic acid
MHV	Murine Hepatitis Virus
Mk	Marker
N7-met GMP	N7-methyl guanosine monophosphate
NCS	Non-crystallographic symmetry
NCT	5-nitrocytidine triphosphate
Ni-NTA	Nickel- nitrilotriacetic acid
NSP	Non-structual protein
sNts	Nucleotides
NTP	Nucleoside triphosphate
NV	Norwalk Virus
OD	Optical density
OPPF	Oxford Protein Production Facility
ORF(s)	Open reading frame(s)
PBS(-T)	Phosphate buffered saline (-with 0.05% Tween-20)
PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulphonylfluoride (serine protease inhibitor)
Pol	Polymerase
PPi	Pyrophosphate
PV	Poliovirus
RMSD	Root mean-square deviation
RNA	Ribonucleic acid
RdDP	RNA-dependent DNA polymerase
RdRP	RNA-dependent RNA polymerase
RHDV	Rabbit Haemorrhagic Disease Virus
RT	Reverse transcriptase
RTP	Ribavirin triphosphate
RV	Reovirus
SARS-CoV	Severe Acute Respiratroy Syndrome Coronoavirus
SDS(-PAGE)	Sodium dodecyl sulphate (-polyacryalmide gel electrophoresis)
SOC	Super optimal broth
SUMO	Small Ubiquitin-like Modifier
SV	Sapovirus
T1 / T2	3'-end / penultimate nucleotide of ϕ6 minus strand RNA
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phoshine hydrochloride
TEV	Tobacco Etch Virus
T_m	Melting temperature
TAE	Tris-Acetate-EDTA buffer
Tris(-HCl)	Tris-hydroxymethyl aminomethane (-hydrochloride) buffer
UMP	Uridine monophosphate
UTP	Uridine triphosphate
Vo / Ve / Vt	Void / Elution / Total bed volume
WNV	West Nile Virus
WT	Wild-type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Chapter 1

Introduction

1.1. Polymerases

Polymerisation of nucleoside triphosphates (NTPs) is vital for biological processes such as: transcription, primer synthesis during deoxyribonucleic acid (DNA) replication, addition of polyadenylate tails to messenger ribonucleic acid (RNA), uridylation in RNA editing, viral RNA replication, and many others. Enzymes that catalyse this polymerisation reaction are classified into two major categories: template dependent and template independent nucleotidyl transferases or polymerases. All template independent RNA transferases share the same fold of the catalytic domain [e.g. oligoA-synthetases, uridyl transferases and CCA-adding enzymes, reviewed in Torralba *et al*., (2008)]. A detailed description of this group of nucleotidyl transferases is beyond this scope of this thesis, since all studied polymerases function in a template dependent manner.

The group of template dependent polymerases refers to enzymes that use nucleic acid template to produce a new molecule, and are classified according to template/product preferences:

(i) DNA-dependent DNA polymerases (DdDps)

(ii) DNA-dependent RNA polymerases (DdRps)

(iii) RNA-dependent DNA polymerases (RdDps), also known as reverse transcriptases (RTs)

(iv) RNA-dependent RNA polymerases (RdRps)

Various authors have used sequence alignment protocols to identify conserved sequence motifs within the four different classes of template dependent polymerase (Delarue *et al*., 1990; Hansen *et al*., 1997; Koonin, 1991; Masters *et al*., 1987; Poch *et al*., 1989; Xiong and Eickbush, 1990). Poch *et al*. (1989) demonstrated the presence of four conserved sequence motifs (motifs A, B, C and D) in single subunit RNA-dependent polymerases. The alignment constructed by Delarue *et al*. (1990) went on to demonstrate that two of these motifs, A and C, are conserved in all single subunit polymerases and provide two strictly conserved aspartate residues, one from motif A (Asp_A) and the other from motif C (Asp_C). Mutagenesis studies indicated that these aspartates are critical for polymerisation [reviewed by O'Reilly and Kao (1998)]. A modified alignment of sequence motifs from all the classes of polymerase, presented by Hansen *et al.* (1997), is shown in **Fig. 1.1.** As structural data has emerged it has become apparent that the structural similarity of these motifs in single subunit polymerases is much more extensive than sequence similarity. Motifs A-D are depicted on representative structures of the four classes of polymerase in **Fig. 1.2.** Furthermore, sequence comparisons of several RNA-dependent polymerases identified a fifth motif (motif E) present only in this subgroup of polymerases (see **Fig. 1.1** and **Fig. 1.2**).

Figure 1.1. Sequence alignment of representative members of the four classes of template dependent polymerases.

Structure-based sequence alignments of poliovirus polymerase (RdRp), HIV-1 RT (RdDP or RT), the Klenow fragment of DNA polymerase I (DdDP), and bacteriophage T7 RNA polymerase (DdRP). In addition to motifs A-D, present in all four classes, motif E that is present only in RNA-dependent polymerases, is also shown. The strictly conserved aspartates in motifs A and C (Asp_A and Asp_C) are highlighted by a red box. The α-helical and β-strand regions of the motifs are indicated on the bottom line. In the Klenow fragment, motif D contains an insert with the sequence TRLDVP which is designated by an asterisk in the alignment. Structural agreements are boxed with a solid line and doubled line areas indicate differences in structure in the different polymerase categories; italic text indicates regions that were not compared due to disorder in the crystals. [Figure modified from (Hansen *et al.*, 1997)].

1.1.1. Structural classification of polymerases

Despite sequence and structural differences, template dependent polymerases determined to date can be grouped into one of three apparently unrelated folds:

(i) Right hand fold

(ii) Polβ fold

(iii) Multisubunit DdRps fold

1.1.1.1. Right hand fold

The Klenow fragment (KF) of DNA-dependent DNA polymerase I from *Escherichia coli* (*E. coli*) was the first polymerase structure to be determined (Ollis *et al*., 1985). It has three different domains: "palm", "fingers" and "thumb" (see **Fig. 1.2, A**) oriented like that of a "right hand". As the first structure of a reverse transcriptase [human immunodeficiency virus type 1 (HIV-1) polymerase], DNA-dependent RNA polymerase [bacteriophage T7 polymerase (T7pol)] and RNA-dependent RNA polymerase [poliovirus polymerase (PV 3Dpol)] emerged, it was clear that they too shared the same organisation of domains and were therefore grouped as "right hand fold" polymerases (see **Fig. 1.2, B, C and D** respectively).

The palm domain, comprising two long α-helices packed against a large β-sheet, is the most well conserved structural feature between polymerases and holds the strictly conserved aspartates (Asp_A and Asp_C) postulated to be essential for a "two metal ion" catalysis (see **Fig. 1.3**), that was later confirmed by structural determination of the first complexes of polymerases, such as T7pol (Doublie *et al*., 1998). In this mechanism, two divalent metal ions accompany the incoming NTP and associate with the phosphates of the nucleotide and carboxylate moieties of the conserved aspartates that anchor the ions of the reaction (Joyce and Steitz, 1995; Steitz, 1998; Steitz and Steitz, 1993). The enzyme catalyses the sequential

addition of (d)NTP monomer (via its α-phosphate group) to the 3'-hydroxyl (OH) group of the daughter chain, establishing a new bond and releasing the β and γ-phosphate groups as a molecule of inorganic pyrophosphate (PPi). Metal ion A (**Fig. 1.3**), lowers the affinity of the 3'-hydroxyl for the hydrogen atom, facilitating the nucleophilic attack on the α-phosphate. Metal ion B (**Fig. 1.3**), coordinates the β and γ-phosphate groups, assisting their release as PPi. Importantly, both ions stabilise a pentacovalent state of the α-phosphate group, essential for the catalysis.

Conversely, the fingers and thumb domains of polymerases across different classes show very little structural similarity. Considering that both domains are important for template and substrate binding (Joyce and Steitz, 1995; O'Reilly and Kao, 1998), differences in these domains likely reflects variation in template preference.

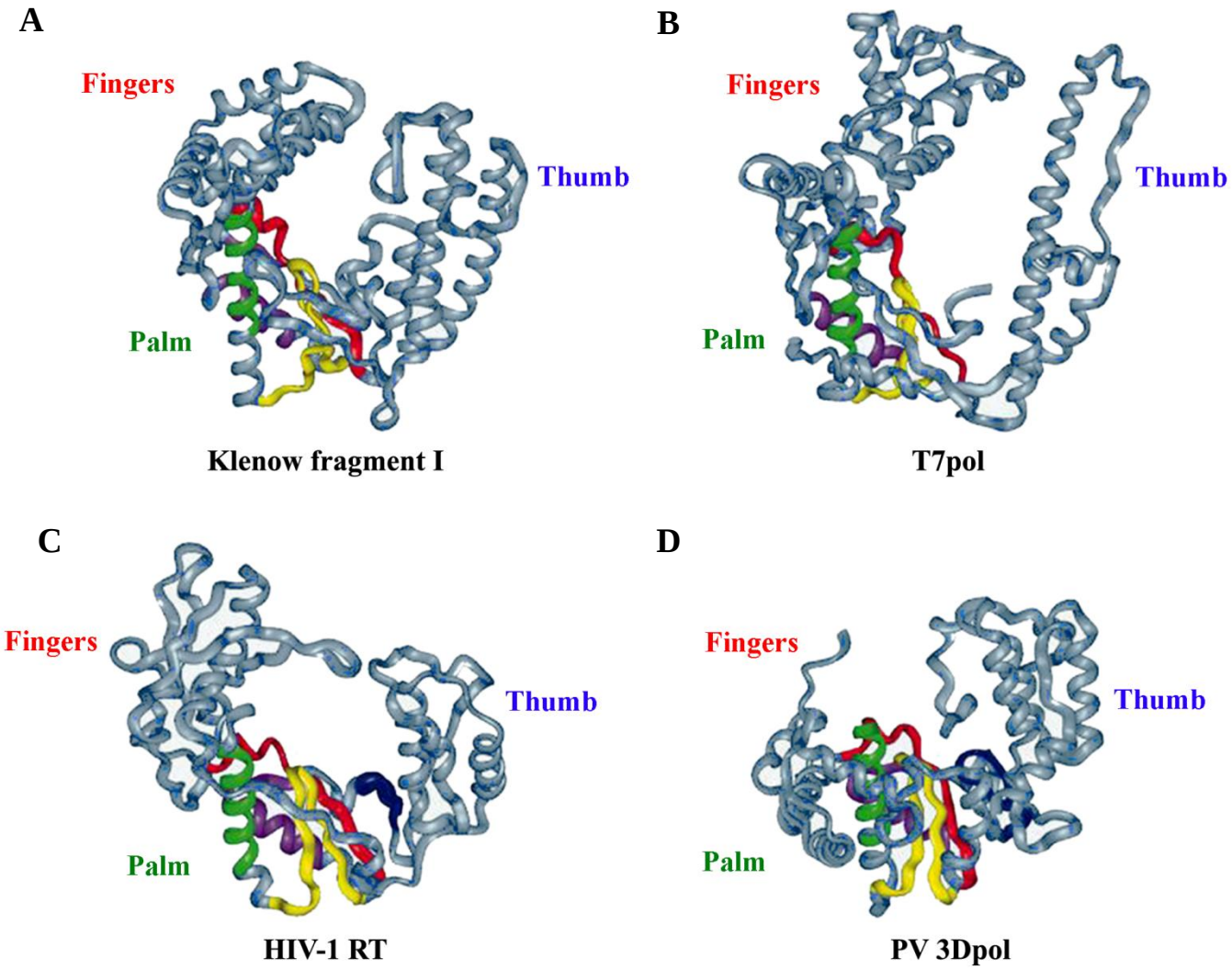
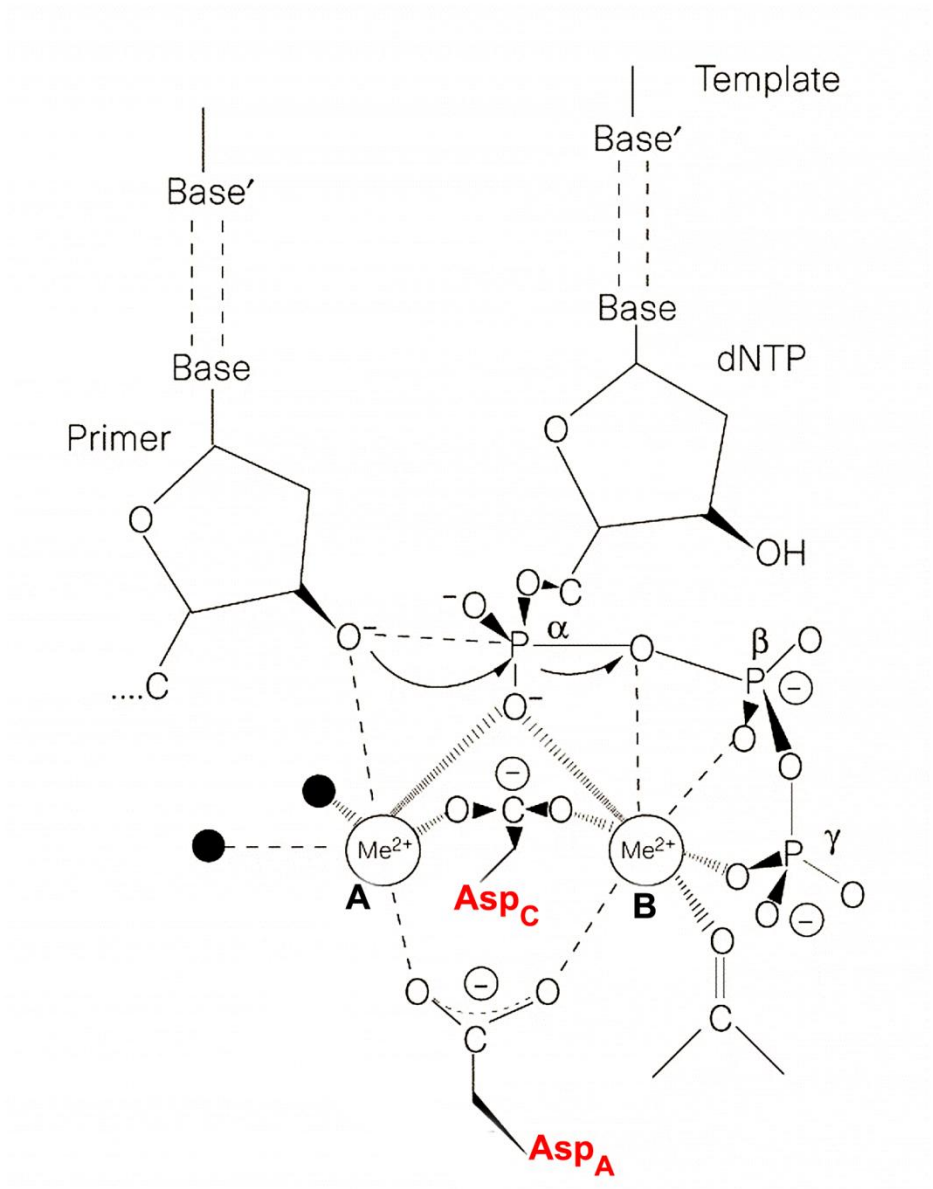


Figure 1.2. Polymerase structure/sequence motifs in representative structures from each of the four categories of polymerases.

Each is positioned with the thumb domain to the right (blue text) and the fingers domain to the left (red text). The structure/sequence motifsof the palm domain (green text) are colour coded: A in red, B in green, C in yellow, D in purple, and E in dark blue. Structure of the polymerase domains of **(A)** DNA-dependent DNA polymerase: large (Klenow) fragment of *E. coli* DNA polymerase I (Ollis *et al*., 1985); **(B)** DNA-dependent RNA polymerase: T7 RNA polymerase (T7pol), (Sousa *et al*., 1993); **(C)** RNA-dependent DNA polymerase: HIV-1 reverse transcriptase (HIV-1 RT) (Kohlstaedt *et al*., 1992); **(D)** RNA-dependent RNA polymerase: poliovirus polymerase (PVpol), (Hansen *et al*., 1997). [Figure modified from (Hansen *et al*., 1997)].

Figure 1.3. Schematic representation of the two metal ion catalysis mechanism.

Polymerase strictly conserved aspartate residues (Asp_A and Asp_C, labelled and coloured red) co-ordinate the two divalent metals ions (Me^{2+}, A and B) essential for catalysis. Water molecules (black circles) are bound to metal ion A. [Figure modified from (Steitz, 1998)].



1.1.1.2. Polβ fold

DNA-dependent DNA polymerase β (Polβ), involved in mammalian base excision DNA repair, belongs to a different class of polymerases, non-homologous to the right hand fold enzymes both in terms of sequence and structure. Although Polβ possesses structural domains equivalent to the aforementioned palm, fingers and thumb [as determined by Davies *et al*., (1994) and Pelletier *et al*., (1994), see **Fig. 1.4, A**], the connectivity of these domains is very different. The fingers and thumb for example have an opposite connectivity to that observed in the right hand fold enzymes, and the palm represents a completely new domain class (**Fig. 1.4, B**). This considered, a similar active site geometry and two metal ion catalysis was reported from structures of Polβ in complex with dideoxy-cytosine triphosphate (ddCTP) and DNA, including the presence of carboxylates coordinating metal ions about the 3'-hydroxyl of the primer strand (Sawaya *et al*., 1994).

It would appear that non-homologous polymerases (Polβ and right hand) with quite distinct palm domain structures can generate an appropriate surface to present catalytic ions in a suitable geometrical arrangement relative to the NTP α-phosphate and primer 3'-OH for efficient catalysis. This suggests that both Polβ and right hand polymerases have achieved this mechanism by convergent evolution from unrelated ancestors.

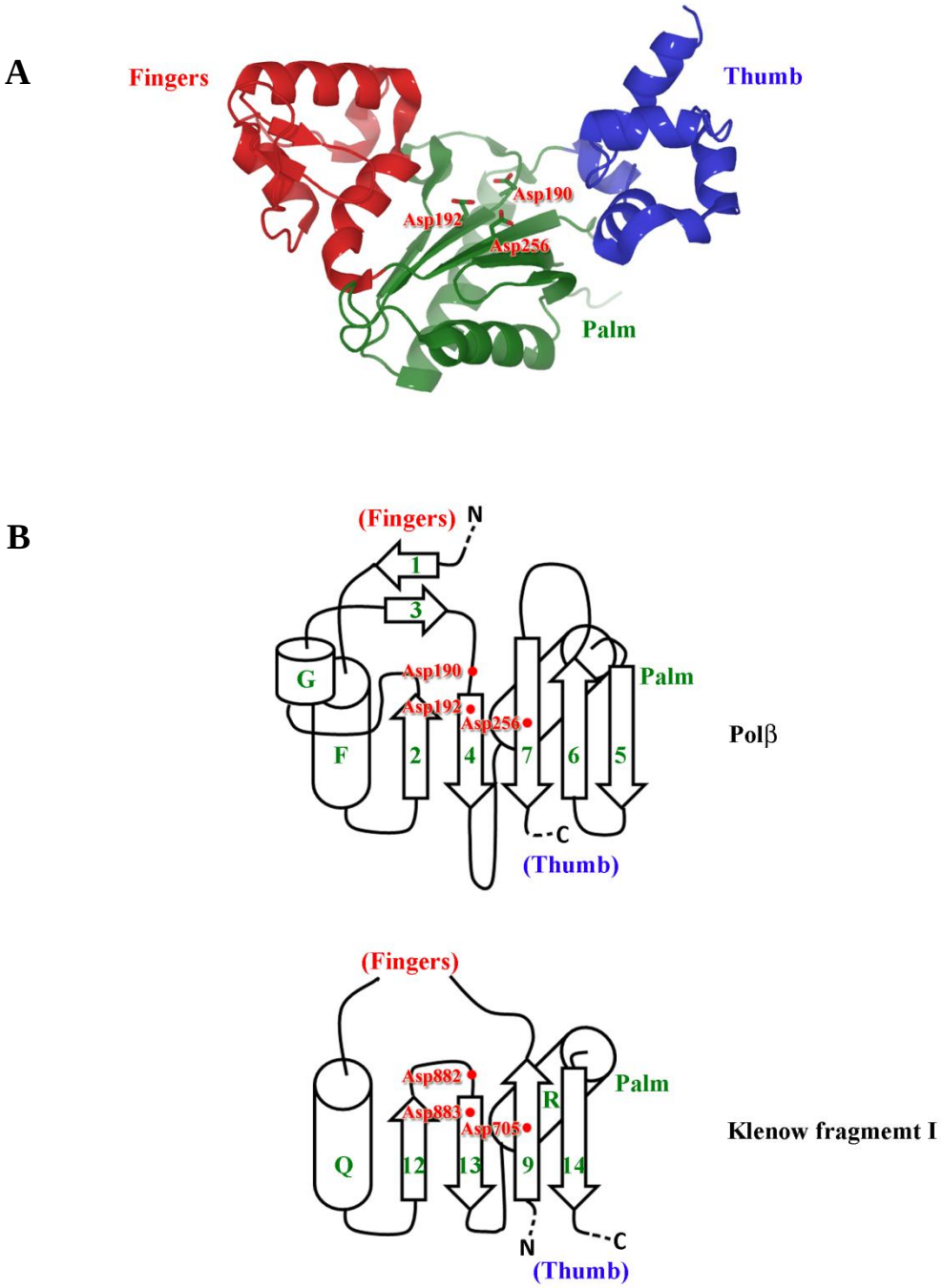


Figure 1.4. Polymerase β (Polβ) fold.

(A) Structure of the rat DNA polymerase β (Davies *et al*., 1994). The divergent equivalents of right hand fold domains are coloured green (palm), blue (thumb) and red (fingers). Only the 31 kDa catalytic domain is shown with residues D190, D192 and D256 indicated by sticks (carbon, green; oxygen, red). **(B)** Schematic representation of the palm domain of Polβ (top) and DNA polymerase I Klenow fragment (bottom) as an example of the right-hand fold. The α-helices (lettered cylinders) and β-strands (numbered arrows) of the central β-sheet are shown. The aspartates at the active site are indicated with a red sphere and labelled. The labels indicate the positions of the fingers and thumb domains (marked by parentheses) with respect to the palm.

1.1.1.3. Multisubunit DdRp fold

DNA-dependent RNA polymerases (DdRp) are large protein complexes composed of many subunits (5 to 17) and are involved in gene transcription. The transcription machinery of eukaryotes is much more complex than that of prokaryotes and archaea, although the general principles of transcription and how it is regulated are conserved. Eukaryotes possess three nuclear enzymes (RNA polymerases I-III), that share five common subunits and are involved in synthesizing different classes of RNA, whereas bacteria and archaea only encode one DdRp. RNA polymerases I, II and III (Pol I-III) comprise 14, 12 and 17 subunits respectively. It has been shown that the ten subunits that make up the core of Pol II are homologous to all cellular DdRps, suggesting that all these enzymes have the same basic structure and mechanism despite differences in subunit number (Cramer *et al*., 2008; Ebright, 2000). A detailed description of the subunits and complex regulation of DdRp transcription is reviewed in Cramer *et al*., (2008) but is out of the scope of this thesis. Instead, focus is given to the organisation of the active site cleft and interactions of elements in the core formed by two large subunits with template and substrate molecules.

The two largest subunits of the multimeric complex form the catalytically active core of DdRps: β' and β in bacteria, RNA binding protein (Rbp) 1 and Rbp2 in yeast, representing the largest and second largest subunits respectively. Unlike right-hand and Polβ folds that have a catalytic palm domain, no corresponding single domain is present in multisubunit DdRps. Instead, interaction of the two largest subunits creates a positively charged nucleic acid binding cleft. The largest subunit (β'/Rbp1) forms a mobile 'clamp' on one side of the cleft, involved in the stabilisation of the DNA-RNA hybrid [reviewed by Cramer *et al*., (2001), see **Fig. 1.5, A**]. The other side is defined by two domains of the β/Rbp2 subunit that form a 'lobe' and 'protrusion', also stabilising the active complex (**Fig. 1.5, A**). This subunit

also provides a 'bridge helix' (β residues 1067-1081 and 1083-1093, Rbp2 residues 810-846) that spans the cleft before the active site (**Fig. 1.5, A and B**) and provides a platform for nucleic acid–protein interactions (Cramer *et al*., 2001; Gnatt *et al*., 2001; Westover *et al*., 2004). The bridge helix lines an opening in the cleft that widens to the exterior creating an inverted funnel (funnel region, **Fig. 1.5, A**) where substrate nucleoside triphosphates enter. Beyond the active site, the cleft is blocked by a 'wall' or 'flap' (**Fig. 1.5, A**) which prevents straight passage of nucleic acids through the DdRp. This wall region forces a 90^{o} bend in the DNA-RNA hybrid with respect to the entering DNA, exposing the end of the DNA-RNA hybrid for the addition of NTPs. Two double-psi topology β-barrels, DPBB (one from each subunit) define the active site (**Fig. 1.5, B**). The β'/Rbp1 subunit contributes three invariant aspartates in a DxDxD motif [e.g. D739, D741, D743 in *Thermus aquaticus* (Zhang *et al*., 1999) and D728, D730, D732 in *Thermus thermophilus* (Vassylyev *et al*., 2002); D481, D483, D485 in yeast (Cramer *et al*., 2001)] that coordinates a Mg^{2+} ion (metal A). The second DPBB (from the β/Rbp2 subunit) contributes positively charged residues that complete the cleft (Artsimovitch *et al*., 2004; Cramer *et al*., 2001; Gnatt *et al*., 2001; Vassylyev *et al*., 2002; Westover *et al*., 2004). Catalysis is thought to follow the two metal ion mechanism identified in single subunit polymerases. In multisubunit polymerases one metal ion (metal A) is intrinsically bound with the second metal ion (metal B) accompanying the NTPs upon entry (Cramer *et al*., 2001; Vassylyev *et al*., 2002; Zhang *et al*., 1999).

There is no sequence or structural conservation between multisubunit and single subunit (right hand and Polβ) polymerases, however all establish an optimal platform for two metal ion catalysis with the correct geometric arrangement of carboxylates from conserved aspartates in the active site. This is a nice example of convergent evolution where structurally unrelated folds have evolved the same optimum mechanism for efficient catalysis.

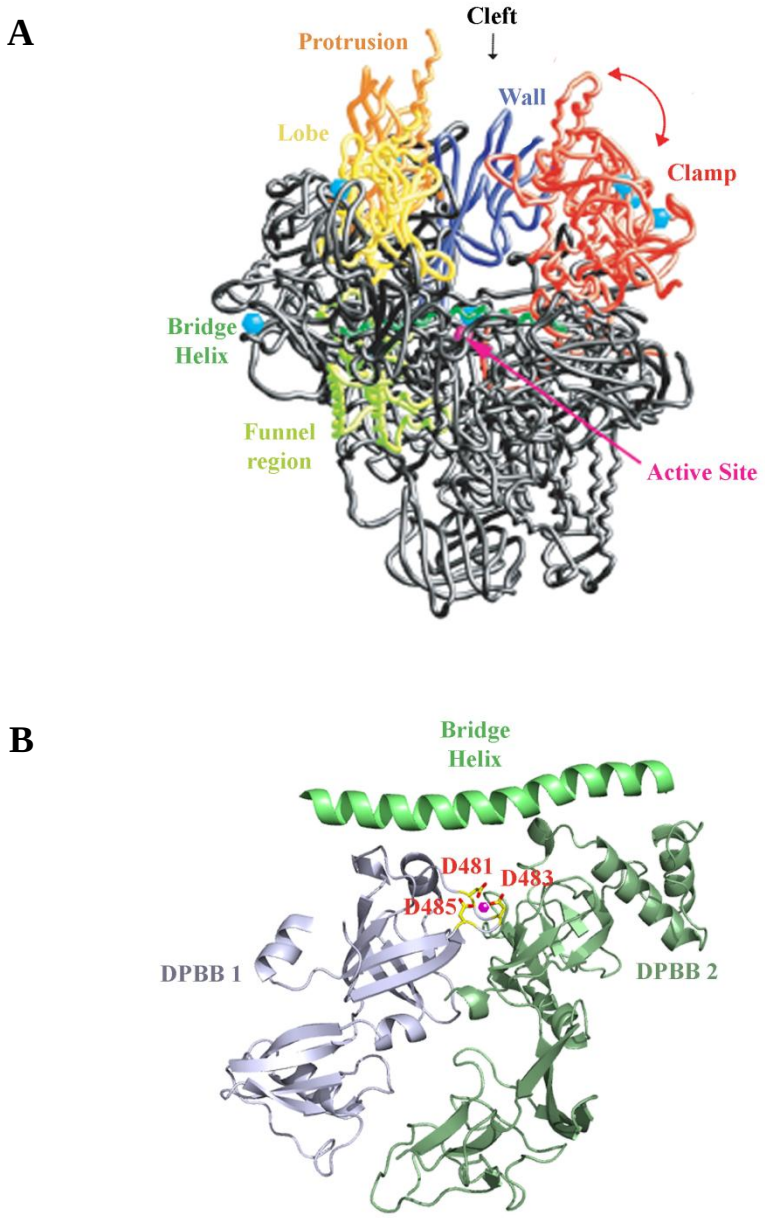


Figure 1.5. Structure of multisubunit DdRp RNA pol II from yeast.

(A) Important functional and structural domains (including the active site, magenta) are highlighted. The Rbp1/Rbp2 domains are coloured as follows: clamp (red), bridge (green), funnel (light green), protrusion (orange), lobe (yellow) and wall (blue). Zinc ions are indicated by light blue spheres. [Figure adapted from (Cramer, 2002)]. **(B)** Close-up of the active site cleft in yeast DdRp. Bridge helix (green) and DPBBs from the Rbp1 subunit (DPBB1, grey) and Rbp2 subunit (DPBB2, dark green) are shown. Catalytic aspartates from the DxDxD motif are depicted by sticks (carbon, yellow; oxygen, red) and labelled accordingly. Bound Mn^{2+} (soaked into the crystal) is depicted by a magenta sphere in this structure (PDB accession no. 1I50) where it has replaced the bound Mg^{2+} ion (Cramer *et al*., 2000).

1.1.2. Viral RNA-dependent RNA polymerases

Virus encoded RNA-dependent RNA polymerases (vRdRps) are vital components in the life cycle of RNA viruses, being responsible for genome replication and mRNA transcription when associated with other proteins of both viral and possibly cellular origin, in a complex generally termed the "polymerase complex" (PC) (Lai, 1998) or replication complex.

To date a number of vRdRps have been structurally determined, some in complex with template, NTP and/or metal ions [see **Table 1.1.**]. They exhibit structural conservation of the canonical right hand motif with palm, fingers and thumb domains, but also possess an extra structural feature linking the fingers and thumb domains, known as the "fingertips", which gives vRdRps a more spherical or 'closed hand' appearance. Conversely, DNA-dependent and RT polymerases have a more 'open hand' appearance. Two positively charged tunnels are created by the vRdRp closed conformation: the template and NTP binding channels [see for example, Bressanelli *et al*., (2002), Butcher *et al*., (2001) and Ferrer-Orta *et al*., (2004)].

In vRdRps the most striking differences are at the fingertips and thumb domains, whereas the palm domain, containing the catalytic aspartates, and fingers domain (a mixed α-helical–β-sheet topology) are well conserved. For example, bacteriophage ϕ6 RdRp (Butcher *et al*., 2001) has an elaborate six-strand fingertip domain, whereas hepatitis C virus polymerase (HCVpol) NS5B (Bressanelli *et al*., 1999) and Norwalk virus polymerase (NVpol) (Ng *et al*., 2004) have only four strands, and the bovine viral diarrhea virus polymerase (BVDVpol) (Choi *et al*., 2004) has an even less elaborate three strands. Differences in the thumb domain are thought to reflect differences in mode of initiation. In addition, ϕ6 RdRp (Butcher *et al*., 2001), HCVpol (Bressanelli *et al*., 1999), BVDVpol (Choi *et al*., 2004) and NVpol (Ng *et al*.,

2004) have an extra C-terminal domain (CTD) that is crucial to the initiation mode adopted by the polymerase.

Structural homologies between vRdRps extend beyond viral classes with striking similarities between dsRNA polymerases, such as ϕ6 RdRp (Butcher *et al*., 2001), and (+)ssRNA virus polymerases, such as HCVpol (Bressanelli *et al*., 1999) and BVDVpol (Choi *et al*., 2004). In fact, these (+)ssRNA *Flaviviridae* polymerases have a higher structural homology with dsRNA *Cystoviridae* ϕ6 RdRp than with other (+)ssRNA polymerases, such as picornavirus polymerases. Picornaviral RdRps structurally determined to date (**Table 1.1**), do not have the additional CTD structural feature exhibited by ϕ6 RdRp, HCVpol and BVDVpol. Instead, they have a much wider template tunnel in order to accommodate a much bulkier template and exploit a primer-dependent initiation mechanism. Conversely, for polymerases with this additional CTD, initiation is primer-independent (*de novo*) and the additional structural feature stabilises the initiation competent complex in the active site (discussed in section 1.1.3.4.). This considered, similarity and conservation in palm and thumb domains at the secondary and tertiary level of the families *Picorna-*, *Flavi-* and *Cystoviridae* suggests that they have evolved from a common ancestor (Butcher *et al*., 2001; O'Farrell *et al*., 2003).

Polymerase	Virus type	Virus Family	PDB code	Resl. (Å)	Structural Information	Reference
PV 3Dpol	(+)ssRNA	*Picornaviridae*	1RDR	2.4	1-12; 38-66; 98-181; 270-290 disordered	(Hansen *et al*., 1997)
PV 3Dpol (mutant)	(+)ssRNA	*Picornaviridae*	1RA6	2.0	L446D/R455D	(Thompson and Peersen, 2004)
PV 3Dpol (mutant) GTP/Mg^{2+} complex	(+)ssRNA	*Picornaviridae*	1RA7	2.0	L446D/R455D + GTP	(Thompson and Peersen, 2004)
PV 3Dpol (mutant) UTP/Mg^{2+} complex	(+)ssRNA	*Picornaviridae*	2IM2	2.4	L446D/R455D + UTP	(Thompson *et al*., 2007)
PV 3Dpol mutant CTP/Mg^{2+} complex	(+)ssRNA	*Picornaviridae*	2IM0	2.3	L446D/R455D + CTP	(Thompson *et al*., 2007)
PV 3Dpol mutant ATP/Mg^{2+} complex	(+)ssRNA	*Picornaviridae*	2ILY	2.6	L446D/R455D + ATP	(Thompson *et al*., 2007)
PV 3Dpol mutant GTP/Mn^{2+} complex	(+)ssRNA	*Picornaviridae*	2ILZ	2.5	L446D/R455D + GTP + Mn^{2+} (2)	(Thompson *et al*., 2007)
PV 3Dpol mutant UTP/Mn^{2+} complex	(+)ssRNA	*Picornaviridae*	2IM3	2.6	L446D/R455D + UTP + Mn^{2+} (2)	(Thompson *et al*., 2007)
PV 3Dpol mutant CTP/Mn^{2+} complex	(+)ssRNA	*Picornaviridae*	2IM1	2.5	L446D/R455D + CTP + Mn^{2+} (2)	(Thompson *et al*., 2007)
HRV1B 3Dpol Serotype 1B	(+)ssRNA	*Picornaviridae*	1XR6	2.3	Serotype 1B + K^{+}	(Love *et al*., 2004)
HRV14 3Dpol Serotype 14	(+)ssRNA	*Picornaviridae*	1XR5	2.3	Serotype 14 + Sm^{+}	(Love *et al*., 2004)
HRV16 3Dpol Serotype 16	(+)ssRNA	*Picornaviridae*	1XR7	2.3	Serotype 16	(Love *et al*., 2004)
HRV16 3Dpol Serotype 16	(+)ssRNA	*Picornaviridae*	1TP7	2.4	Full length Serotype 16	(Appleby *et al*., 2005)
FMDV 3Dpol	(+)ssRNA	*Picornaviridae*	1U09	1.9	Full length	(Ferrer-Orta *et al*., 2004)
FMDV 3Dpol RNA complex	(+)ssRNA	*Picornaviridae*	1WNE	3.0	template:primer RNA	(Ferrer-Orta *et al*., 2004)
FMDV 3Dpol VPg complex	(+)ssRNA	*Picornaviridae*	2D7S	3.0	+ VPg	(Ferrer-Orta *et al*., 2006)
FMDV 3Dpol VPg/UTP complex	(+)ssRNA	*Picornaviridae*	2F8E	2.9	+ VPg + UMP + Mg^{2+} + Mn^{2+}	(Ferrer-Orta *et al*., 2006)
FMDV 3Dpol RNA/ATP complex	(+)ssRNA	*Picornaviridae*	2ECO	2.7	template:primer RNA[a] + AMP + PPi	(Ferrer-Orta *et al*., 2007)
FMDV 3Dpol RNA/ATP/UTP complex	(+)ssRNA	*Picornaviridae*	2E9Z	3.0	template:primer RNA[a] + ATP + UTP	(Ferrer-Orta *et al*., 2007)
FMDV 3Dpol RNA/FUTP complex	(+)ssRNA	*Picornaviridae*	2E9T	2.6	template:primer RNA[a] + FUMP + PPi	(Ferrer-Orta *et al*., 2007)

Polymerase	Virus type	Virus Family	PDB code	Resl. (Å)	Structural Information	Reference
FMDV 3Dpol RNA/RTP complex	(+)ssRNA	*Picornaviridae*	2E9R	2.8	template:primer RNA[a] + RTP	(Ferrer-Orta *et al*., 2007)
CV 3Dpol Serotype B3	(+)ssRNA	*Picornaviridae*	3DDK	2.2	Full length	(Campagnola *et al*., 2008)
CV 3Dpol Serotype B3 GTP(PPi) complex	(+)ssRNA	*Picornaviridae*	3CDU	2.1	+ PPi	(Gruez *et al*., 2008)
CV 3Dpol Serotype B3 3′dUTP/VPg complex	(+)ssRNA	*Picornaviridae*	3CDW	2.5	+ PPi + VPg	(Gruez *et al*., 2008)
HCVpol [b] Genotype 1B BK strain	(+)ssRNA	*Flaviviridae*	1QUV	2.5	21 residues deleted at C-term.	(Ago *et al*., 1999)
HCVpol 1B BK strain	(+)ssRNA	*Flaviviridae*	1C2P	1.9	21 residues deleted at C-term.	(Lesburg *et al*., 1999)
HCVpol 1B BK strain	(+)ssRNA	*Flaviviridae*	1CSJ	2.8	55 residues deleted at C-term.	(Bressanelli *et al*., 1999)
HCVpol 1B BK strain GTP complex	(+)ssRNA	*Flaviviridae*	1GX5	1.7	55 residues deleted at C-term. + GTP + Mn^{2+}	(Bressanelli *et al*., 2002)
HCVpol 1B BK strain UTP complex	(+)ssRNA	*Flaviviridae*	1GX6	2.5	55 residues deleted at C-term. + UTP + Mn^{2+}	(Bressanelli *et al*., 2002)
HCVpol 1B J4 strain	(+)ssRNA	*Flaviviridae*	1NB4	2.0	21 residues deleted at C-term.	(O'Farrell *et al*., 2003)
HCVpol 1B J4 strain UTP complex	(+)ssRNA	*Flaviviridae*	1NB6	2.6	21 residues deleted at C-term. + UTP	(O'Farrell *et al*., 2003)
HCVpol 1B J4 strain RNA complex	(+)ssRNA	*Flaviviridae*	1NB7	2.9	21 residues deleted at C-term. 5'-UUUU-3' RNA	(O'Farrell *et al*., 2003)
HCVpol Genotype 2a BK strain	(+)ssRNA	*Flaviviridae*	1YUY	1.9	21 residues deleted at C-term. Crystal form I	(Biswal *et al*., 2005)
HCVpol 2a BK strain	(+)ssRNA	*Flaviviridae*	1YV2	2.5	21 residues deleted at C-term. Crystal form II	(Biswal *et al*., 2005)
BVDVpol construct 1	(+)ssRNA	*Flaviviridae*	1S48	3.0	Construct I: 71-679 71-91 residues disordered	(Choi *et al*., 2004)
BVDVpol construct 1 GTP complex	(+)ssRNA	*Flaviviridae*	1S49	3.0	Construct 1: 71-679 71-91 residues disordered + GTP	(Choi *et al*., 2004)
BVDVpol construct 2	(+)ssRNA	*Flaviviridae*	1S4F	3.0	Construct 2: 79-678 71-91; 675-678 disordered	(Choi *et al*., 2004)

Polymerase	Virus type	Virus Family	PDB code	Resl. (Å)	Structural Information	Reference
DENVpol Serotype 3 Mg^{2+} complex	(+)ssRNA	*Flaviviridae*	2J7U	1.8	Catalytic domain (273-900) + Zn^{2+} (2); + Mg^{2+}	(Yap *et al*., 2007)
DENVpol Serotype 3 3'dGTP complex	(+)ssRNA	*Flaviviridae*	2J7W	2.6	Catalytic domain (273-900) + Zn^{2+} (2); + 3'dGTP	(Yap *et al*., 2007)
WNVpol	(+)ssRNA	*Flaviviridae*	2HFZ	3.0	Longer construct (273-905) + Zn^{2+} + Mg^{2+}	(Malet *et al*., 2007)
WNVpol	(+)ssRNA	*Flaviviridae*	2HCS	2.5	Shorter construct (317-905) + Zn^{2+}	(Malet *et al*., 2007)
WNVpol Ca^{2+} complex	(+)ssRNA	*Flaviviridae*	2HCN	2.3	Shorter construct (317-905) + Zn^{2+} + Ca^{2+}	(Malet *et al*., 2007)
RHDVpol	(+)ssRNA	*Caliciviridae*	1KHV	2.5	1-4; 181-184; 502-516 disordered + Lu^{3+}	(Ng *et al*., 2002)
RHDVpol	(+)ssRNA	*Caliciviridae*	1KHW	2.7	1-4; 181-184; 502-516 disordered + Mn^{2+}	(Ng *et al*., 2002)
NVpol crystal form I	(+)ssRNA	*Caliciviridae*	1SH0	2.2	*P*1 1-5; 508-510 disordered	(Ng *et al*., 2004)
NVpol crystal form II	(+)ssRNA	*Caliciviridae*	1SH2	2.3	$C222_1$ 1-5; 508-510 disordered	(Ng *et al*., 2004)
NVpol crystal form III	(+)ssRNA	*Caliciviridae*	1SH3	3.0	$P2_12_12_1$ 1-5; 508-510 disordered + Mg^{2+}	(Ng *et al*., 2004)
NVpol RNA/CTP complex crystal form III	(+)ssRNA	*Caliciviridae*	3BSO	1.7	$P2_12_12_1$ 1-4; 467-471; 489-510 disordered 5'-UGCCCGGG-3' + CTP + Mn^{2+} (3)	(Zamyatkin *et al*., 2008)
NVpol RNA/NCT complex crystal form III	(+)ssRNA	*Caliciviridae*	3BSN	1.8	$P2_12_12_1$ 1-4; 467-471; 489-510 disordered 5'-UGCCCGGG-3' + NCT + Mn^{2+} (3)	(Zamyatkin *et al*., 2008)
NVpol RNA/ACT complex crystal form III	(+)ssRNA	*Caliciviridae*	3H5X	1.8	$P2_12_12_1$ 1-4; 467-471; 489-510 disordered 5'-UGCCCGGG-3' + ACT + Mn^{2+} (4)	(Zamyatkin *et al*., 2009)
SVpol	(+)ssRNA	*Caliciviridae*	2CKW	2.3	Full length	(Fullerton *et al*., 2007)
ϕ6 RdRp	dsRNA	*Cystoviridae*	1HI8	2.5	Full length (SeMet $P3_2$) + Mg^{2+}	(Butcher *et al*., 2001)
ϕ6 RdRp	dsRNA	*Cystoviridae*	1HHS	2.0	Full length (Native $P2_1$) + Mn^{2+}	(Butcher *et al*., 2001)

Polymerase	Virus type	Virus Family	PDB code	Resl. (Å)	Structural Information	Reference
ϕ6 RdRp ATP complex	dsRNA	*Cystoviridae*	1H11	3.0	+ ATP; + Mn^{2+}	(Butcher *et al*., 2001)
ϕ6 RdRp DNA complex	dsRNA	*Cystoviridae*	1HHT	2.5	5'-TTTCC-3' DNA + Mn^{2+}	(Butcher *et al*., 2001)
ϕ6 RdRp Initiation complex	dsRNA	*Cystoviridae*	1HI0	2.5	5'-TTTCC-3' DNA + GTP (2); + Mg^{2+} (2); + Mn^{2+}	(Butcher *et al*., 2001)
ϕ6 RdRp Ca^{2+} inhibited complex	dsRNA	*Cystoviridae*	1UVN	3.0	5'-UUUUCC-3' RNA + GTP (2); + Ca^{2+} (2); + Mn^{2+}	(Salgado *et al*., 2004)
ϕ6 RdRp Dead end complex	dsRNA	*Cystoviridae*	1UVK	2.5	5'-UUUCC-3' RNA + PPP-G-P-G; + PPi; + Mg^{2+} (2); + Mn^{2+}	(Salgado *et al*., 2004)
ϕ6 RdRp 7nt RNA complex	dsRNA	*Cystoviridae*	1UVJ	2.2	5'-UUUUUCC-3' RNA + Mn^{2+}	(Salgado *et al*., 2004)
ϕ6 RdRp 6nt RNA complex	dsRNA	*Cystoviridae*	1UVI	1.9	5'-UUUUCC-3' RNA + Mn^{2+}	(Salgado *et al*., 2004)
ϕ6 RdRp 5nt RNA complex (conf. A)	dsRNA	*Cystoviridae*	1UVM	2.0	buried conformation A 5'-UUUCC-3' RNA + Mn^{2+}	(Salgado *et al*., 2004)
ϕ6 RdRp 5nt RNA complex (conf. B)	dsRNA	*Cystoviridae*	1UVL	2.0	intermediate conformation B 5'-UUUCC-3' RNA + Mn^{2+}	(Salgado *et al*., 2004)
ϕ6 RdRp Back-priming SG mutant	dsRNA	*Cystoviridae*	1WAC	3.0	SG mutant Q629S/Y630G/K631T/W632E	(Laurila *et al*., 2005)
Reovirus λ3pol type 3	dsRNA	*Reoviridae*	1MUK	2.5	Full length (apo) 1; 957-964; 1266-1267 disordered	(Tao *et al*., 2002)
Reovirus λ3pol serotype 3 mRNA cap analogue	dsRNA	*Reoviridae*	1MWH	2.5	1; 957-964; 1266-1267 disordered + mRNA cap analogue GPPG; + Mn^{2+}	(Tao *et al*., 2002)
Reovirus λ3pol serotype 3 Initiation complex	dsRNA	*Reoviridae*	1N1H	2.5	1; 957-964; 1266-1267 disordered + 5'-AUUAGC-3' RNA + 3'dCTP; + 3'dGTP; + GDP; + N7-met GMP; + Mn^{2+}	(Tao *et al*., 2002)
Reovirus λ3pol serotype 3 Elong. complex I	dsRNA	*Reoviridae*	1N38	2.5	1; 957-964; 1266-1267 disordered 5'-GC-3' 3'-AUUGC-5' + 3'dUTP; + 3'dCTP; + Mn^{2+} (2)	(Tao *et al*., 2002)

Polymerase	Virus type	Virus Family	PDB code	Resl. (Å)	Structural Information	Reference
Reovirus λ3pol serotype 3 Elong. complex II	dsRNA	*Reoviridae*	1N35	2.8	1; 957-964; 1266-1267 disordered 5'-GGGG-3' 3'-AUUGCCCCC-5' + 3'dCTP; + Mn^{2+} (2)	(Tao *et al*., 2002)
IBDV pol (VP1)	dsRNA	*Birnaviridae*	2PGG	2.5	1-30; 603-611; 805-878 disordered	(Pan *et al*., 2007)
IBDV pol (VP1)	dsRNA	*Birnaviridae*	2PUS	2.4	1-26 disordered	(Garriga *et al*., 2007)
IBDV pol (VP1) Mg^{2+} complex	dsRNA	*Birnaviridae*	2R72	3.1	1-26 disordered Mg^{2+} (3)	(Garriga *et al*., 2007)
IBDV pol (VP1) regulatory VP3 peptide complex	dsRNA	*Birnaviridae*	2R70	2.7	1-26 disordered VP3 peptide: GRLGRWIRTVSDEDLE	(Garriga *et al*., 2007)

Table 1.1. Structural models of vRdRps determined to date by X-ray crystallography.

Viruses: PV – Poliovirus; HRV – Human Rhinovirus; FMDV – Foot and Mouth Disease Virus; CV – Coxsackievirus; HCV – Hepatitis C Virus; BVDV – Bovine Viral Diarrhoea Virus; RHDV – Rabbit Haemorrhagic Disease Virus; NV – Norwalk Virus (or more recently Norovirus); ϕ6 – Bacteriophage ϕ6; IBDV – Infectious Bursal Disease Virus; WNV – West Nile Virus; SV – Sapovirus; DENV – Dengue Virus. **Molecules:** N7-metGMP – N7-methyl guanosine monophosphate; NCT – 5-nitrocytidine triphosphate; ACT – 2′-amino-2′-deoxycytidine-5′-triphosphate; FUTP – 5-fluorouridine triphosphate; FUMP – 5-fluorouridine monophosphate; RTP – ribavirin triphosphate. [a] – FMDV template-primer sequence: 5′-GCAUGGGCCC-3′ and 5′-CGUAGGGCCC-3′ respectively. [b] – PDB accession codes for recent HCVpol crystal structures in complex with clinically relevant inhibitors (not included in the table for clarity): 3G86, 3H59, 3H5S, 3H5U, 2WCX, 2WHO, 3FQK, 3FQL, 3CIZ, 3CJ0, 3CJ2, 3CJ3, 3CJ4, 3CJ5, 2GIQ, 2GIR, 2JC0, 2JC1 and 2FVC.

1.1.3. Bacteriophage ϕ6

Bacteriophage ϕ6 infects the plant pathogenic bacterium *Pseudomonas syringae* and is the prototype virus of the *Cystoviridae* family. The virion is made up of three concentric layers (see **Fig. 1.6.**):

(i) Core or polymerase complex (PC)

(ii) Nucleocapsid (NC)

(iii) Lipid envelope

The dsRNA genome of ϕ6 is internalised within the core particle and consists of three dsRNA segments: small (S), medium (M) and large (L) that are surrounded by the PC in the virion. Each virus particle contains one copy of each segment (Day and Mindich, 1980) with the following lengths respectively: 2948bp, 4063bp and 6374bp (Gottlieb *et al*., 1988; McGraw *et al*., 1986; Mindich, 1988). A conserved regulatory sequence is found at the 5' and 3' termini of each segment. Segment L codes for all core particle components, whereas the other genes are shared between the S and M segments.

The core or PC layer comprises 120 copies of major structural protein P1 (Ktistakis and Lang, 1987), 12 monomers of RNA-dependent RNA polymerase P2 (Makeyev and Grimes, 2004), 12 hexamers of RNA translocase P4 (Gottlieb *et al*., 1992; Juuti *et al*., 1998) and 30 dimers of assembly factor P7 (Juuti and Bamford, 1997). The 120 copies of P1 are arranged as 60 dimers on a T=1 icosahedral lattice (Butcher *et al*., 1997; de Haas *et al*., 1999). The organisation with two molecules of the same protein in two different conformations in an icosahedral asymmetric unit is also referred to as the T=2 structure (Grimes *et al*., 1998). Bluetongue virus VP3 layer as well as the rotavirus VP2 layer are organized similarly (Mertens *et al*., 2004). This arrangement has only been observed in the core particles of the dsRNA viruses.

The major structural protein P1 has been shown to play an important role in specific recognition of ϕ6 RNA packaging signals. In addition the RNA translocase P4, which sits at the 5-fold vertices, has been shown to unwind dsRNA by NTP hydrolysis and use this energy to drive packing of ssRNA genome precursors into the PC. A likely mechanism for this highly efficient unidirectional motor in RNA translocation has been proposed from the determination of several catalytically relevant structures of the P4 protein from related bacteriophage ϕ12 (Mancini *et al*., 2004). The positions of P2 (RNA-dependent RNA polymerase) and P7 (thought to modulate RNA metabolism and packing) within the PC are less certain, although it is thought that P2 might also sit at the 5-fold vertices (de Haas *et al*., 1999; Poranen *et al*., 2001).

The NC layer, which forms a shell around the core, is represented by P8 trimers that are arranged onto an incomplete T=13 lattice (Butcher *et al*., 1997; de Haas *et al*., 1999; Ktistakis and Lang, 1987). A similar arrangement has also been found in the intermediate layer of reoviruses (Grimes *et al*., 1998; Reinisch *et al*., 2000). In addition, an endopeptidase P5 is present at T=13 sites of the P8 layer. The outermost layer (envelope) comprises integral membrane proteins P6, P9, P10 and P13 embedded in the phospholipid bilayer. The cell attachment protein P3 is anchored to P6 of the lipid envelope forming the outer spikes of the virion. **Fig. 1.6** provides an illustration of the virion structure highlighting the of positions key proteins.

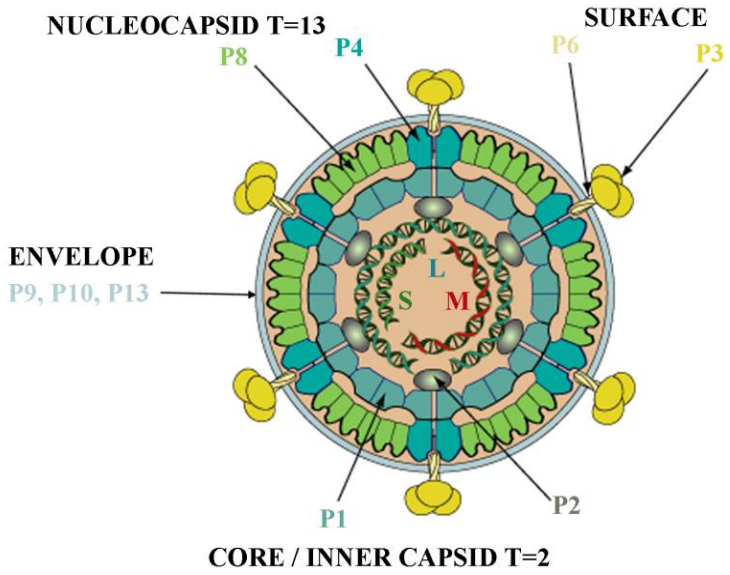


Figure 1.6. Bacteriophage ϕ6 virion structure.

Cross-section schematic of the virion showing key viral proteins. dsRNA genomic segments are shown in the core (small, S – green; medium, M – red; large, L - blue) with polymerase complex (PC) proteins P1 (shell constituent, light blue), P2 (RdRp, dark grey) and P4 (packaging NTPase, dark blue); P7 (modulator protein) is not shown. The nucleocapsid (NC) protein P8 (green), lipid envelope proteins P9, P10, P13 (integral membrane proteins and phospholipids, light grey) and outer spike proteins P3 (cell attachment protein) anchored to P6 (outer spikes) are also highlighted on the structure; P5 (endopeptidase) of the NC is not shown. [Figure adapted from ExPASy proteomics server database (Gasteiger *et al*., 2003)].

1.1.3.1. ϕ6 life cycle

The ϕ6 life cycle (**Fig. 1.7.**) has been reviewed previously (Mindich, 1988; Poranen and Tuma, 2004) and only a brief overview is given below. The outer spike protein P3 from bacteriophage ϕ6 attaches to the bacterial pilus of *Pseudomonas syringae* causing it to retract. P6 mediates the fusion of the virion with the outer membrane (OM). Once attached, the endopeptidase P5 degrades the peptidoglycan layer (PG) leading to the concomitant penetration of the plasma membrane (PM) by the viral nucleocapsid (NC). Interactions between P8 and the PM then mediate NC entry into the cytoplasm by membrane voltage dependent and independent interactions [**Fig. 1.7. (i)** and **(ii)**]. Once in the cytoplasm, active core particles (CPs) are formed by loss of the external P8 shell and surrounding membrane vesicles [**Fig. 1.7. (ii)**]. These particles transcribe the three genome segments S, M and L, extruding the (+) sense transcripts (s+, m+ and l+) into the cytoplasm [**Fig. 1.7. (iii)**]. The cellular transcription machinery is then exploited for viral protein synthesis. For replication, s+, m+ and l+ transcripts are packaged by RNA translocase P4 into newly synthesised and preformed PCs [**Fig. 1.7. (vi)** and **(vii)**]. The polymerase protein P2 (ϕ6 RdRp) can then synthesises the daughter strand of the (+)-sense transcripts, thereby reforming the dsRNA genome segments [**Fig. 1.7. (viii)**]. These PCs can then direct additional rounds of transcription or produce new NC leading to mature virion production via a complex morphogenesis pathway not detailed here [**Fig. 1.7. (ix)**]. At this stage, proteins P10 and P5 are involved in the disruption of the cell membrane leading to new infective particles being released by cell lysis [**Fig. 1.7. (x)**].

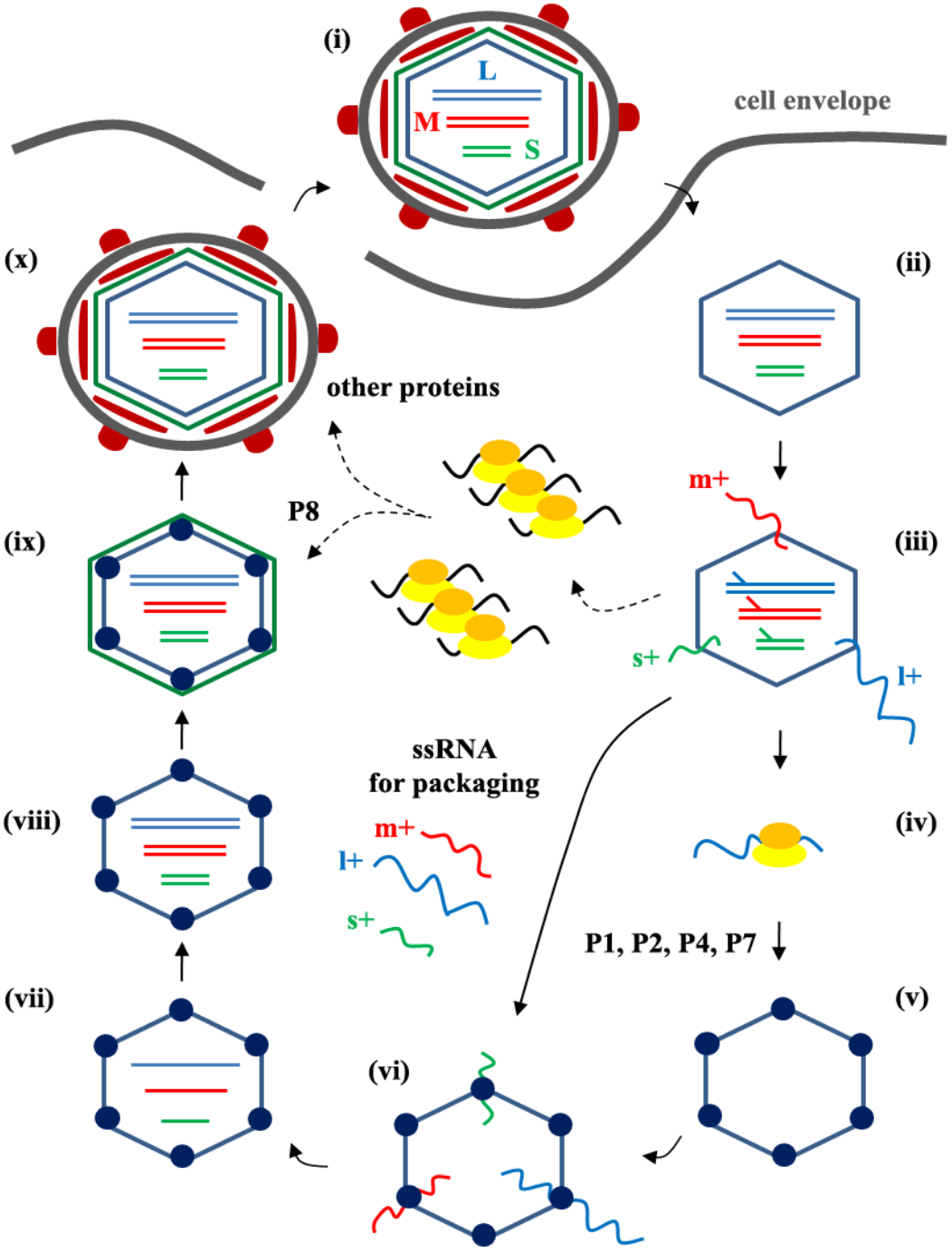


Figure 1.7. Bacteriophage ϕ6 life cycle.

(i) P3 interaction with bacterial pilus. P6 mediated fusion of envelope with OM and loss of envelope. P8 binding to PM and entry of NC. **(ii)** Loss of P8 shell (core particles). **(iii)** Transciption of dsRNA and extrusion of s+, m+ and l+. **(iv)** l+ translation into P1, P2, P4 and P7; s+ and m+ translation. **(v)** Assembly of new PC. **(vi)** Sequential packaging of (+)ssRNA segments. **(vii)** Formation of replicative active PC. **(viii)** Replication of s+, m+ and l+. **(ix)** P8 assembly to PC. **(x)** Lipid envelope, phospholipids and capsid proteins assembly. Release of new viral particle, accompanied by cell lysis. [Figure modified from Makeyev, (2001)].

1.1.3.2. ϕ6 RNA-dependent RNA polymerase

Bacteriophage ϕ6 encodes only one protein species carrying RdRp motifs, the P2 protein (referred herein as ϕ6 RdRp), that catalyses both transcription of genomic dsRNA into (+)-sense copies and the replication of the (+)ssRNA back into the dsRNA form. *In vitro* studies revealed differences between replication and transcription that may account for the regulation (Makeyev and Bamford, 2000a), however the molecular basis of both these processes is the same, involving the threading of ssRNA template into ϕ6 RdRp and synthesis of a complementary strand. Therefore, during transcription unwinding of dsRNA template is necessary for transcript synthesis.

A. Replication

Packaging of RNA molecules by ϕ6 is carried out sequentially: s+ is packaged first, followed by m+ and then l+. A mechanism was proposed by Qiao *et al*., (1997) based on *in vitro* PC preparations (Frilander and Bamford, 1995; Juuti and Bamford, 1995) and *in vivo* studies (Ewen and Revel, 1990). In brief, empty PCs only contain binding sites for s+ ssRNA. Particle expansion upon s+ entry forms m+ recognition sites. A similar change to reveal l+ recognition sites occurs upon m+ entry. Final expansion by packaged l+ triggers replication, presumably by RdRp stimulation. Interestingly, replication requires full packaging of only one copy of each of these three genome segments. This is also true of viruses of the *Reoviridae* family, whose genomes have 10 to 12 dsRNA genome segments. The mechanism of particle formation and RNA packaging in the family *Reoviridae* is still not very well understood. However, virus-encoded non-structural proteins have been shown to facilitate RNA packaging within cytoplasmic inclusions, known as viroplasms [see Patton and Spencer (2000) and references therein].

ϕ6 PCs carry out end-to-end polymerisation, where ssRNA and dsRNA templates are copied without any loss of genetic information. All genomic segments have a similar efficiency rate of initiation of replication, implying no specific dependence on signal sequences at the 3'-end of (+)ssRNA. Furthermore, *in vitro* isolated vRdRps from ϕ6 and other cystoviruses accept many heterologous ssRNA templates. Interestingly, the yield of dsRNA product is directly related to the template 3'-end secondary structure and terminal nucleotide. Studies by Makeyev and Bamford (2001) and Yang *et al*., (2001) showed that ϕ6, ϕ8 and ϕ13 polymerases prefer pyrimidine-rich 3'-terminal initiation sites (C-3' > U-3'). A similar template preference has been observed for HCVpol (Sun *et al*., 2000; Zhong *et al*., 2000) and BVDVpol (Kao *et al*., 1999). Further studies showed that ϕ6 RdRp requires a single-stranded 3'-terminus to efficiently initiate the polymerisation reaction (Laurila *et al*., 2002; Makeyev and Bamford, 2000b). An additional *in vitro* observation was the ability for ϕ6 RdRp to accept ssDNA as a template, although ssRNA was preferred (Makeyev and Bamford, 2001).

B. Transcription

Transcription is a semi-conservative reaction in bacteriophage ϕ6, with the old (+) strand being displaced by the newly synthesised (+) strand, using (-)ssRNA as the template. Displaced (+)ssRNA is extruded into the cytoplasm, where it is used as a mRNA for protein synthesis. Work carried out by Makeyev and Bamford (2000a) demonstrated that transcription activity *in vitro* is at least one order of magnitude less efficient than replication, presumably due to inefficient initiation from a dsRNA that will need to be unwound to reveal the 3'-end of the template (-)ssRNA. However, when assembled into the PC particle, ϕ6 RdRp can initiate multiple rounds of transcription from dsRNA and does not seem to require assistance of other proteins to unwind the duplex during elongation (Makeyev and Bamford,

2000a). Non-template strand displacement has also been shown for poliovirus and alfalfa mosaic virus polymerases (Cho *et al*., 1993; de Graaff *et al*., 1995), suggesting that this may be a common feature amongst RdRps.

ϕ6 RdRp transcribes genomic segments S and M more efficiently than the L segment. This is likely to be related to the preference for 3' terminal cytidines (described previously), given the following differences in (-)ssRNA 3' terminal sequences: 5'-...CC-3' for S and M segments, and 5'-...AC-3' for the L segment. The L segment encodes for all proteins of the polymerase complex and so differences in initiation efficiency likely reflect a control mechanism to switch from NC production to complete virion assembly and cell lysis.

C. ϕ6 RdRp stimulation and inhibition

The addition of calcium ions (Ca^{2+}) has been shown to inhibit transcription of dsRNA viruses such as bluetongue virus (BTV) and bacteriophage ϕ6 (Ojala and Bamford, 1995; van Dijk *et al*., 1995; Van Dijk and Huismans, 1980). In ϕ6 RdRp, transcription is inhibited at Ca^{2+} concentrations as low as 0.2 mM and replication activity lost at concentrations higher that 0.5 mM (Makeyev and Bamford, 2000a). The crystal structure of an inhibited initiation complex in the presence of Ca^{2+} (Salgado *et al*., 2004) demonstrated that the inhibition involved structural differences in the ligand coordination of the Ca^{2+} in comparison to magnesium (Mg^{2+}) at the active site. For Ca^{2+}, the average number of ligands is seven compared to the more rigid six ligand octagonal coordination of Mg^{2+} (Glusker *et al*., 1999; Saris *et al*., 2000). In the inhibited complex where Ca^{2+} has replaced the two Mg^{2+} ions, one of the bound Ca^{2+} ions is coordinated at a position 4.6 Å from the catalytically crucial Mg^{2+} position and causes a rotation of Y630 that provides the stabilising platform for the first incoming NTP [denoted D1 (Butcher *et al*., 2001), see section 1.1.3.3.]; Y630 is not involved in the ligand

coordination of Mg^{2+} in the catalytically active complex. Ca^{2+} at this new positions thereby causes a displacement of the incoming D1 NTP from its standard base pairing position with the 3'-end nucleotide of the template [denoted T1 (Butcher *et al*., 2001), see section 1.1.3.3.], which in turn alters the normal base pairing positioning of the second incoming NTP [denoted D2 (Butcher *et al*., 2001), see section 1.1.3.3.]. This new NTP positioning prevents the nucleophilic attack from O3' of D1 onto the O-P bond of D2 resulting in enzyme inhibition (Salgado *et al*., 2004).

Conversely, *de novo* and primer-independent RNA synthesis can be stimulated in ϕ6 RdRp by the addition of Mg^{2+} and manganese ions (Mn^{2+}) ions (Ranjith-Kumar *et al*., 2002b; Shim *et al*., 2002; Yi *et al*., 2003). The addition of Mn^{2+} ions (in Mg^{2+}-containing reaction buffer) increases the specific activity of ϕ6 RdRp by over an order of magnitude (Makeyev and Bamford, 2000b; Poranen *et al*., 2008b; Yang *et al*., 2001). Similar stimulatory effects have been described for a number of viral RdRps including polymerases from HCV (and other members of the family *Flaviviridae*), poliovirus (*Picornaviridae*) sapovirus (*Caliciviridae*), reovirus (*Reoviridae*), brome mosaic virus (*Bromoviridae*), and bacteriophages ϕ8, ϕ12 and ϕ13 (*Cystoviridae*) (Alaoui-Lsmaili *et al*., 2000; Arnold *et al*., 1999; Ferrari *et al*., 1999; Fullerton *et al*., 2007; Luo *et al*., 2000; Ranjith-Kumar *et al*., 2002a; Starnes and Joklik, 1993; Sun *et al*., 1996; Yang *et al*., 2001; Yang *et al*., 2003; Yi *et al*., 2003; Zhong *et al*., 2000). The mechanism of this stimulation is not well understood, although Mn^{2+} ions are known to decrease template specificity, lower the K_m for nucleotide binding, and modulate nucleotide incorporation in several polymerases (Palmenberg and Kaesberg, 1974; Ranjith-Kumar *et al*., 2002a; Tabor and Richardson, 1989; Wang and Korn, 1982). Mn^{2+} ions also support polymerization in the absence of Mg^{2+} ions (Luo *et al*., 2000; Poranen *et al*., 2008b) suggesting that Mn^{2+} can replace Mg^{2+} as the catalytic ion. Indeed initiation complexes with

Mn^{2+} ions coordinating triphosphate moieties of NTPs have been reported for RV and NV polymerases (Tao *et al*., 2002; Zamyatkin *et al*., 2008).

Previous structural studies on ϕ6 RdRp have identified a high affinity Mn^{2+} binding site in the palm domain at a position ~6 Å from the catalytic Mg^{2+} ions, co-ordinated by E491, A495 and one of the invariant catalytic aspartates D454 (Butcher *et al*., 2001). Interestingly, this non-catalytic ion binding site is conserved in a number of vRdRps [(Poranen *et al*., 2008b), **Table 1.2.**, **Fig. 1.8.** and discussed in detail in Chapter 3, section 3.3.9], suggesting that it may play an essential role in viral RNA polymerization. However, because of the low physiological concentration of Mn^{2+} [~10 μM compared to >10 mM Mg^{2+} (Finney and O'Halloran, 2003; Zhang and Ellis, 1989)], it remains uncertain as to whether it has a general physiological role in polymerisation.

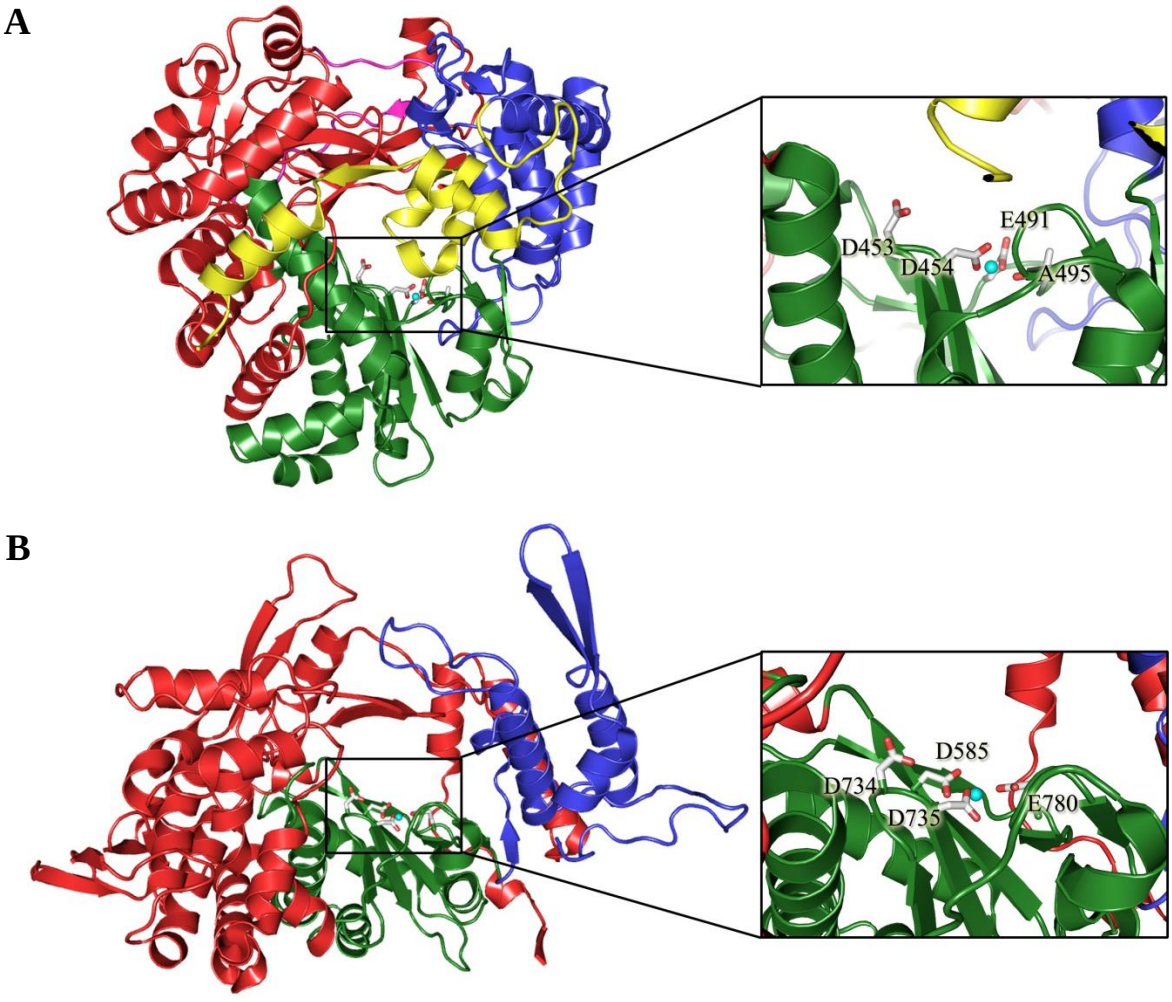
Table 1.2. Comparison of equivalent divalent cations bound in viral RNA polymerases to the high affinity non-catalytic Mn^{2+} binding site of ϕ6 RdRp.

vRdRp	PDB ID	Published ion at non-catalytic site
WNVpol	2HCN	Ca^{2+} [a]
	2HFZ	Mg^{2+}
FMDV 3Dpol	1WNE	Mg^{2+}
RHDVpol	1KHV	H_2O [b]
	1KHW	Density difficult to interpret
NVpol	1SH3	Mg^{2+}
	1BS0	H_2O
DENVpol	2J7U	Mg^{2+}
PV 3Dpol	1RDR	Ca^{2+} [a]
RVpol	1MWH	Mn^{2+}
IBDVpol	2R72	Mg^{2+}

[a] From map and coordination geometry it is unlikely to be Ca^{2+} (Poranen *et al*., 2008b).
[b] Evidence for Mg^{2+} in place of H_2O at equivalent site (Poranen *et al*., 2008b).
[Table adapted from (Poranen *et al*., 2008b)].

Figure 1.8. Comparison of equivalent non-catalytic Mn^{2+} binding sites in viral RNA polymerases from bacteriophage ϕ6 and reovirus.

(A) Structure of bacteriophage ϕ6 RdRp showing the non-catalyic Mn^{2+} binding site (PDB accession no. 1UVJ). *Left panel*: secondary structure elements of ϕ6 RdRp are shown with right-hand fold domains coloured green (palm), blue (thumb) and red (fingers). The connecting chains (fingertips) and C-terminal domain are coloured magenta and yellow respectively. *Right panel*: close-up of the non-catalytic Mn^{2+} binding site region. Residues D454, E491 and A495 that form the site are indicated by sticks (carbon, white; oxygen, red) and are labelled. The position of the second conserved asparate (D453) is also shown for completeness. Bound Mn^{2+} is indicated by a cyan sphere. **(B)** Polymerase region of reovirus λ3 polymerase (RV pol) showing an equivalent non-catalyic Mn^{2+} binding site (PDB accession no. 1MWH). *Left panel:* secondary structure elements of RV pol are shown with the conserved right-hand fold domains coloured green (palm), blue (thumb) and red (fingers). Only the polymerase region is depicted (the bracelet and N-terminal domain are not drawn for clarity). *Right panel*: close-up of the equivalent non-catalytic Mn^{2+} site region in RV pol. Residues D585, D735 and E780 that form the site are indicated by sticks (carbon, white; oxygen, red) and are labelled. The position of the second conserved asparate (D734) is also shown for completeness. Bound Mn^{2+} is indicated by a cyan sphere.



1.1.3.3. ϕ6 RdRp Structure

The structure of ϕ6 RdRp was determined in its apo form, in complex with NTP and template, and in an initiation state (Butcher *et al*., 2001). The structure reveals two positively charged tunnels allowing access of NTP substrates and RNA template to the active site (**Fig. 1.9.**). The 664 residues of the monomeric protein form 25 α-helices and 21 β-strands that fold into the right hand polymerase structure with fingers, thumb and palm domains with the additional fingertips and CTD domains that impose a more spherical architecture (**Fig. 1.9.** and **Fig. 1.10, A**).

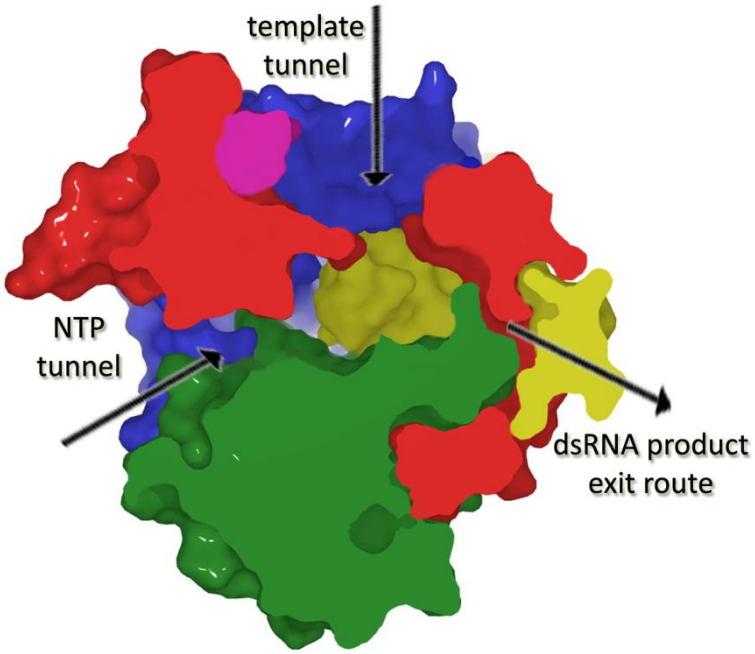


Figure 1.9. ϕ6 RdRp structure showing NTP and template tunnels.

Cross section of a surface representation of ϕ6 RdRp. Right hand fold domains are coloured as follows: palm (green), thumb (blue) and fingers (red), connecting chains / fingertips (magenta) and C-terminal domain (yellow). The positions of the NTP tunnel, template entry tunnel and dsRNA product exit route are indicated with a black arrow and labelled. Note the C-terminal domain blocks the exit route.

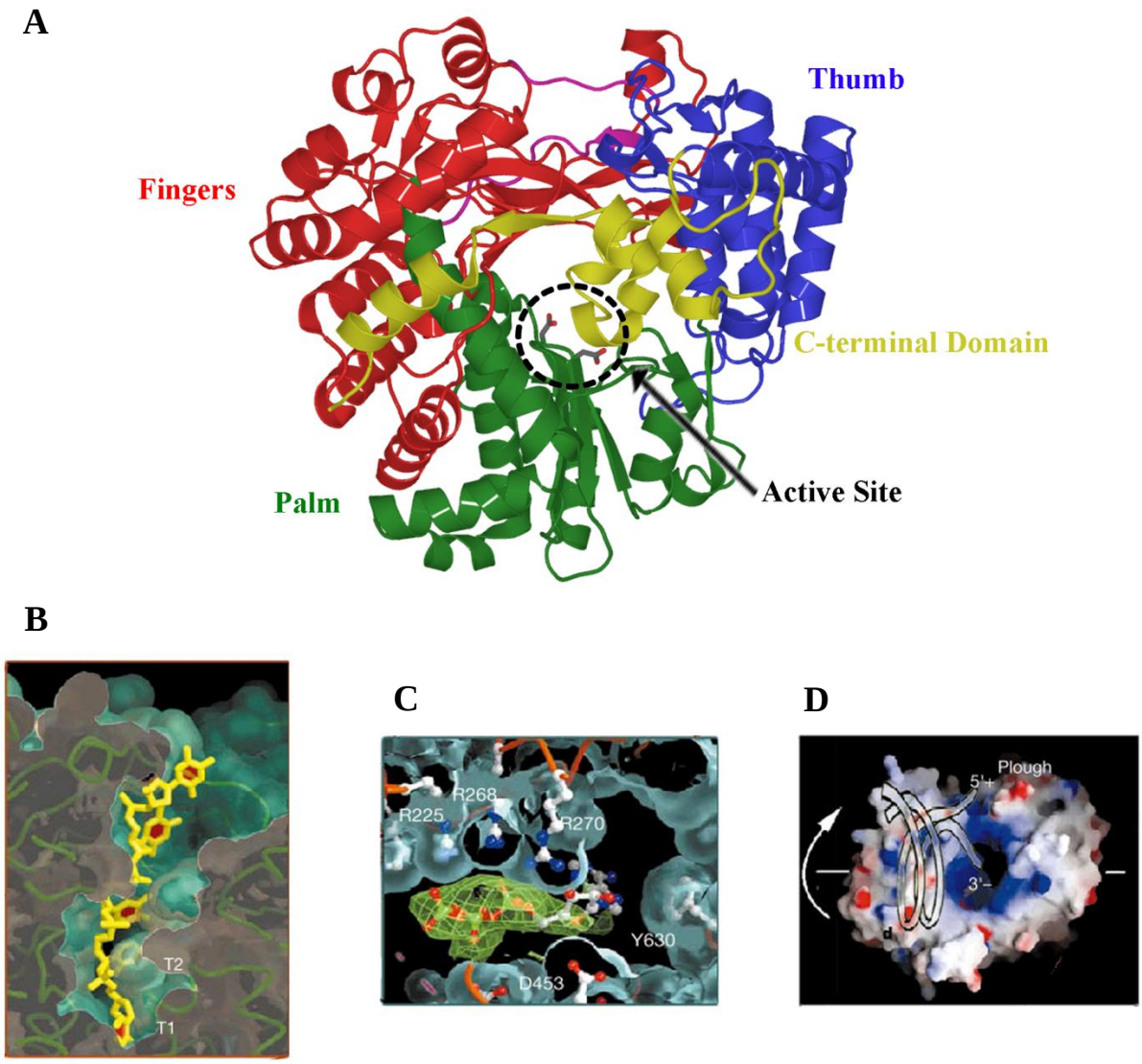


Figure 1.10. ϕ6 RdRp structure.

(A) Cartoon representation of ϕ6 RdRp structure. Right hand fold domains are coloured as follows: palm (green), thumb (blue) and fingers (red). The C-terminal domain is coloured yellow and connecting chains are coloured magenta. The active site is indicated by a dotted black circle and labelled; catalytic aspartates (D453 and D454) are depicted by sticks (carbon, white; oxygen, red).
(B) A section through the template channel with bound 5nt DNA oligonucleotide drawn in yellow. The surface of the polymerase and the embedded polypeptide chain are coloured green. T1 and T2 nucleotides are labelled.
(C) Difference electron density map for the NTP bound to site I. Conserved residues R225, R268 and R270 that 'interrogate' the phosphate backbone of the bound ATP molecule are shown. Catalytic aspartate D453 and priming platform residue Y630 are drawn. All residues and the ATP molecule are shown in ball-and-stick representation (carbon, white; nitrogen, blue; oxygen, red).
(D) Surface charge representation of ϕ6 RdRp viewed from above, showing the entrance to the template tunnel. Putative positions for the strands before initiation are shown.
[Modified from (Butcher *et al*., 2001)].

The NTP (substrate) tunnel is lined by the fingertips (or connecting chains) comprising six β-strands that strap together the fingers and thumb domains. The structure of ϕ6 RdRp in complex with an ATP molecule (Butcher *et al*., 2001) identified a number of interrogation residues (K223, R225, R268 and R270) shown to be important in co-ordinating the negatively charged triphosphate backbone of incoming NTPs. This interrogation site (I) is ~5 Å away from the catalytically relevant binding site for incoming NTPs (C) – nomenclature according to Butcher *et al*. (2001); see **Fig. 1.10, C.**

The template tunnel is only wide enough to accommodate ssRNA and not dsRNA. The distance from the active site to the surface can be spanned by 5 nucleotides (nts) of ssRNA, which has been observed in a complex of ϕ6 RdRp with a ssRNA oligonucleotide (Salgado *et al*., 2004) and also demonstrated with a 5nt ssDNA oligonucleotide [(Butcher *et al*., 2001), see **Fig. 1.10, B**]. Although it has been shown that ϕ6 RdRp can utilize both RNA and DNA templates, RNA templates are preferred (Makeyev and Bamford, 2000b). Complexes with RNA and DNA showed that sugar rings of RNA oligonucleotides assume a different conformation to those of DNA and make additional stabilizing interactions with template tunnel residues due to the presence of the extra hydroxyl group (O2' of ribose sugar ring, missing in DNA). From the published structures, the template oligonucleotides (T) have been numbered sequentially from 3' to 5', with the 3' end of the template used to initiate *de novo* RNA synthesis named T1. Nucleotides of the daughter strand (D) are numbered sequentially from 5' to 3': D1 base pairs with T1, D2 base pairs with T2, and so on [(Butcher *et al*., 2001) and section 1.1.3.2, C]. Structures of ϕ6 RdRp in complex with template have shown that T1 is positioned well past the expected catalytic site residues with the base threaded into a specific binding pocket within the CTD. This specificity pocket (S) is too small to accommodate a purine base and is likely to select for pyrimidines cytidine and uracil.

Preference for a 3' cytidine over uracil may result from specific hydrogen bond interactions within the S pocket, such as O^{ε} of E634 with N4 of cytidine, which is not possible with the O4 of uracil.

The edge of the template channel is shaped like a plough, adjacent to a positively charged path over the surface of the polymerase (**Fig. 1.10, D**). Site-directed mutagenesis studies showed that residues R30 and K541, at the entrance to the tunnel, are essential for efficient template binding (Sarin *et al*., 2009). These residues sit at the centre of two highly charged basic residue patches on either side of the entrance to the tunnel and presumably make non-specific interactions with the phosphate backbone of RNA and direct the template down into the tunnel. In addition, it was proposed that the plough-like protuberance along the surface of the polymerase might separate the strands of RNA, feeding one directly into the positively charged tunnel and the other out of the capsid, possibly through a positively charged surface channel (Butcher *et al*., 2001); see **Fig. 1.10, D**. This helicase activity is likely to be facilitated by residues K247, D248 and R251 and exploit the AU-rich region at the 5' end of the (+) strand of ϕ6 dsRNA (Bruenn, 2003), although the mechanism remains elusive. In addition, discrimination at T1 by structural elements present at the entrance to the template tunnel and again at the S pocket explain how the virus may regulate the level of transcription of each genomic strand and promote production of (+)ssRNA over (-)ssRNA from the dsRNA template.

The *de novo* initiation priming platform of the active site, that facilitates the base-pairing of incoming NTPs with bound template, is formed by residues Q629, Y630, W631 and K632 of the CTD. Mutation of this loop to smaller residues (SG) gives rise to a mutant polymerase that is prone to back-priming initiation (Laurila *et al*., 2005). This type of initiation is

deleterious *in vivo* since the newly produced daughter strand would remain covalently attached to the template strand. This suggests that ϕ6 RdRp has developed a discreet set of residues at the CTD that protrude into the active site and ensure *de novo* RNA synthesis *in vivo*, preventing back-priming from occurring. The bulky side chain of W632 is thought to hold the priming platform in place by bridging interactions with the 305-308 loop region of the palm domain, and Y630 is essential for stabilising the initiation complex in the active site by base stacking with the first incoming D1 NTP (Laurila *et al*., 2005), see section 1.1.3.4 for details. Back-priming has also been observed for HCVpol and BVDVpol and is favoured *in vitro* over *de novo* initiation (Behrens *et al*., 1996; Kim *et al*., 2000; Lampio *et al*., 2000; Zhong *et al*., 1998). Interestingly, Hong *et al*., (2001) showed that the β-hairpin (part of the HCVpol CTD) acts as a 'gate' preventing RNA template slipping through the RdRp and ensuring only the single-stranded 3'-end of bound template can enter the active site (Hong *et al*., 2001). In addition it was noted that mutations in this structural feature could alter the mode of initiation as observed for ϕ6 RdRp. For example, mutation of the HCVpol β-hairpin (^{443}LDCQIYGACYSI^{454}) to LGGI increased the propensity of HCVpol to initiate from an internally annealed primer as proposed to the wild-type that could only utilize short primers complementary the 3'-end of bound template (Hong *et al*., 2001). The important role of these CTD protrusions in controlling the mode of initiation in primer-independent RdRps provides an explanation for the absence of this additional domain feature in vRdRps that utilise primer-dependent mechanisms.

1.1.3.4. Mechanism of initiation and polymerisation

RNA synthesis by RdRps involves template binding, initiation, elongation and finally termination. The first insights into the molecular details of initiation were provided by structural studies of ϕ6 RdRp in complex with a 5nt DNA template and NTP substrates (Butcher *et al*., 2001). These studies provide an atomic model for *de novo* initiation of RNA polymerisation (Butcher *et al*., 2001; Salgado *et al*., 2004), see **Fig. 1.11**. In brief, the template RNA is bound in the template tunnel, with the 3' cytidine base (T1) being recognised by the S pocket within the CTD. An interrogation site (I) is defined by a set of positively charged residues at the entrance of the NTP tunnel (see section 1.1.3.3.) which lock the phosphate moieties of incoming NTPs in position, allowing the base component to interrogate the template sequence. Incoming GTP (D2) is stabilised as it base pairs to a penultimate 3' cytidine (T2) by packing against Y630, defined as site P, of the initiation stabilizing platform of the CTD (629QYWK632). The template then moves backwards, releasing the 3'-end cytidine (T1) from pocket S. This allows a second GTP (D1) to enter the active site and base pair with T1 at site P, stabilized again by Y630 and locking the initiation complex into its final conformation. Catalysis can then occur, with release of pyrophosphate (PPi). The next NTP slips in, setting the ratchet, and chain elongation is underway. In ϕ6 RdRp, the double-stranded RNA (dsRNA) exit path is blocked by the CTD (**Fig. 1.9.**). Therefore, displacement of this domain is necessary for elongation to occur. When the nascent dsRNA product reaches a critical length, the CTD must be repositioned so that the extending product can exit from the contained active site of the polymerase (Makeyev and Grimes, 2004; van Dijk *et al*., 2004). The mechanism of how the CTD and therefore the priming platform is destabilised and subsequently displaced from the main body of the polymerase is still ambiguous and will be explored in detail in this thesis.

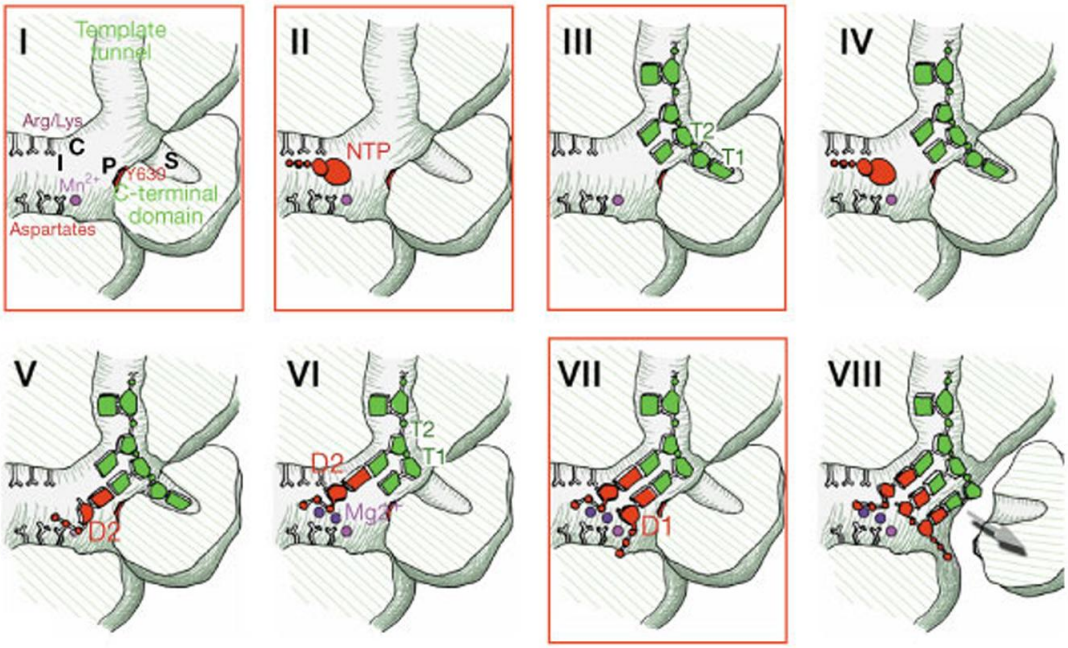


Figure 1.11. Model for ϕ6 RdRp initiation.

I-IV – Template enters the tunnel, slotting into site S (specificity pocket). Site I (interrogation) is occupied by NTPs, presumably in rapid exchange. **V** – GTP binds to site P (initiation stabilising platform). **VI** – Template ratchets backwards moving T1 away from the S pocket. **VII** – A second GTP binds into site P locking the initiation complex in its final state. **VIII** – Catalysis occurs with release of PPi. **VII-VIII** – The next NTP slips into site C from site I, setting the ratchet, and chain elongation is underway. Displacement of the C-terminal domain (CTD) is then necessary for elongation to occur. Red boxes indicate stages captured by X-ray crystallography. [Figure from Butcher *et al*. (2001)].

1.2. Directed evolution, DNA fragmentation & Structural Biology

In recent years, it has become clear that only a small fraction of natural proteins produced from whole open reading frames (ORFs) of cloned cDNA are suitable for structure determination because of bottlenecks such as: poor expression, aggregation, misfolding of proteins, and difficulties in solubilisation and crystallisation. In an effort to obtain tractable fragments of such proteins it is common to use bioinformatics prediction to identify stable globular domains and express only these regions, or for difficulties with crystallisation to carry out limited proteolysis of the isolated full length protein in attempt to remove disordered regions.

A wide range of techniques has been developed that attempt to bypass 'traditional' approaches by exploiting high-throughput (HTP) screening of DNA fragment libraries to identify tractable fragments of proteins for structural studies [reviewed by Hart and Tarendeau (2006), Prodromou *et al*., (2007) and Savva *et al*., (2007)]. The library can be generated by either systematic (ordered) DNA fragmentation using multiple designed PCR primers in a sequential fashion (termed primer pair walking), or by random DNA fragmentation using physical (e.g. sonication), enzymatic or PCR-based methods (such as tagged-PCR). For expression of protein fragments produced from the library-based methods a number of expression vectors have been developed that provide tags that allow for both detection of protein expression and subsequent generic affinity-based purification (Savva *et al*., 2007; Waldo, 2003). Expression of the DNA fragment library typically leads to tens of thousands of independent clones, only a subset of which express the fragment DNA in the correct reading frame and orientation leading to a protein product (**Fig. 1.12.** and **Fig. 1.13.**).

A screening and selection step is introduced to assess whether the protein produced is tractably soluble, a soluble aggregate, or expressed in inclusion bodies. Examples include: novel screens for soluble protein expression (Cornvik *et al*., 2005) and directed evolution approaches to enhance protein solubility (Pedelacq *et al*., 2002; Tarendeau *et al*., 2007; Waldo, 2003; Waldo *et al*., 1999). A general outline of the library-based DNA fragmentation approach (also termed 'domain trapping') is shown in **Fig. 1.12.** and a comparison with conventional methods shown as a schematic in **Fig. 1.13.**

Figure 1.12. Ten step work-flow for soluble protein domain trapping experiments starting from a single template DNA.

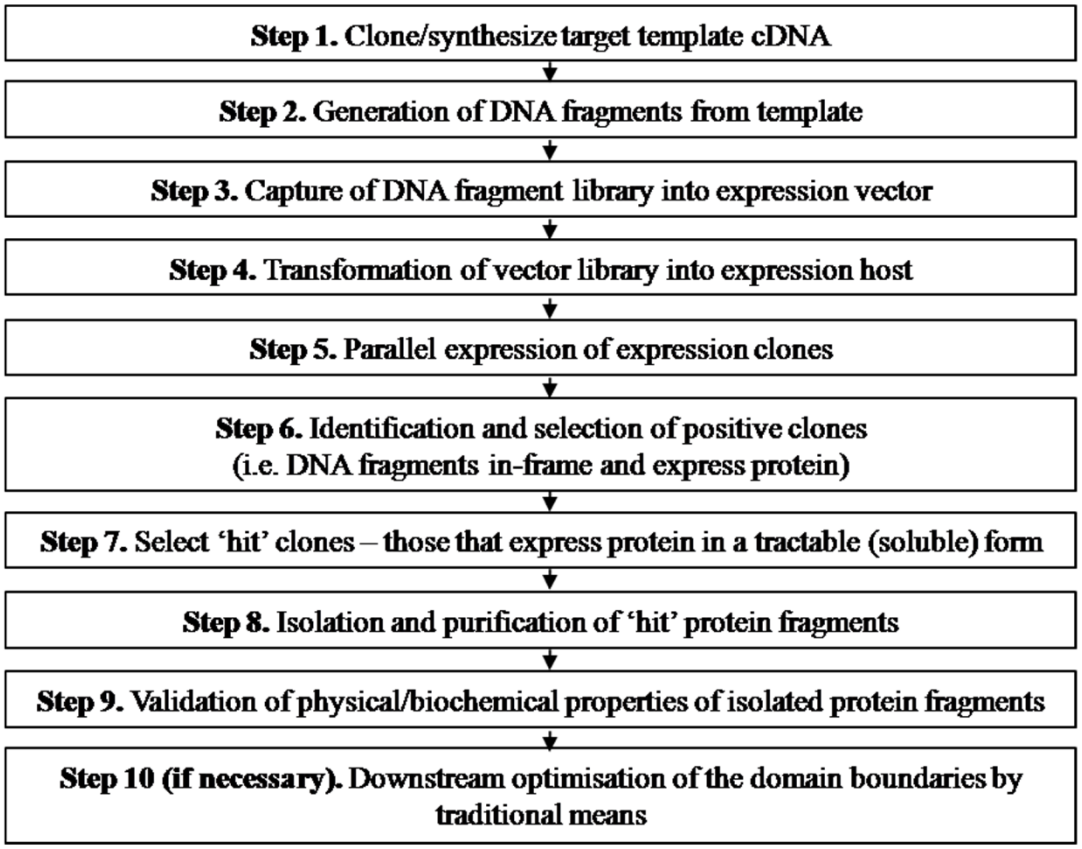
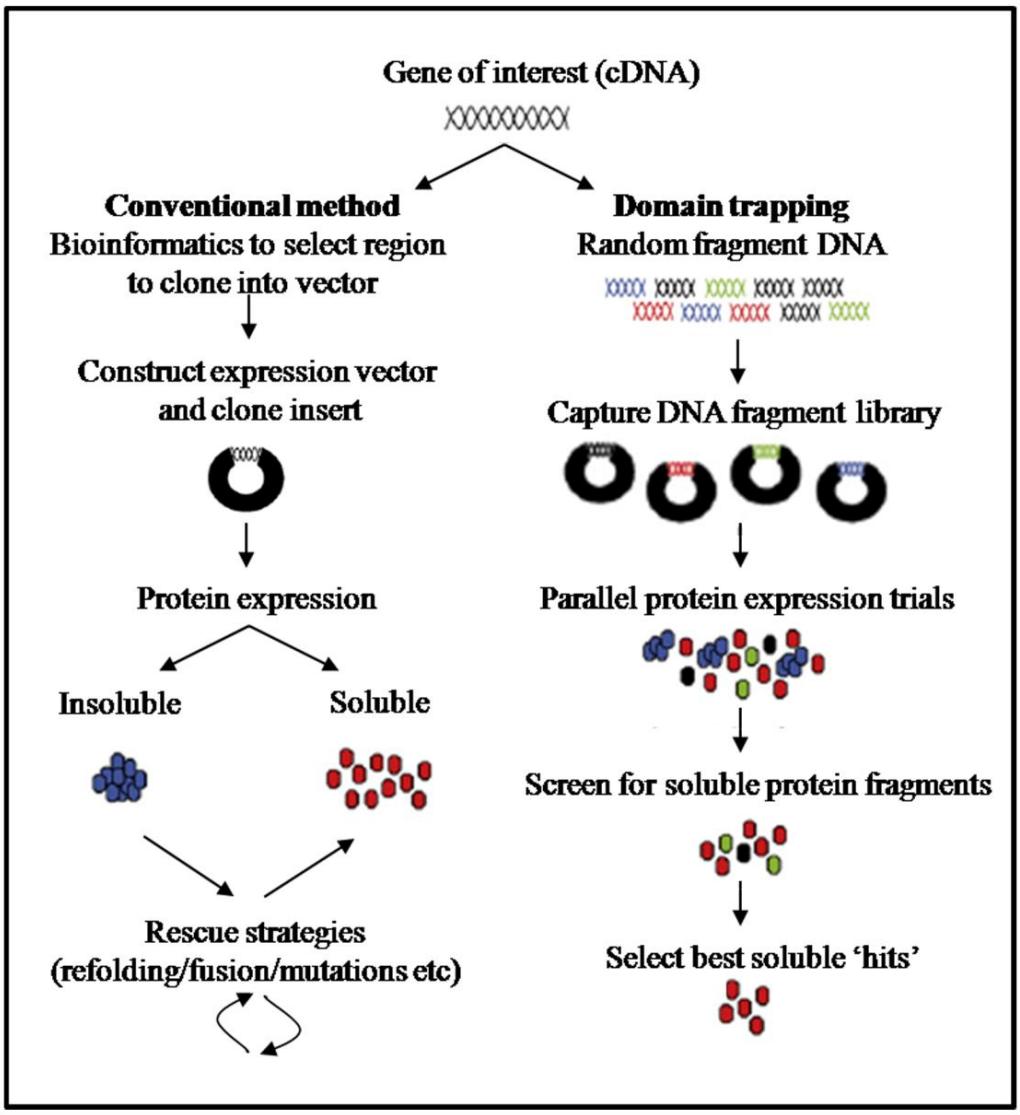


Figure 1.13. Schematic indicating the conventional approach to targeted expression of protein products compared with protein expression screening approaches based upon random or combinatorially generated DNA fragments. [Figure modified from (Prodromou *et al*., 2007; Savva *et al*., 2007)].



1.2.1. ESPRIT - Expression of Soluble Protein by Incremental Truncation

The ESPRIT method (Alzari *et al*., 2006; Hart, 2004) is designed to introduce directed evolution methods into expression screening of a random DNA fragmentation library produced by enzymatic digestion of the gene of interest (see **Fig. 1.12 and 1.13.**). Based on well-established molecular biological protocols (Henikoff, 1984; Ostermeier *et al*., 2002), exonuclease III (ExoIII) is used to specifically digest DNA from a 5´ protruding or blunt-end; the gene is cloned into an expression vector with restriction sites at either end that can leave ExoIII sensitive sites upon cleavage. The ExoIII reaction conditions are such that the ExoIII cuts at a slow rate (typically 20-50 bp/min) into the gene of interest. At regular intervals during the digest, samples are removed from the reaction mixture and quenched giving rise to a population of truncations of different lengths. This can be done in a unidirectional fashion, digesting from only one end of the gene, or bi-directionally, truncating one end after the other (as illustrated in **Fig. 1.14.**).

Detection of soluble expression relies on a novel reporter system (Hart, 2004), which consists of a short peptide (GLNDIFEAQKIEWHE) that is provided by the expression vector and becomes attached to the C-terminus of the protein fragments produced by ESPRIT. This peptide is a substrate for *E. coli* biotin ligase (BirA) enzyme (Beckett *et al*., 1999), which covalently attaches biotin to a specific lysine residue (Chapman-Smith and Cronan, 1999; Schatz, 1993); herein this sequence will be referred to as a biotin acceptor peptide (BAP) tag. The principle of this reporter system is illustrated in **Fig. 1.15** and relies on the assumption that efficient biotinylation on the BAP tag is only possible for correctly folded (and likely to be soluble) protein domains (i.e. ‘domain trapping’). In addition the expression vector provides a cleavable hexahistidine tag (His_6-tag) that becomes attached to the N-terminus of

protein fragments produced by ESPRIT, allowing biotinylated hits to be purified by affinity chromatography.

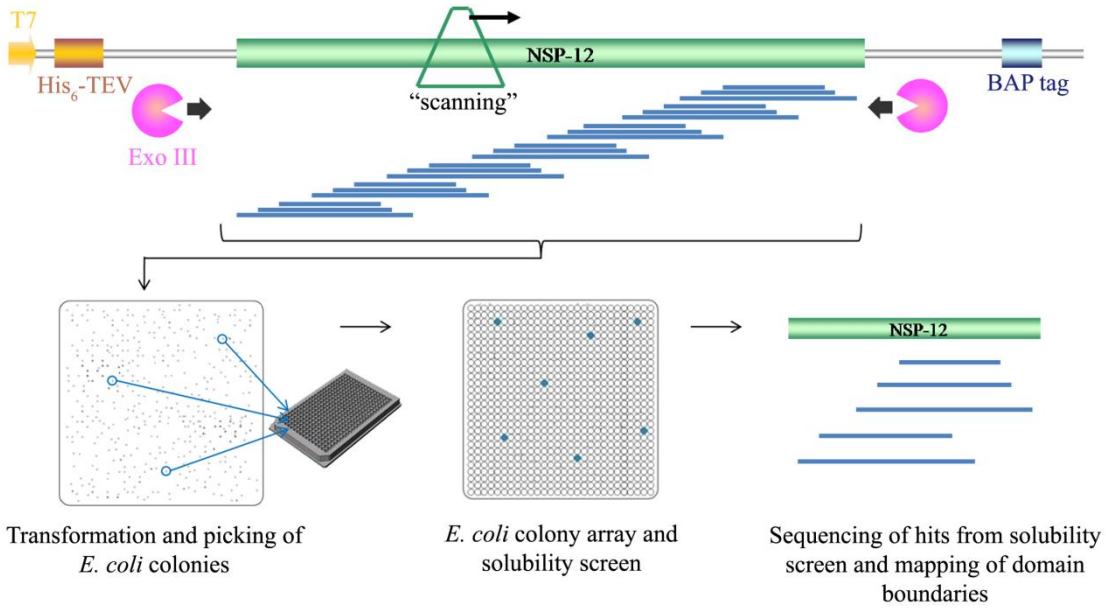
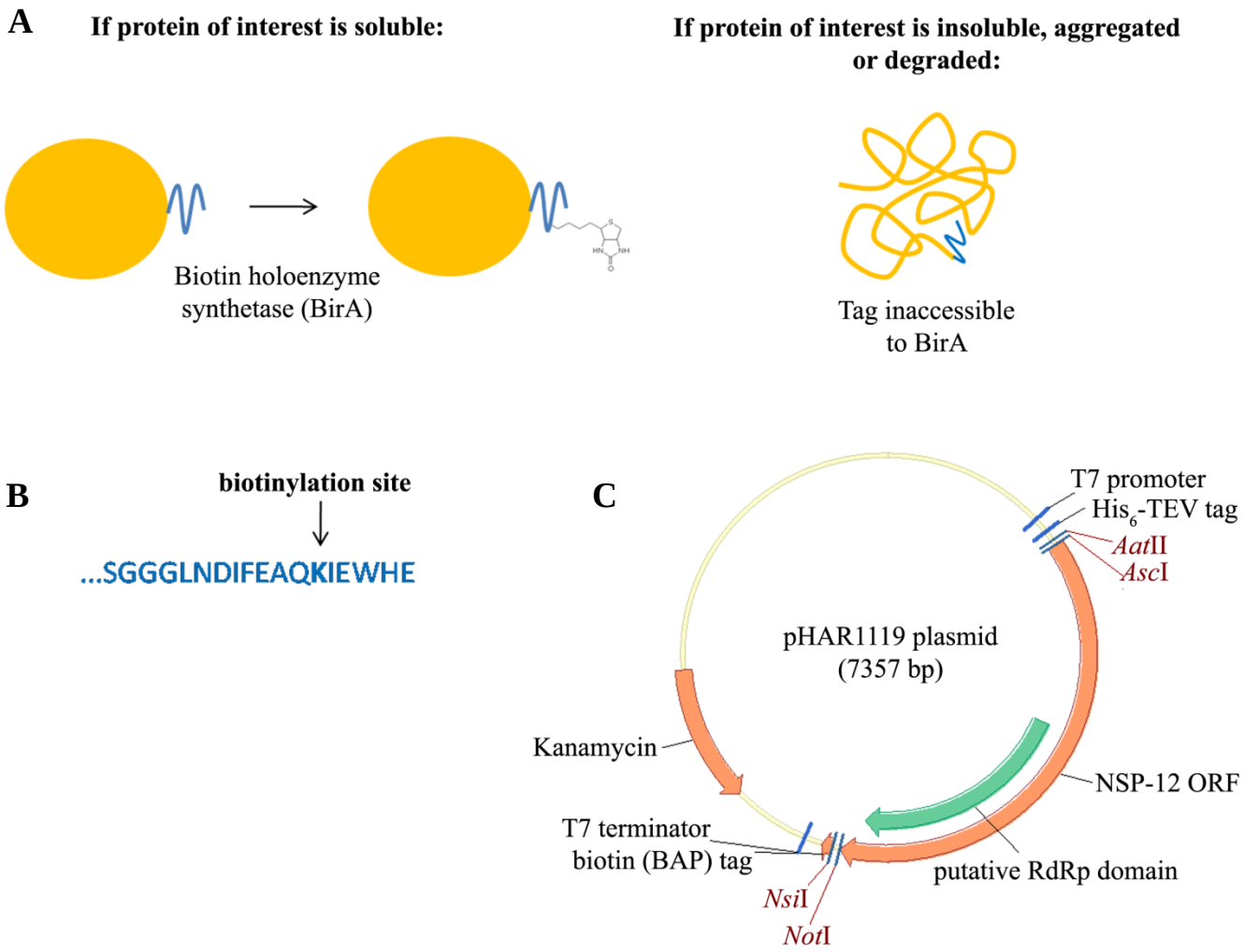


Figure 1.14. An overview of the ESPRIT method.

The ESPRIT incremental truncation experiment consists of four stages. First, the gene of interest (green) is amplified and cloned into the appropriate screening vector (pHAR1119 with T7 promotor, see **Fig. 1.15.**), which contains a 5' hexahistidine tag (His_6-TEV, orange) cleavable with Tobacco Etch Virus (TEV) protease, and a 3' biotin acceptor peptide (BAP tag, cyan) sequence. Second, a library of unidirectionally or bidirectionally truncated mutants is generated by ExoIII digestion with the reaction quenched a regular intervals; the library is then quality-tested using molecular biological methods. Third, the library is transformed into *E. coli* and screened (in a large colony array format) for truncations of the protein which give rise to 'soluble' His_6-tagged protein that can be purified. The His_6-tag is probed with an anti-His_6 antibody followed by an anti-mouse antibody conjugated to a fluorophore. Then, the biotinylated BAP tag is probed with streptavidin conjugated to a different fluorophore. Fourth, selected best 'hits' are sequenced and domain boundaries identified.

Figure 1.15. The ESPRIT solubility reporter technology

(A) An illustration of the solubility reporter peptide. If the protein of interest (shown in orange) is soluble, the reporter peptide (shown in blue) will be accessible to the *E. coli* biotinylation enzyme BirA and becomes biotinylated. If the protein of interest is insoluble, the reporter is inaccessible and remains unmodified. **(B) The sequence of the reporter peptide.** The biotinylation site is indicated. **(C) The screening vector (pHAR1119) used in this experiment** The vector contains the full-length SARS polymerase NSP-12 gene, under the control of a T7 promoter. The biotin acceptor peptide (BAP tag) is fused to the 3' end of the gene and His_6-TEV tag at the 5' end. Restriction enzyme sites *Aat*II, *Asc*I, *Not*I and *Nsi*I that flank the NSP-12 ORF (see Chapter 4 for details) and the position of the kanamycin antibiotic resistance gene are also indicated. [Image generated using Vector NTI Invitrogen].



1.2.1.1. SARS-CoV NSP-12 as a test-bed for ESPRIT

Structural work on coronaviruses, particularly severe acute respiratory syndrome coronavirus (SARS-CoV) is still one of the most competitive fields in life sciences with laboratories worldwide devoting significant resources to this work. Although structures of SARS-CoV replicase proteins NSP-9 (Sutton *et al*., 2004), NSP-15 (Ricagno *et al*., 2006; Xu *et al*., 2006), NSP-7/NSP-8 hexadecamer (Zhai *et al*., 2005), and the adenosine diphosphate-ribose-1-phosphatase (ADRP) domain of NSP-3 (Saikatendu *et al*., 2005) have been published, there are still many functionally well-understood open reading frames of SARS-CoV and other coronaviruses that lack high-resolution structures; the 3C-like protease is an exception, where crystal structures are available in several coronaviruses, including SARS-CoV (Anand K *et al*., 2002; Yang H *et al*., 2003).

For more than six years now researchers have struggled to obtain the crystal structure of the RdRp (NSP-12) from SARS-CoV because of first stage bottlenecks such as poor protein expression and solubility. In order to test whether the development of HTP library-based DNA fragmentation and complementary screening methods could be used to isolate domains of NSP-12 that could be expressed solubly, the ESPRIT technique was used in a collaborative project with Darren Hart and co-workers in Grenoble, France (see Chapter 4). For background, a brief overview of SARS-CoV is given that focuses on the current understanding of the function and domain organisation of NSP-12.

1.3. SARS Coronavirus

Coronaviruses are a family of higher animal disease viruses that infect both avian and mammalian species and are typically associated with respiratory, enteric, hepatic and central nervous system diseases. In man, coronaviruses are responsible for a large proportion of common colds and upper respiratory tract infections. In 2003, a novel coronavirus was identified as the causative agent of SARS (Drosten *et al*., 2003), an infectious respiratory tract infection reported in Asia, North America, and Europe. The global outbreak of the SARS-CoV affected more than 8000 people, caused more than 700 deaths and demonstrated the enormous pathogenic potential associated with this family of viruses.

1.3.1. Coronavirus Structure and Genome Organisation

Coronaviruses (family *Coronaviridae,* order *Nidovirales*) are enveloped viruses with a linear, non-segmented, positive-sense, single-stranded RNA genome [(+)ssRNA] of approximately 16 to 31 kb. They have been subdivided into 3 major groups based on serological and genetic properties. SARS CoV is most closely related to group 2 coronaviruses (Snijder EJ *et al*., 2003) and has a 29.7 kb genome. The replicative polyprotein (5' two-thirds of the genome) is divided into two overlapping open-reading frames (ORF1a and ORF1b) separated by a +1 frameshift (**Fig. 1.16, B**). The translated replicase polyprotein is proteolytically processed by two main proteases (papain-like protease and 3C-like cysteine protease) to give rise to the SARS CoV non-structural proteins (**Fig. 1.16, B**). The remaining third of the genome encodes SARS CoV structural proteins. The SARS CoV virion is made up of a helical nucleocapsid, which consists of nucleocapsid protein (N) and the viral genomic RNA, that is surrounded by a lipid bilayer (derived from intracellular membranes) containing the

membrane glycoprotein (M, also known as Matrix), the envelope protein (E) and the spike glycoprotein (S); the S protein forms a crown-like *corona* that gives the virus its name.

Figure 1.16. SARS-CoV structure and genome organisation.

(A) Schematic of the SARS CoV virion. The key structural proteins N, S, E and M are indicated. **(B) Genome organisation of SARS CoV.** The position of the replicase polyprotein and structural proteins are indicated. The replicative polyprotein is divided into two overlapping open-reading frames (ORF1a and ORF1b) separated by a +1 frameshift (indicated by an arrow). The translated replicase polyprotein is proteolytically processed by two main proteases; triangles represent sites that are cleaved by papain-like proteinases (orange) or the 3C-like cysteine proteinase (blue). [Modified from Snijder EJ *et al*., (2003)].

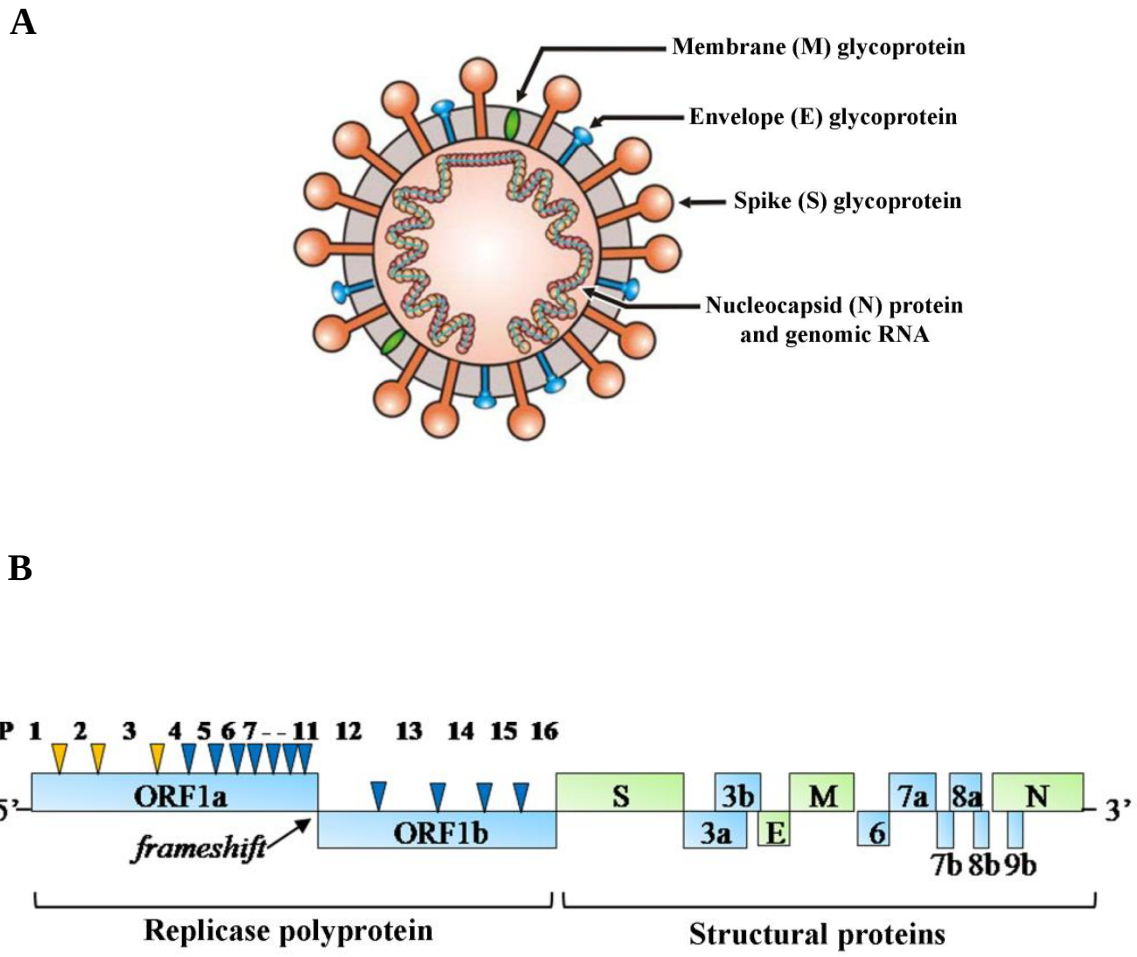


Table 1.3. Summary of the enzymatic functions associated with the SARS CoV replicative polyprotein.

Replicase region	SARS protein	Length (αα)	Position in polyprotein	Functional and structural predictions	Crystal structure available?
ORF1a	NSP-1	180	1–180	Host mRNA degradation	NMR only (Almeida *et al*., 2006)
	NSP-2	638	181-818	–	No
	NSP-3	1922	819–2740	Papain-like cysteine protease cleavage of NSP 1-4; ADRP; 2 TMD	Yes (Saikatendu *et al*., 2005)
	NSP-4	500	2741–3240	3 TMD	No
	NSP-5	306	3241–3546	3C-like cysteine protease cleavage of NSP 4-16	Yes (Anand K *et al*., 2002; Yang H *et al*., 2003)
	NSP-6	290	3547–3836	5 TMD	No
	NSP-7	83	3837–3919	Processivity factor for RdRp	Yes (Zhai *et al*., 2005)
	NSP-8	198	3920–4117	Processivity factor for RdRp	Yes (Zhai *et al*., 2005)
	NSP-9	113	4118–4230	ssRNA-binding protein	Yes (Sutton *et al*., 2004)
	NSP-10	139	4231–4369	Transcription factor (novel dodecamer zinc-finger protein); GFL	Yes (Joseph *et al*., 2006; Su *et al*., 2006)
	NSP-11	13	4370–4382	–	No
ORF1b	NSP-12	932	4370–5301	RdRp	No
	NSP-13	601	5302–5902	Helicase (HEL1), zinc-binding domain,; NTPase	No
	NSP-14	527	5903–6429	3' → 5' exonuclease (ExoN homologue)	No
	NSP-15	346	6430–6775	Uridylate specific endonuclease (XendoU homologue)	Yes (Ricagno *et al*., 2006; Xu *et al*., 2006)
	NSP-16	298	6776–7073	mRNA cap-1 methyltransferase	No

Table Abbreviations: TMD = transmembrane domain(s); ADRP = adenosine diphosphate-ribose-1-phosphatase; GFL = growth factor-like domain.
[Modified from (Stadler K *et al*., 2003) and (Bartlam *et al*., 2005)].

1.3.2. SARS-CoV RNA-dependent RNA polymerase (NSP-12)

Non-structual protein 12 (NSP-12) is the RdRp of the virus, which is well-conserved in all known coronaviruses (Ziebuhr, 2005), and is primarily responsible for the transcription of the viral RNA genome. It is derived from a replicative polyprotein that is processed extensively by viral proteases to produce the functional subunits of the virus replication machinery (see **Fig. 1.16.** and **Table 1.3.** for a summary table of the enzymatic functions associated with the SARS CoV replicase polyprotein).

NSP-12 consists of 932 amino acids (full length molecular weight 105 kDa), significantly larger than typical viral polymerase enzymes (Bruenn, 2003). Sequence alignments and comparisons indicate that SARS-CoV RdRp has a high sequence identity with other coronavirus RdRps (~62-73%), (Xu *et al*., 2003). However, it shares <10% sequence identity with other viral RdRps and RTs, including RdRps from PV, HCV, RHDV, RV, ϕ6, and the HIV-1 RT whose structures are known. At present there is contradictory evidence in the literature about the structural organisation of the protein. Three different models describing the putative domain organisation of NSP-12 have been developed, and suggestions about functional roles, based mostly on bioinformatic analysis, have been made (summarised in **Fig. 1.17**).

Figure 1.17. Three models for the domain organisation of NSP-12.

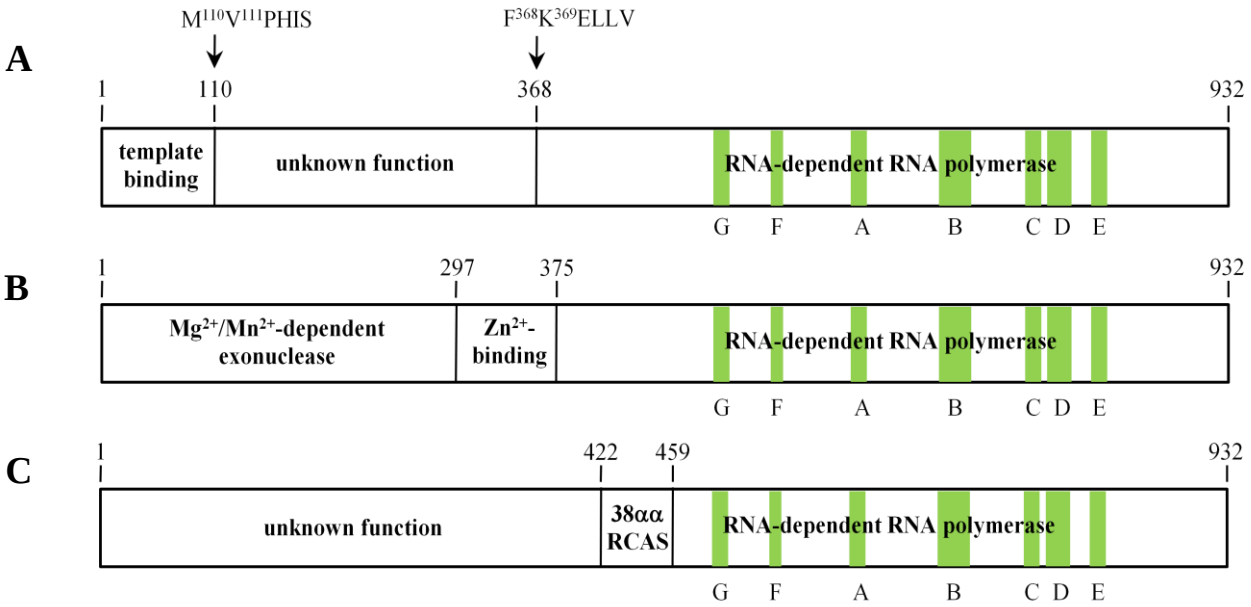
Polymerase signature motifs have been identified by bioinformatic methods (Bruenn, 2003; Xu *et al*., 2003) and are marked in green on all three models (A-C).

(A) Model derived from proteolytic studies (Cheng *et al*., 2005). The protein is thought to have three domains: an N-terminal domain with putative template primer binding function; a central domain with unknown function; and a C-terminal region which contains the RNA-dependent RNA polymerase activity. N-terminal sequencing of the proteolytic fragments was used to identify the cleavage sites shown by the arrows and respective sequence.

(B) Model derived from bioinformatic analysis and expression studies (Moll and Hilgenfeld, 2004). In this model, NSP-12 is thought to consist of three domains: an N-terminal Mg^{2+}/Mn^{2+}-dependent exonuclease, (homologous to the RNAse H enzyme from bacteriophage T4); a central Zn^{2+}-binding domain; and a C-terminal RNA-dependent RNA polymerase domain.

(C) Model derived from bioinformatic analysis (Brockway *et al*., 2003). RCAS is the 38 amino acid replicative complex assembly sequence first identified in Mouse Hepatitis Virus (MHV) and its identification in SARS is reported in Brockway *et al* (2003) and has the following sequence: ^{422}FNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISD^{459}.

(D) Sequences of conserved RNA-dependent RNA polymerase motifs A-G.



D

Motif G: ^{499}DKSAGFPFNKWGK^{511}
Motif F: ^{544}LKYAISAKNRARTVAGV^{560}
Motif A: ^{612}PHLMGWDYPKCDRAM^{626}
Motif B: ^{678}GGTSSGDATTAYANSVFNICQAVTANVNALLST^{710}
Motif C: ^{753}FSMMILSDDAVVCYN^{767}
Motif D: ^{771}AAQGLVASIKNFKAVLYYQNNVFMSE^{796}
Motif E: ^{810}HEFCSQHTMLV^{820}

All reports published so far agree that the RdRp catalytic domain occupies the C-terminal 50-60% of the protein (approximately residues 376-932), even though the precise domain boundaries are contentious. The remaining N-terminal portion (residues 1-375) of SARS-CoV and other coronavirus RdRps is large and has been named the N-terminal domain (NTD), but has no counterpart in other (+)ssRNA virus RdRps of known structures. Perhaps the main reason for this lack of definitive knowledge regarding this protein lies in the fact that it is not readily expressable; indeed, there are only two reports about expression of parts of the protein at extremely low levels (Cheng *et al*., 2005; Moll and Hilgenfeld, 2004), summarised in **Fig. 1.17**. Work by Cheng *et al*. (2005) also demonstrated the high susceptibility of full length NSP-12 to proteolysis by bacterial proteases and suggested that the enzyme has many surface-exposed cleavage sites, presumably due to many individual domains connected by flexible regions.

Analysis of the amino acid composition of NSP-12 demonstrates that it is quite a hydrophobic protein (43% non-polar and hydrophobic, 32.6% neutral and polar, 11.5% acidic and 12.8% alkaline) and also encodes 31 cysteine residues (3.32% of full length protein with 2.83% in the putative catalytic domain); these features may contribute to its poor solubility and may also reflect an important biological function. The RdRp of closely related coronavirus Murine Hepatitis Virus (MHV) for example, associates with membranes during cell lysis, centrifugation and fractionation (Brockway *et al*., 2003; Gosert *et al*., 2002) due to a similarly high hydrophobic content. Immunoblotting and immunofluorescence of cell lysates infected with SARS-CoV also demonstrated that replicase proteins co-localised to cytoplasmic complexes containing markers for autophagosome membranes (Prentice *et al*., 2004), suggesting that the hydrophobic nature of NSP-12 may indeed play a key role in replication complex association with membranes in the cytoplasm of the infected cell.

Despite numerous attempts by several groups, SARS-CoV NSP-12 has proven to be difficult to produce in sufficient quantities for crystallization due to this poor solubility. As a result of this, there has been little success in developing coronavirus anti-virals that specifically target the RdRp and block viral replication. It is worth noting that RdRps of other RNA viruses have been major targets for antiviral compounds, such as HCV with non-nucleoside inhibitors (NNI), (Biswal *et al*., 2005). The lack of structural information for SARS-CoV replicase proteins has resulted in a shift towards structure-based discovery of anti-coronavirus therapeutics that target the spike (S) protein (where the crystal structure is available) and prevent virus entry. The research reported herein explores new high-throughput library-based methodologies, such as ESPRIT, in an attempt to isolate sufficient soluble NSP-12 protein, or domains thereof, to set up crystallization trials and hopefully provide structural information that can shine some light on the underlying mechanisms of NSP-12 polymerisation. In addition, this structural information would assist the development of a broader spectrum of inhibitors that could act as a first line of defence against emerging coronaviruses.

Chapter 2

ϕ6 RdRp – Experimental Procedures

Although much is known about the initiation mechanism of ϕ6 RdRp from structures published by Butcher *et al*., (2001), Laurila *et al*., (2005) and Salgado *et al*., (2004), many questions still remain unanswered:

(i) What are the key structural features on the surface of the polymerase that direct separation and threading of ssRNA from genomic dsRNA into the ϕ6 RdRp active site during preinitiation?

(ii) Can NTPs stably bind at the active site and *de novo* priming platform in the absence of Watson-Crick base pairing with template during preinitiation?

(iii) After formation of the initiation complex and synthesis of the dinucleotide daughter strand, how does the polymerase transition to an elongation phase?

(iv) The CTD containing the *de novo* priming platform is a structural barrier for the exit path of the newly synthesized dsRNA. This CTD must be displaced for the addition of the third nucleotide of the daughter strand. How are the strong interactions between the CTD and main body of the protein destabilized?

(v) What is the role of the non-catalytic divalent cation binding site near the active site and is Mn^{2+} bound in the purified polymerase?

(vi) Can Mg^{2+} and Mn^{2+} functionally replace one another at both the non-catalytic and catalytic sites?

To answer these questions a number of structural and biochemical studies were undertaken in a collaborative project with Prof. Dennis Bamford's group at the University of Helsinki. Crystal structures of WT ϕ6 RdRp, an E634 to Q mutant (lacking a salt bridge between E634 and K144 connecting the CTD to the main body of the protein) and C-terminally trypsin-cleaved RdRp ($P2^{T1}$) were determined, in most cases in complex with different oligonucleotide templates, several NTP substrate conditions and in the presence of different divalent cations. In addition, the effect of Mg^{2+} and Mn^{2+} on the thermal stability of WT, E634Q and an E491 to Q mutant [with reduced affinity for Mn^{2+} at the non-catalytic ion site (Poranen *et al*., 2008b)] was assayed by thermal denaturation experiments.

Structural and biochemical studies of ϕ6 RdRp described in this thesis (**Table 2.1**) are grouped into different subcategories according to the questions being addressed:

(i) Detailed structural information on template entry, NTP positioning and role of Mg^{2+} during preinitiation was obtained from WT ϕ6 RdRp crystals soaked with linear or hairpin DNA, ATP and/or GTP and Mg^{2+} (WT-Tri4T-ATP-Mg^{2+}, WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}, WT-5'-AATCT-3'-GTP-Mg^{2+}) and E634Q mutant crystals soaked with linear RNA or DNA, GTP, Mg^{2+} and/or Mn^{2+} (E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} and E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+}).

(ii) The strong interactions of the CTD with the main body of the polymerase were evident from the crystal structure of C-terminally trypsin-digested WT ϕ6 RdRp ($P2^{T1}$).

(iii) The apo structure of the E634Q mutant was determined, together with a set of complexes from crystals soaked with hairpin DNA, ATP analog (AMPPCP) and Mg^{2+} (E634Q-Tri2T-AMPPCP-Mg^{2+}, E634Q-Tri4T-AMPPCP-Mg^{2+}, E634Q-Tetra2T-AMPPCP-Mg^{2+}, E634Q-Tetra4T-AMPPCP-Mg^{2+}) and co-crystals with hairpin DNA in the presence of EDTA (E634Q-Tetra2T-coxtal). The results shed light on the molecular basis for the transition to elongation and importance of Mn^{2+} and the non-catalytic ion site.

(iv) To further understand the role of Mn^{2+} and Mg^{2+} on the stability of ϕ6 RdRp, thermal denaturation experiments were carried out under different divalent cation conditions for WT ϕ6 RdRp, E634Q and E491Q [a non-catalytic ion site mutant with reduced affinity for Mn^{2+} (Poranen *et al*., 2008b)].

In this chapter, the experimental procedures are described for each structure determination. A detailed analysis of the results obtained and how they contribute to a better understanding of ϕ6 RdRp-catalysed polymerisation is described in Chapter 3.

Table 2.1. ϕ6 RdRp structure determination described in this thesis.

Structure	ϕ6 RdRp	Nucleotide	Ion bound*	NTP
WT-5'-AATCT-3'-GTP-Mg^{2+}	WT	5nt (+)DNA mimic	Mg^{2+}	GTP
WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}	WT	5nt (+)DNA mimic	Mg^{2+}	ATP, GTP
WT-Tri4T-ATP-Mg^{2+}	WT	DNA hairpin	Mg^{2+}	ATP
Trypsin-digest WT ($P2^{T1}$)	WT	-	-	-
E634Q apo	E634Q	-	Mn^{2+}	-
E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}	E634Q	6nt (-)RNA mimic	Mg^{2+}	GTP
E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+}	E634Q	6nt (-)DNA mimic	Mn^{2+}	GTP
E634Q-Tri2T-AMPPCP-Mg^{2+}	E634Q	DNA hairpin	Mg^{2+}	AMPPCP
E634Q-Tri4T-AMPPCP-Mg^{2+}	E634Q	DNA hairpin	Mg^{2+}	AMPPCP
E634Q-Tetra2T-AMPPCP-Mg^{2+}	E634Q	DNA hairpin	Mg^{2+}	AMPPCP
E634Q-Tetra4T-AMPPCP-Mg^{2+}	E634Q	DNA hairpin	Mg^{2+}	AMPPCP
E634Q-Tetra2T-coxtal (EDTA)	E634Q	DNA hairpin	-	-

* ion bound in at least one molecule of the asymmetric unit (molecules I, II and III)

2.1. Protein Expression and Purification

Protein expression and purification of WT ϕ6 RdRp and E634Q was jointly conducted by myself in Oxford and by Minna Poranen as part of a collaboration with Prof. Dennis Bamford at the University of Helsinki. Expression and purification of a C-terminally truncated polymerase ($P2^{T1}$, cleaved at residue 607) was obtained by trypsin digestion of the purified WT ϕ6 RdRp; these experiments were carried out by Peter Sarin at the University of Helsinki. All structural and biophysical studies were carried out in Oxford as our main contribution to the project. The ϕ6 RdRp expression and purification procedures are detailed below.

2.1.1. Bacterial Strains and Plasmids

Escherichia coli (*E. coli*) DH5α (Gibco-BRL) containing plasmid pEMG2 [WT, (Makeyev and Bamford, 2001)] or pSVe2 [E634Q, (Sarin *et al*., 2009)] were donated as glycerol stocks by our collaborators in Dennis Bamford's group at the University of Helsinki. To isolate plasmid DNA, 5 mL of LB medium containing 200 μg/mL ampicillin was inoculated with 1 μL of glycerol stock in a 50 mL Falcon tube and incubated for 16 h at 37 ^{o}C in a shaking incubator at 225 rpm. Cell pellet was harvested by centrifugation at 6800 x g for 5 min at 4 ^{o}C and the supernatant discarded. Plasmid DNA was isolated by DNA miniprep (QIAprep Spin Miniprep Kit, QIAGEN) following the manufacturer's instructions. 50 μL of plasmid DNA was eluted from QIAprep spin columns in 10 mM Tris-HCl, pH 8.5 and the concentration measured from a 2 μL sample using a NanoDrop 1000 (Thermo Scientific) spectrophotometer at 260 nm.

2.1.2. Transformation

1-2 μL of 100 ng/μL isolated plasmid DNA (pEMG2 or pSVe2) were added to *E. coli* BL21(DE3) cells (Novagen) pre-thawed on ice (50μL cells/tube). After a 30 min incubation, *E. coli* BL21(DE3) cells were heat-shocked for 30 s at 42 ^{o}C in a pre-heated water bath and then returned to ice for 2 min. 500 μL of LB medium (no antibiotic selection) was then added to each tube and the cells allowed to recover for 1 h at 37 ^{o}C. 150 μL of cell mix was then plated out on LB Agar supplemented with 200 μg/mL ampicillin. Plates were allowed to dry upright (with lids removed) at 37 ^{o}C for 20 min. Lids were then replaced and the plates inverted and incubated overnight at 37 ^{o}C. Colonies were picked and used for protein expression and purification. To create a back-up stock of cells containing the desired expression vectors, single colonies were used to inoculate 5 mL of LB medium supplemented with 200 μg/mL ampicillin and grown overnight at 37 ^{o}C in a shaking incubator at 225 rpm. 100 μL of each culture was mixed with 100 μl of LB-glycerol (30% v/v glycerol) and stored at -80^{o}C.

2.1.3. Protein purification

Protein purification from *E. coli* BL21 (DE3) containing plasmid pEMG2 (WT) or pSVe2 (E634Q) was carried out as described by Makeyev and Bamford, (2000b) and Poranen *et al*., (2008b). Briefly, *E. coli* BL21 (DE3) cells containing the appropriate expression plasmid were streaked out from a -80°C stock on a LB-agar plate containing 200 μg/mL ampicillin. The plate was incubated overnight at 37°C. To prepare a starter culture, 5 colonies were transferred to 50 mL of LB-medium containing 200 μg/mL ampicillin, and incubated at 37°C while shaking at 225 rpm overnight. A 12.5 mL volume of starter culture was added to 500 mL of LB-medium containing 200 μg/mL ampicillin (40 fold dilution), and the cultures were

grown at 37°C shaking at 225 rpm until the OD_{600} reached 0.6. The cultures were then moved onto ice for 10 minutes. A 20 μL sample was taken of which 0.5 μL was analysed by sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) as a pre-induction control. The expression of WT or E634Q polymerase was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to each culture to a final concentration of 0.5 mM. The IPTG induced cells were then incubated at 20°C, shaking at 225 rpm for a further 20 h. After incubation the cell pellets were collected by centrifugation at 8000 x *g* at 4°C for 20 min. The cell pellets were then resuspended in 50 mL of buffer A1 (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 1 mM EDTA). A 20 μL sample was taken of which 0.5 μL was analysed by SDS-PAGE (see section 2.1.4).

All the following steps were performed at 4 °C. The purification of WT or E634Q polymerase was monitored by SDS-PAGE throughout the process. One protease inhibitor tablet (complete EDTA free, Roche Diagnostics) per 50mL was added to the cell suspension alongside DNaseI to a final concentration of 20 μg/mL. The cell suspension was passed twice through a pre-cooled Basic-Z cell disruptor (Constant systems) at 20 Kpsi to break up the bacterial cell walls. The lysate was centrifuged at 30,000 x *g* for 45 min to pellet cell debris. The supernatant was collected and filtered through a 0.45 μm sterile filter. The filtered supernatant was loaded on a dye affinity column containing Cibacron Blue 3GA agarose (Sigma) equilibrated with buffer A1. Blue 3GA agarose binds NTP-binding proteins. The column was washed with buffer A1. Protein was eluted from the column using 500 mM NaCl buffered with 50 mM Tris-HCl pH 8.0, 1 mM EDTA (**Fig. 2.1.**).

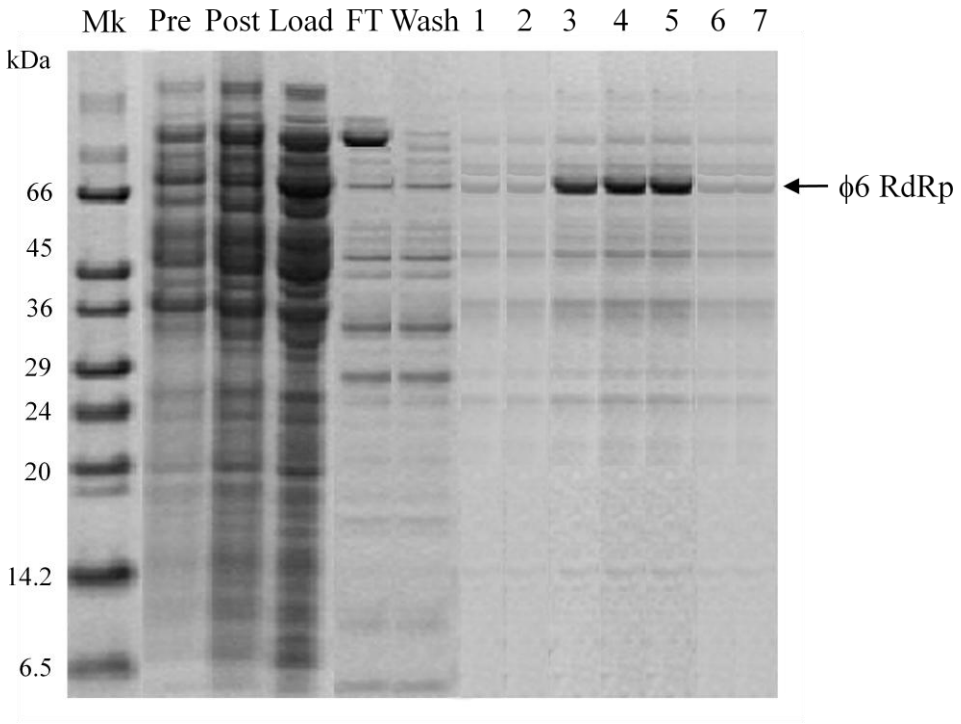


Figure 2.1. SDS-PAGE analysis of 1st step of ϕ6 RdRp purification (blue agarose affinity column).

SDS-PAGE gel of samples after blue agarose affinity column purification. Samples analysed: before and after cell disruption (pre and post, respectively); after centrifugation at 30,000 x *g* for 45 min (load); flow-through after loading (FT); washing steps (wash) and eluted fractions (1 to 7). SigmaMarker High Range (Sigma) was used as the molecular size marker (Mk). The position of ϕ6 RdRp is marked with a black arrow.

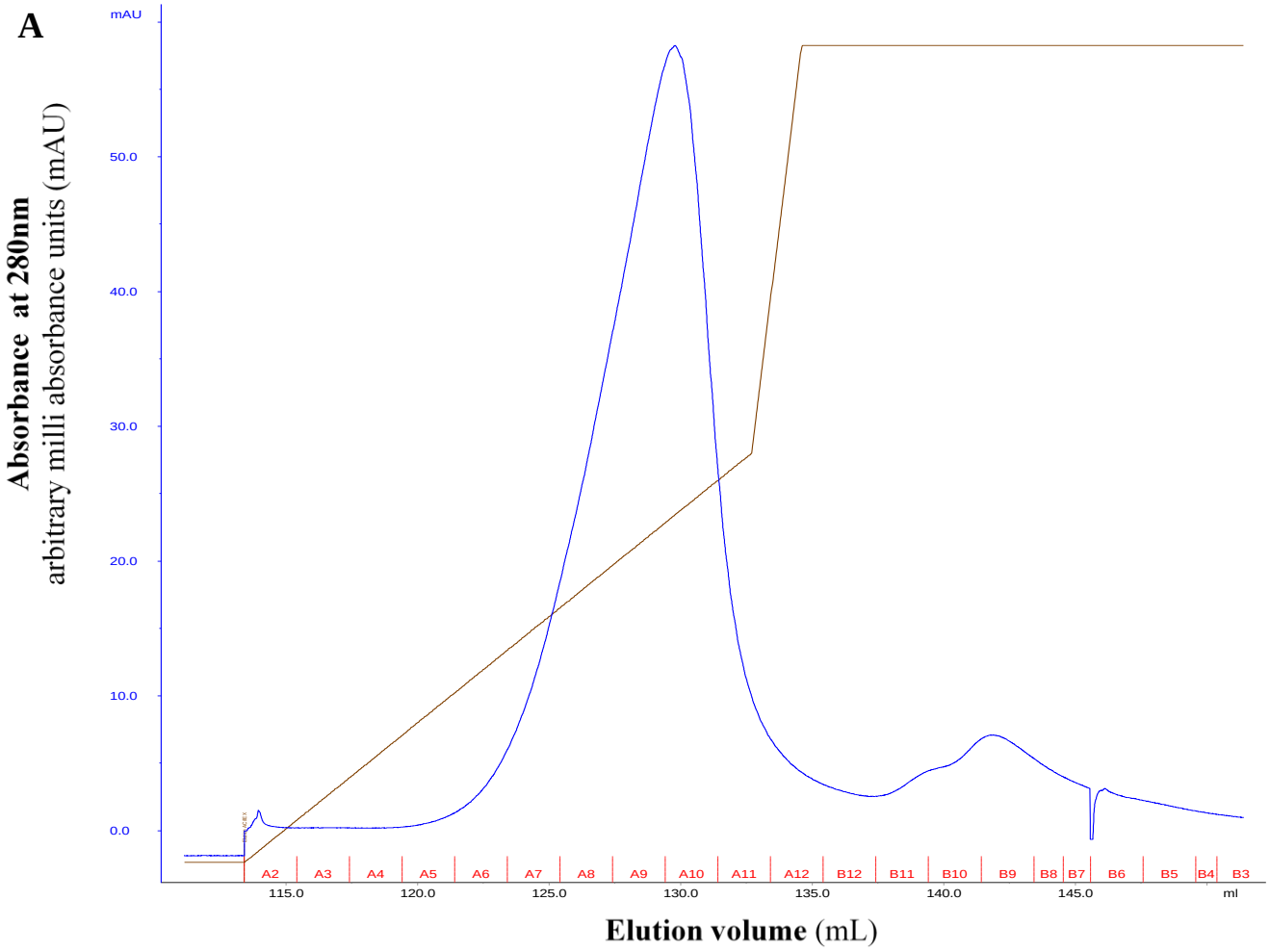
Column fractions containing WT or E634Q polymerase as analysed by SDS-PAGE were pooled, diluted 5-fold with ice-cold distilled water and filtered through 0.45 μm filter. The filtrate was loaded on a 5 ml Hi-Trap™ Heparin column (Amersham Biosciences) equilibrated with A1 buffer. Heparin binds proteins with affinity for oligonucleotides. Bound proteins were eluted with a linear 0.1-1 M NaCl gradient buffered with 50 mM Tris-HCl, pH 8.0 and 1 mM EDTA (**Fig. 2.2.**). Fractions containing WT or E634Q polymerase were pooled and diluted 10-fold with buffer containing 20 mM Tris-HCl, pH 8.0. The diluted protein was again filtered through a 0.45 μm filter, and loaded on a Hi-Trap™ Q Sepharose HP-column (Amersham Biosciences) equilibrated with QA-buffer (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA). This column performs separation by ion exchange. The column was washed with QA-buffer. The proteins were eluted with linear 0–1 M NaCl gradient buffered with 50 mM Tris-HCl pH 8.0 and 0.1 mM EDTA (**Fig. 2.3.**).

Figure 2.2. (overleaf). Analysis of ϕ6 RdRp purification products after heparin column.

(A) The elution profile (A_{280}) from the Hi-Trap™ Heparin column (Amersham Biosciences) with a peak corresponding to ϕ6 RdRp containing fractions. Elution gradient is indicated by the brown diagonal and the plateau indicates 100% elution buffer. Fractions are numbered and lettered in red. **(B)** SDS-PAGE gel of fractions eluted from the heparin column (fractions are named in accordance with elution profile in (A). SigmaMarker High Range (Sigma) was used as the molecular size marker (Mk). The position of ϕ6 RdRp is marked with a black arrow.

Figure 2.3. (overleaf). Analysis of ϕ6 RdRp purification products after Q ion exchange column.

(A) The elution profile (A_{280}) from the Hi-Trap™ Q Sepharose HP-column (Amersham Biosciences) with a peak corresponding to ϕ6 RdRp containing fractions. Elution gradient is indicated by the brown diagonal and the plateau indicates 100% elution buffer. Fractions are numbered and lettered in red. **(B)** SDS-PAGE gel of fractions eluted from the Q column (fractions are named in accordance with elution profile in (A). SigmaMarker High Range (Sigma) was used as the molecular size marker (Mk). The position of ϕ6 RdRp is marked with a black arrow.


A
mAU
50.0
40.0
30.0
20.0
10.0
0.0
Absorbance at 280nm
arbitrary milli absorbance units (mAU)
A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 B12 B11 B10 B9 B8 B7 B6 B5 B4 B3
115.0 120.0 125.0 130.0 135.0 140.0 145.0 ml
Elution volume (mL)

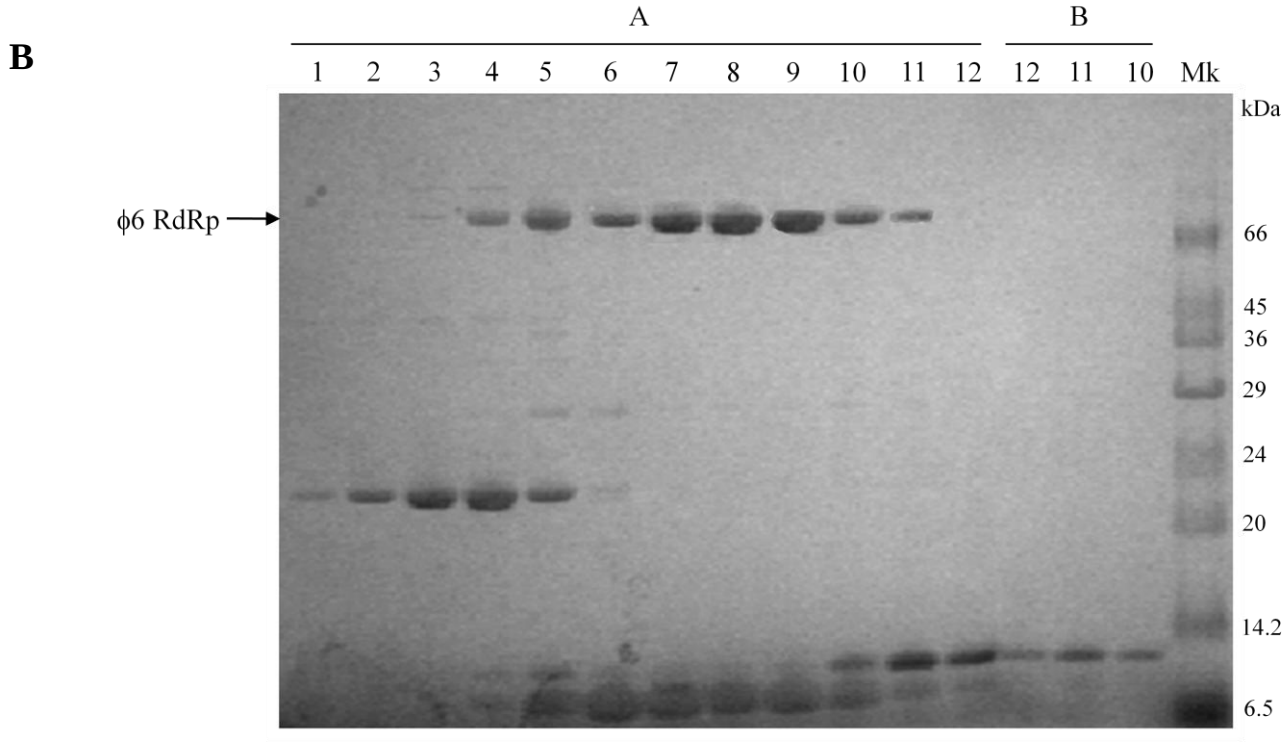

B
A
B
1 2 3 4 5 6 7 8 9 10 11 12 12 11 10 Mk
kDa
ϕ6 RdRp
66
45
36
29
24
20
14.2
6.5

Figure 2.2.

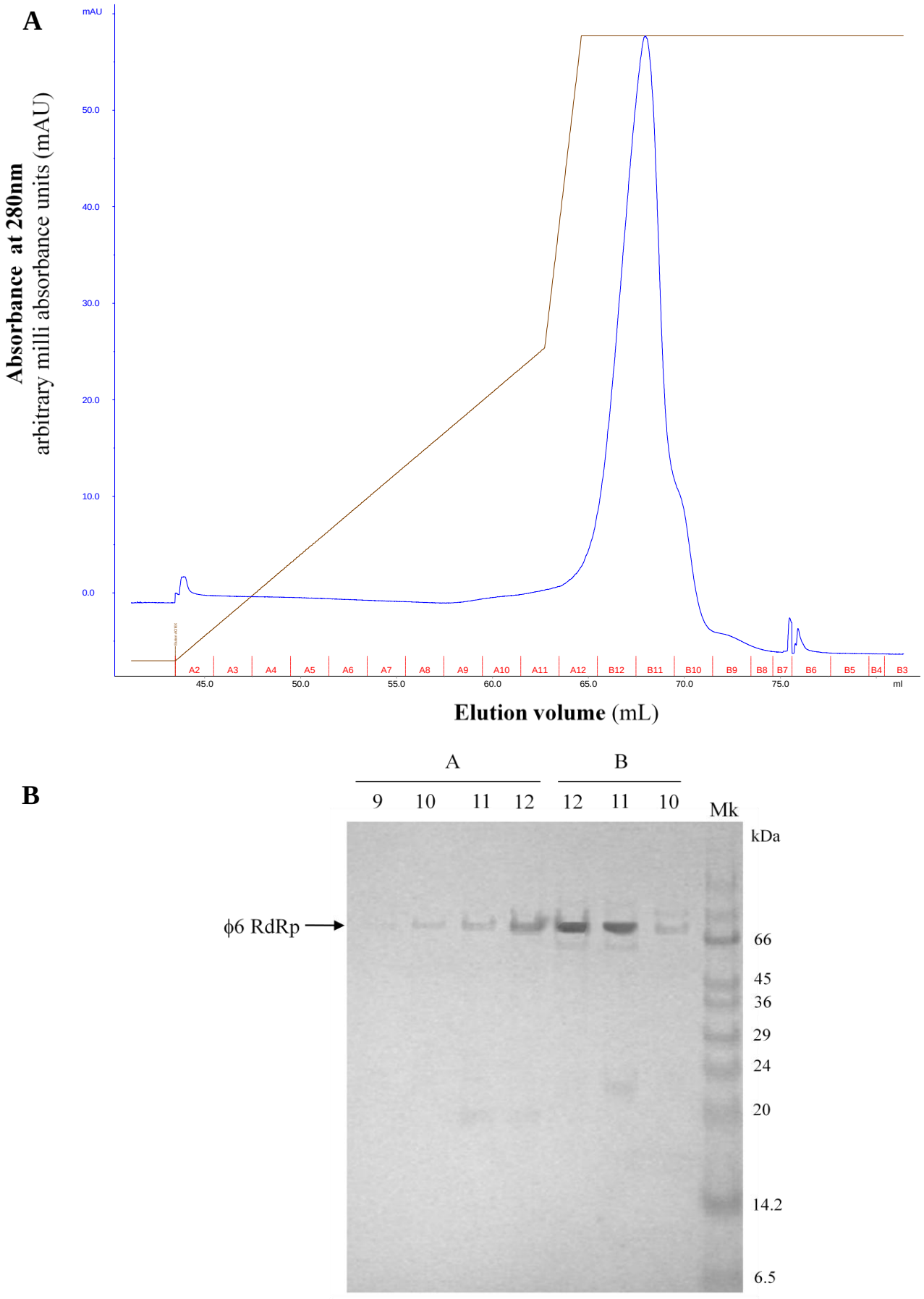

A
mAU
50.0
40.0
30.0
20.0
10.0
0.0
Absorbance at 280nm
arbitrary milli absorbance units (mAU)
A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 B12 B11 B10 B9 B8 B7 B6 B5 B4 B3
45.0 50.0 55.0 60.0 65.0 70.0 75.0 ml
Elution volume (mL)
B
A
B
9 10 11 12 12 11 10
Mk
kDa
ϕ6 RdRp
66
45
36
29
24
20
14.2
6.5

Figure 2.3.

2.1.4. SDS-PAGE

The NuPAGE® Novex Midi gel system (Invitrogen) was used to monitor the purification of WT or E634Q polymerase throughout the process. For each sample analysed, 10 µL was added to a 96-well PCR plate containing 10 µL of SDS loading buffer (2X stock, 100 mM Tris pH 6.4, 4% w/v SDS, 0.2 % w/v Bromophenol blue, 20 % v/v glycerol). After incubation at 95^{o}C for 5 min, 10 µL of each sample were loaded onto pre-cast NuPAGE® Novex 10 % Bis-Tris Midi Gels (Invitrogen). The gels were electrophoresed in 1X MES [2-(N-morpholino)-ethane sulphonic acid] buffer and stained with SimplyBlue™ SafeStain following the manufacturer's instructions (Invitrogen). The gels were visualized and documented using the Gene Genius bioimaging system (Syngene).

2.1.5. Protein Concentration

Fractions containing pure protein as determined by SDS-PAGE were pooled and typically concentrated to 10 mg/mL by centrifugation in VivaSpin concentrators (VivaScience). A Bio-Rad Micro Bio-Spin 6 Chromatography Column was then used to exchange the protein buffer to 10 mM Tris-HCl (pH 8.0), 100 mM NaCl. The protein concentration was monitored by measurement of the UV absorbance at 280 nm from a 2 µL sample using a NanoDrop 1000 (Thermo Scientific) spectrophotometer using the predicted extinction coefficient of the protein. For storage, concentrated fractions were diluted 1:4 with storage buffer (62.5% glycerol, 100 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0,125% Triton-X 100) and stored at -20 °C. A sample of the concentrated protein was also taken for QA using liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS).

2.1.6. Serine Protease Digestion of WT ϕ6 RdRp to form $P2^{T1}$

This is experiment was carried out by Peter Sarin at the University of Helsinki. Protease digestion of WT ϕ6 RdRp (produced as in section 2.1.2 and 2.1.3) was carried out by incubating 1.0 mg/mL purified WT protein with 0.5 µg/mL Trypsin Gold (mass spectrometry grade, Promega) in reaction buffer T (50mM Tris-HCl pH 8.0, 1mM $CaCl_2$) at 23°C for 16 h. The protease digestion was terminated by a 20-fold dilution with buffer T. The digestion mixture was filtered through a sterile 0.22 µm filtration unit (Millipore) and loaded onto a Hi-Trap™ Q Sepharose HP-column (Amersham Biosciences) equilibrated with QA-buffer (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA) and proteins eluted with a linear 0–1 M NaCl gradient buffered with 50 mM Tris-HCl pH 8.0 and 0.1 mM EDTA. The peak fractions were analysed by SDS-PAGE and fractions containing $P2^{T1}$ were pooled and sent to Oxford for crystallization trials.

2.2. Thermal denaturation assay

Thermal denaturation assays (ThermoFluor) were carried out in a real-time PCR machine (BioRad DNA Engine Opticon 2) where buffered solutions of protein and fluorophore (SYPRO Orange; Molecular Probes, Invitrogen), with and without additives, were heated in a stepwise fashion from 25°C to 99°C. 5 µL protein (1 mg/mL) and 5 µL SYPRO Orange (Molecular Probes, Invitrogen) were made up to a total assay volume of 50 µL with reference buffer (50 mM Tris–HCl, pH 8.0, 50 mM NaCl) in white low profile thin-wall PCR plates (Abgene) sealed with microseal 'B' films (BioRad). The fluorophore was excited in the range of 470–505 nm and fluorescence emission was measured in the range of 540–700 nm every 0.5°C after a 10 s hold. All experiments were carried out in triplicate. GraphPad Prism software and the Boltzmann Distribution function was used to calculate the inflection point or

midpoint of the unfolding transition [herein denoted the melting temperature (T_m)], of the sigmoidal melting curves; a standard error for T_m values was also calculated. The effect of Mg^{2+} and Mn^{2+} ions on the thermal stability of WT ϕ6 RdRP, E634Q and E491Q mutants, was assayed by comparing melting temperatures of protein in reference buffer with those containing the following additives: 0.1 mM, 1 mM and 5 mM EDTA, 5 mM EDTA with 10 mM $MgCl_2$ or $MnCl_2$, and finally 10 mM $MgCl_2$ or $MnCl_2$ without EDTA.

2.3. Protein crystallization

The optimal crystallisation conditions of ϕ6 RdRp had been determined previously by (Butcher *et al*., 2001) and similar conditions were used in all experiments. Typically, 100 nL of precipitant solution [50-150 mM HEPES-Na (pH 6.9-7.6), 10-13% (v/v) PEG 20K, 2% (v/v) ethylene glycol, 2 mM $MnCl_2$] was mixed with 100 nL protein (10 mg/mL) in a 96-well format using the sitting drop vapour diffusion method at room temperature, prepared using a Cartesian Technologies pipetting system available at the Oxford Protein Production Facility (Walter *et al*., 2005). Initial hits were subjected to a 3-row optimisation (varying the drop size, ratio of protein to precipitant, and pH) and/or additive screen (introducing 96 different additives to the initial hit); see Walter *et al*., (2005).

2.3.1. WT ϕ6 RdRp crystals

For WT ϕ6 RdRp, optimization around the initial hits (Walter *et al*., 2005) resulted in crystals up to 200x110x80 μm in 20% PEG 4k, 100 mM HEPES-Na (pH 7.5), 8.5% isopropanol, 15% glycerol, 2 mM $MnCl_2$.

2.3.1.1. WT ϕ6 RdRp soaking experiments (WT-Tri4T-ATP-Mg^{2+}, WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}, WT-5'-AATCT-3'-GTP-Mg^{2+})

WT ϕ6 RdRp crystals were used for soaking experiments with DNA oligonucleotide, NTPs and $MgCl_2$. DNA oligonucleotide soaked into the crystal was either a short hairpin (Operon), 5'-TTTTCGCGTAGCG-3' (Tri4T) designed to mimic the trinucleotide base-paired step with a four nucleotide 3'-overhang (underlined nucleotides base pair to the hairpin), or a short linear oligonucleotide mimic of the 3'-end of the positive-sense strand of ϕ6 genomic RNA used during replication (5'-AATCT-3'). Typically, 0.5 μL of soak solution containing 10 mM $MgCl_2$ was added first to the crystal drop, followed a few seconds later with 1 μL of soak solution containing 200 μM DNA (hairpin or linear), 25 mM ATP and/or 25 mM GTP (both lithium salts, Sigma) and 20% (v/v) glycerol. The solutions were left to diffuse through the crystal for 1-10 min or until cracking was observed.

2.3.1.2. Trypsin-digested WT ϕ6 RdRp ($P2^{T1}$) crystals

For trypsin-digested $P2^{T1}$ (cleaved at residue 607), optimisation around the initial hits resulted in crystals up to 200x70x40 μm in 20% PEG 6k, 100 mM HEPES (pH 7.0) using a drop size ratio of 1:2 with 100 nL protein to 200 nL precipitant condition. Crystals were swept through a drop containing 20% (v/v) glycerol (as cryo-protectant) made up in mother liquor, prior to freezing and subsequent data collection.

2.3.2. E634Q mutant crystals

For the E634Q mutant, optimization around the initial hits (Walter *et al*., 2005) resulted in crystals up to 400x200x100 μm in 12% PEG 20k, 100 mM MES (pH 6.5). E634Q apoenzyme crystals (no soaks) were swept through a drop containing 20% (v/v) glycerol (as cryo-protectant) made up in mother liquor, prior to freezing and subsequent data collection.

2.3.2.1. E634Q hairpin soaking (E634Q-hairpin-AMPPCP-Mg^{2+})

ϕ6 E634Q mutant crystals were used for soaking experiments with one of four different DNA hairpins, β,γ-methyleneadenosine 5'-triphosphate (AMPPCP, an ATP analog) and $MgCl_2$. The online Integrated DNA Technologies SciTools OligoAnalyzer 3.1 with hairpin function was used to design DNA hairpins with high melting temperatures that mimic the trinucleotide base-paired step of the polymerisation reaction. The hairpins obtained from Operon were: Tetra2T (5'-TTCGCGTAAGCG-3') and Tetra4T (5'-TTTTCGCGTAAGCG-3') with hairpin melting temperatures of 72.8^{o}C, and Tri2T (5'-TTCGCGTAGCG-3') and Tri4T (5'-TTTTCGCGTAGCG-3') with hairpin melting temperatures of 72.1^{o}C (underlined residues base pair to form short hairpins with a two or four nucleotide 5'-overhangs, as described in section 2.3.1.1). Typically, 0.5 μL of soak solution containing 10 mM $MgCl_2$ was added first to the crystal drop, followed a few seconds later with 1 μL of a soak solution containing 200 μM DNA hairpin, 25 mM AMPPCP (lithium salt, Sigma) and 20% (v/v) glycerol. The solution was left to diffuse through the crystal prior to data collection for 1-10 min or until cracking was observed.

2.3.2.2. E634Q soaking experiments with DNA or RNA negative-sense strand mimics (E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} and E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+})

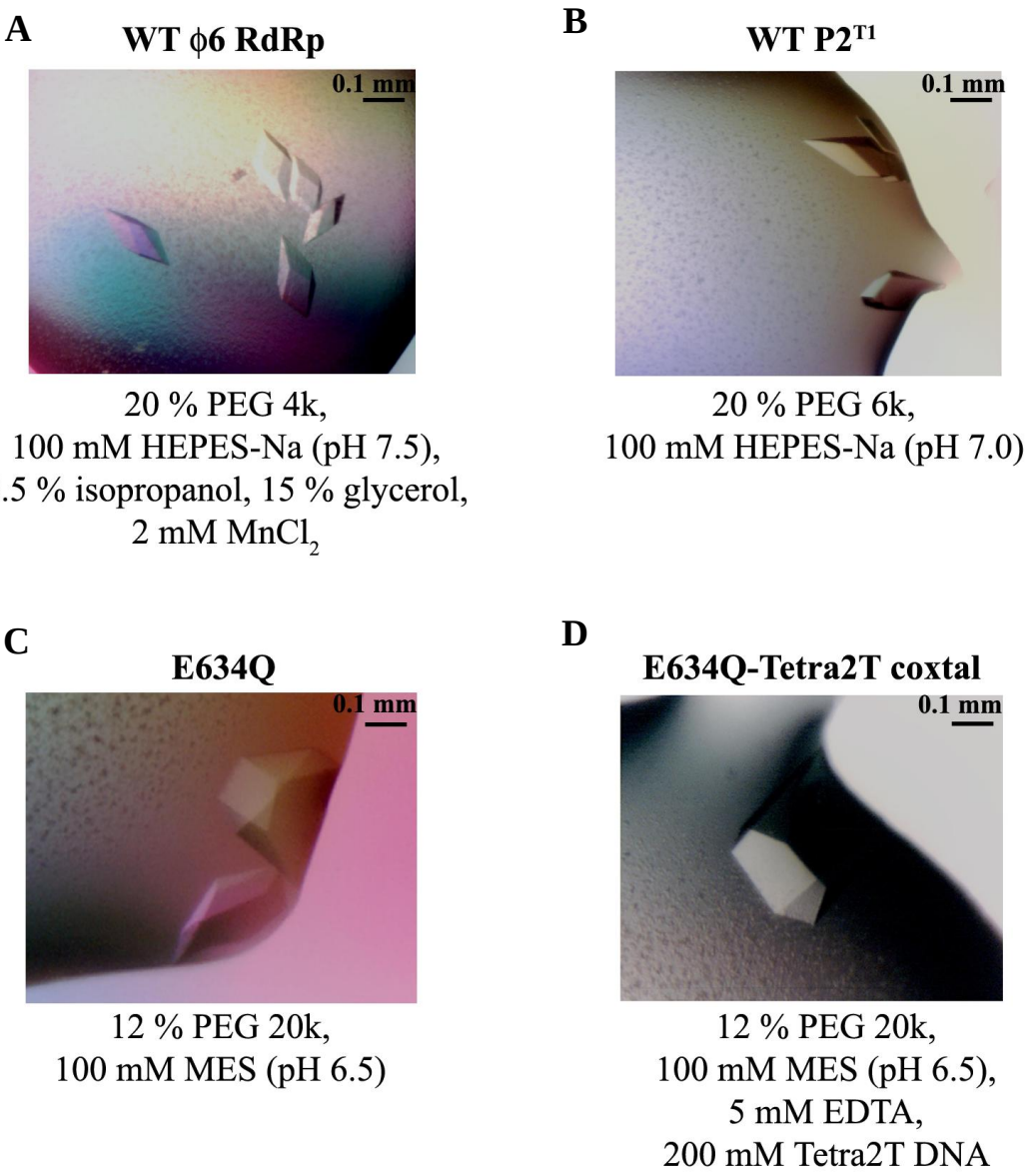
ϕ6 E634Q mutant crystals were used for soaking experiments with linear RNA and DNA 6mer oligonucleotides (5'-UUCCCC-3' and 5'-TTCCCC-3' respectively) that mimic the 3'-end of ϕ6 genomic RNA used in transcription, alongside GTP and divalent cations. For E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}, 0.5 μL of soak solution containing 10 mM $MgCl_2$ was added first to the crystal drop, followed a few seconds later with 1 μL of a soak solution containing 200 μM 5'-UUCCCC-3' RNA, 25 mM GTP (lithium salt, Sigma) and 20% (v/v) glycerol. For E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+}, 0.5 μL of soak solution containing 10 mM $MgCl_2$ and 5 mM $MnCl_2$ was added first to the crystal drop, followed a few seconds later with 1 μL of a soak solution containing 200 μM 5'-TTCCCC-3' DNA, 25 mM GTP (lithium salt, Sigma) and 20% (v/v) glycerol. The solution was left to diffuse through the crystal prior to data collection for 1-10 min or until cracking was observed.

2.3.2.3. E634Q hairpin co-crystallisation (E634Q-Tetra2T-coxtal)

E634Q mutant co-crystals were grown in the presence of 200 μM DNA Tetra2T hairpin (5'-TTCGCGTAAGCG-3') and 5 mM EDTA. Optimization around the initial hits (Walter *et al*., 2005) resulted in crystals up to 300x200x100 μm in 12% PEG 20k, 100 mM MES (pH 6.5), 5 mM EDTA; crystals did not grow in the absence of EDTA. A volume of 1 μL soak solution [containing 200 μM DNA and 20 % (v/v) glycerol, made up in mother liquor] was added to the co-crystal drop and allowed to soak for 2 min prior to data collection.

Figure 2.4. Example of WT ϕ6 RdRp, E634Q and P2^{T1} cystals grown.

The crystal conditions are indicated below each image and a scale bar shown to indicate size. **(A)** WT ϕ6 RdRp crystals grown after 20 h; **(B)** Trypsin-digested WT P2^{T1} crystals grown after 35 h; **(C)** E634Q crystals grown after 48 h; **(D)** E634Q-Tetra2T co-crystal grown after 72 h.



2.4. Data collection and processing

A summary of data collection and processing statistics for crystal structures presented in this thesis are given in **Table 2.2** for WT structures and **Table 2.3** for E634Q structures. Unless otherwise stated, all data were collected from crystals cooled in a stream of gaseous nitrogen at a temperature 100K after briefly washing the crystals with 20% (v/v) glycerol as cryoprotectant.

2.4.1. WT-Tri4T-ATP-Mg^{2+}, WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}, WT-5'-AATCT-3'-GTP-Mg^{2+}

X-ray diffraction data for WT ϕ6 RdRp complex structures were collected on station I03 at the Diamond Light Source, Oxfordshire, UK using an ADSC Q315 3x3 CCD detector. Each image covered an oscillation of 1^{o} and exposure times were typically 1-2 seconds. All data were collected from crystals at 100 K, after soaking the crystals for 1-10 min in the specified condition in the presence of 20% (v/v) glycerol as cryoprotectant. WT-Tri4T-ATP-Mg^{2+} and WT-5'-AATCT-3'-GTP-Mg^{2+} data were scaled and indexed using HKL-2000 suite of software (Otwinowski, 1997), but for WT-5'-AATCT-3'-ATP-GTP-Mg^{2+} the automated data reduction program XIA2 with the 3D option [using Labelit, XDS and XSCALE algorithms; (Kabsch, 1993)] was used. All WT datasets collected were approximately isomorphous to the WT crystals previously described (Butcher *et al*., 2001; Butcher *et al*., 2000; Salgado *et al*., 2004) and belong to space group $P2_1$ (with unit cell dimensions of a≈106Å, b≈92Å, c≈142Å, β≈102º).

2.4.2. Trypsin-digested WT $P2^{T1}$

X-ray diffraction data for $P2^{T1}$ were collected on station BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France using a MAR CCD detector. Each image covered an oscillation of 1^{o} and exposure times were typically 45-60 seconds. Data were processed using the automated data reduction program XIA2 with the 3D option [using Labelit, XDS and XSCALE algorithms; (Kabsch, 1993)]. The dataset collected for $P2^{T1}$ was approximately isomorphous to the $P3_2$ SeMet crystal structure determined by Butcher *et al*., (2001) and Poranen *et al*., (2008b) with unit cell dimensions a≈b≈109Å, c≈159Å.

2.4.3. E634Q apoenzyme and E634Q-hairpin-AMPPCP-Mg^{2+} complexes

X-ray diffraction data for E634Q apoenzyme and the different hairpin complex structures were collected on station ID14-EH2 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France using an ADSC Q4R CCD detector. Each image covered an oscillation of 1^{o} and exposure times were typically 45-60 seconds. All data were collected from crystals at 100 K after briefly washing the crystals with 20% (v/v) glycerol as cryoprotectant [for the soaks, 20% (v/v) glycerol was already present in the soaking condition as described previously]. Data were processed using the automated data reduction program XIA2 with the 3D option [using Labelit, XDS and XSCALE algorithms; (Kabsch, 1993)].

2.4.4. E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} and E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}-Mn^{2+} complexes

X-ray diffraction data for E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} and E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+} were collected on station I03 at the Diamond Light Source, Oxfordshire, UK using an ADSC Q315 3x3 CCD detector. Each image covered an oscillation of 1^{o} and exposure times were typically 1-2 seconds. All data were collected at liquid nitrogen temperatures after soaking the crystals for 1-10 min in the specified condition in the presence of 20% (v/v) glycerol as cryoprotectant. Both datasets were scaled and indexed using the HKL-2000 suite of software (Otwinowski, 1997) and were approximately isomorphous to the WT crystals previously described (Butcher *et al*., 2001; Butcher *et al*., 2000; Salgado *et al*., 2004) and belong to space group $P2_1$ (with cell dimensions of a≈106Å, b≈92Å, c≈142Å, β≈102º).

2.4.5. E634Q co-crystallisations with hairpin oligonucleotide (E634Q-Tetra2T-coxtal)

X-ray diffraction data for E634Q-Tetra2T-coxtal were collected on station I03 at the Diamond Light Source, Oxfordshire, UK using an ADSC Q315 3x3 CCD detector. Each image covered an oscillation of 1^{o} and exposure times were typically 1-2 seconds. All data were collected at liquid nitrogen temperatures after soaking the crystals for 1-10 min in the specified condition in the presence of 20% (v/v) glycerol as cryoprotectant. E634Q-Tetra2T-coxtal data were scaled and indexed using the HKL-2000 suite of software (Otwinowski, 1997) and was approximately isomorphous to the WT crystals previously described (Butcher *et al*., 2001; Butcher *et al*., 2000; Salgado *et al*., 2004) and belong to space group $P2_1$ (with cell dimensions of a≈106Å, b≈92Å, c≈142Å, β≈102º).

2.5. Structure determination and refinement

ϕ6 RdRp, including different complexes and mutants, crystallise in one of two space groups:

A – $P2_1$ with unit cell dimensions a≈106Å, b≈92Å, c≈142Å, β≈102^o

B – $P3_2$ with unit cell dimensions a≈ b≈109Å, c≈159Å

Most of the crystal structures described in this chapter belong to the A crystal form and were isomorphous with crystal structures determined previously by Butcher *et al*., (2001), Laurila *et al*., (2005) and Salgado *et al*., (2004). Rigid body refinement of monomers in the asymmetric unit (ASU) was sufficient to obtain initial phases of these structures. The structures that belong instead to crystal forms B were solved by molecular replacement as described below. Final refinement statistics for all WT and E634Q structures are summarised in **Table 2.2** and **Table 2.3** respectively.

Table 2.2. Data collection, processing and refinement statistics for WT complexes including soaking conditions.

	WT-$^{5'}$AATCT$^{3'}$-GTP-Mg^{2+}	WT-$^{5'}$AATCT$^{3'}$-ATP-GTP-Mg^{2+}	WT-Tri4T-ATP-Mg^{2+}	Trypsin-digested WT $P2^{T1}$
Soaking Conditions				
DNA oligonucleotide	0.2 mM $^{5'}$AATCT$^{3'}$	0.2 mM $^{5'}$AATCT$^{3'}$	0.2 mM Tri4T	-
NTP [a]	25 mM GTP	25 mM ATP	25 mM ATP	-
	-	25 mM GTP	-	-
$MgCl_2$	10 mM	10 mM	10 mM	-
$MnCl_2$	-	-	-	-
EDTA	-	-	-	-
Data Collection				
X-ray source	Diamond I03	Diamond I03	Diamond I03	ESRF BM14
Wavelength (Å)	0.978	0.978	0.978	0.978
Space group	$P2_1$	$P2_1$	$P2_1$	$P3_2$
Unit cell [a,b,c (Å); $\alpha = \gamma = 90^o$; β (^{o})]	106.5, 93.0, 140.8; 101.4	106.3, 92.1, 140.8; 101.3	105.9, 91.8, 140.5; 101.3	105.1, 105.1, 157.5
Resolution range (Å)[b]	50.0−3.5 (3.6−3.5)	58.7−2.9 (3.0−2.9)	50.0−3.3 (3.4−3.3)	34.4−3.3 (3.4−3.3)
Observations	258,004	201,019	302,147	138,515
Unique reflections	34,154	55,288	39,855	28,332
Completeness (%)[b]	100.0 (100.0)	95.2 (95.5)	100.0 (100.0)	96.5 (87.0)
I/σI (%)[b]	4.9 (2.0)	11.7 (1.6)	6.3 (1.7)	6.7 (1.6)
R_{merge} (%)[c]	32.5	9.1	26.7	23.8
Refinement Statistics				
Resolution range (Å)	46.5−3.5	46.1−2.9	46.0−3.3	29.7−3.3
No. of reflections (working/test)	29,302/1,718	50,625/2,553	34,770/1,769	25,572/1,301
R_{work}[d]/R_{free}[e]	19.9/29.4	17.9/27.6	19.6/30.4	23.3/30.6
No. of atoms: (protein/ligand/water)	15,795/235/0	15,795/282/0	15,795/308/0	15,644/0/6
RMSD bond length (Å)	0.008	0.008	0.008	0.008
RMSD bond angle (^{o})	1.3	1.2	1.3	1.2
Mean B-factor (Å^2)	96.4	62.7	77.4	49.8

[a] β,γ-methyleneadenosine 5'-triphosphate (AMPPCP), adenosine triphosphate (ATP) and guanosine triphosphate (GTP) were all lithium salts from Sigma.
[b] Values in parenthesis refer to the highest resolution shell.
[c] $R_{merge} = \Sigma_i\Sigma_h(|I_{j,h} - \langle I_h \rangle|) / \Sigma_j\Sigma_h(\langle I_h \rangle)$, where h are unique reflections indices, $I_{j,h}$ are intensities of symmetry-related reflections and $\langle I_h \rangle$ is the mean intensity
[d] R_{work} and [e] R_{free} are defined by $R=\Sigma_{hkl}||F_{obs}|- |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where *h,k,l* are the indices of the reflections (used in refinement for R_{work}; 5%, not used in refinement, for R_{free}), F_{obs} and F_{calc} are the structure factors, deduced from measured intensities and calculated from the model, respectively.

Table 2.3. Data collection, processing and refinement statistics for E634Qapo and E634Q complexes including soaking conditions.

	E634Q	E634Q-Tri2T-AMPPCP-Mg^{2+}	E634Q-Tri4T-AMPPCP-Mg^{2+}	E634Q-Tetra2T-AMPPCP-Mg^{2+}	E634Q-Tetra4T-AMPPCP-Mg^{2+}	E634Q-Tetra2T-coxtal	E634Q-$^{5'}$UUCCCC$^{3'}$-GTP-Mg^{2+}	E634Q-$^{5'}$TTCCCC$^{3'}$-GTP-Mg^{2+}
Soaking Conditions								
DNA oligonucleotide	-	0.2 mM Tri2T	0.2 mM Tri4T	0.2 mM Tetra2T	0.2 mM Tetra4T	0.2 mM Tetra2T	0.2 mM UUCCCC	0.2 mM TTCCCC
NTP [a]	-	25 mM AMPPCP	25 mM AMPPCP	25 mM AMPPCP	25 mM AMPPCP	-	25 mM GTP	25 mM GTP
$MgCl_2$	-	10 mM	10 mM	10 mM	10 mM	-	10 mM	10 mM
$MnCl_2$	-	-	-	-	-	-	-	5 mM
EDTA	-	-	-	-	-	5 mM	-	-
Data Collection								
X-ray source	ESRF ID14-EH2	ESRF ID14-EH2	ESRF ID14-EH2	ESRF ID14-EH2	ESRF ID14-EH2	Diamond I03	Diamond I03	Diamond I03
Wavelength (Å)	0.978	0.978	0.978	0.978	0.978	0.978	0.978	0.978
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P3_2$	$P2_1$	$P2_1$
Unit cell [a,b,c (Å); $\alpha = \gamma = 90^o$; β (^{o})]	106.2, 92.5, 141.0; 101.5	107.4, 92.9, 141.2; 101.9	107.5, 91.7, 142.1; 102.0	106.9, 91.4, 141.6; 101.9	105.8, 89.3. 141.0; 101.9	111.6, 111.6, 159.1	106.2, 93.1, 141.1; 101.1	105.1, 89.4, 140.5; 100.1
Resolution range (Å)[b]	69.1−2.7 (2.8−2.7)	77.2−3.0 (3.1−3.0)	76.5−2.9 (3.0−2.9)	104.6−3.0 (3.1−3.0)	67.6−3.4 (3.5−3.4)	50.0−3.1 (3.2−3.1)	50.0−3.4 (3.5−3.4)	50.0−3.8 (4.0−3.8)
Observations	553,323	377,954	318,362	528,644	189,936	232,464	142,977	90,454
Unique reflections	74,265	50,754	37,930	51,249	30,875	80,438	70,560	47,275
Completeness (%)[b]	97.5 (94.0)	98.4 (96.9)	98.5 (96.4)	99.1 (98.1)	86.8 (61.6)	100.0 (100.0)	95.6 (94.3)	99.3 (99.1)
I/σI (%)[b]	17.0 (2.3)	13.3 (2.0)	12.1 (2.0)	14.7 (2.4)	9.7 (2.1)	8.4 (1.4)	5.1 (1.2)	4.7 (1.4)
R_{merge} (%)[C]	9.2	18.1	19.1	17.5	17.0	21.1	25.3	26.8
Refinement Statistics								
Resolution range (Å)	69.1−2.7	69.6−3.0	64.5−2.9	104.6−3.0	67.6−3.4	48.3−3.1	46.5−3.4	45.6−3.8
No. of reflections (working/test)	67,815/3,432	46,590/2,371	45,681/2,320	46,879/2,384	26,049/1,320	40,167/2,009	33,064/1,751	20,782/1,667
R_{work}[d]/R_{free}[e]	18.5/25.3	19.0/28.0	21.0/31.9	19.7/28.4	21.0/34.0	20.1/30.8	20.1/31.1	21.4/34.8
No. of atoms: (protein/ligand/water)	15,795/3/0	15,672/253/1	15,690/294/1	15,631/294/1	15,615/274/1	15795/85/0	15,795/251/0	15,795/278/0
RMSD bond length (Å)	0.008	0.009	0.008	0.008	0.009	0.007	0.009	0.009
RMSD bond angle (^{o})	1.2	1.3	1.3	1.3	1.5	1.1	1.4	1.5
Mean B-factor (Å^2)	57.0	74.6	83.3	74.2	111.4	64.3	54.8	72.9

[a] β,γ-methyleneadenosine 5'-triphosphate (AMPPCP), adenosine triphosphate (ATP) and guanosine triphosphate (GTP) were all lithium salts from Sigma.
[b] Values in parenthesis refer to the highest resolution shell.
[c] $R_{merge} = \Sigma_i\Sigma_h(|I_{j,h} - \langle I_h \rangle|) / \Sigma_j\Sigma_h(\langle I_h \rangle)$, where h are unique reflections indices, $I_{j,h}$ are intensities of symmetry-related reflections and $\langle I_h \rangle$ is the mean intensity
[d] R_{work} and [e] R_{free} are defined by $R = \Sigma_{hkl}||F_{obs}| - |F_{calc}|| / \Sigma_{hkl}|F_{obs}|$, where *h,k,l* are indices of the reflections (used in refinement for R_{work}; 5%, not used in refinement for R_{free}), F_{obs} (deduced from measured intensities) and F_{calc} (calculated from the model) are the structure factors.

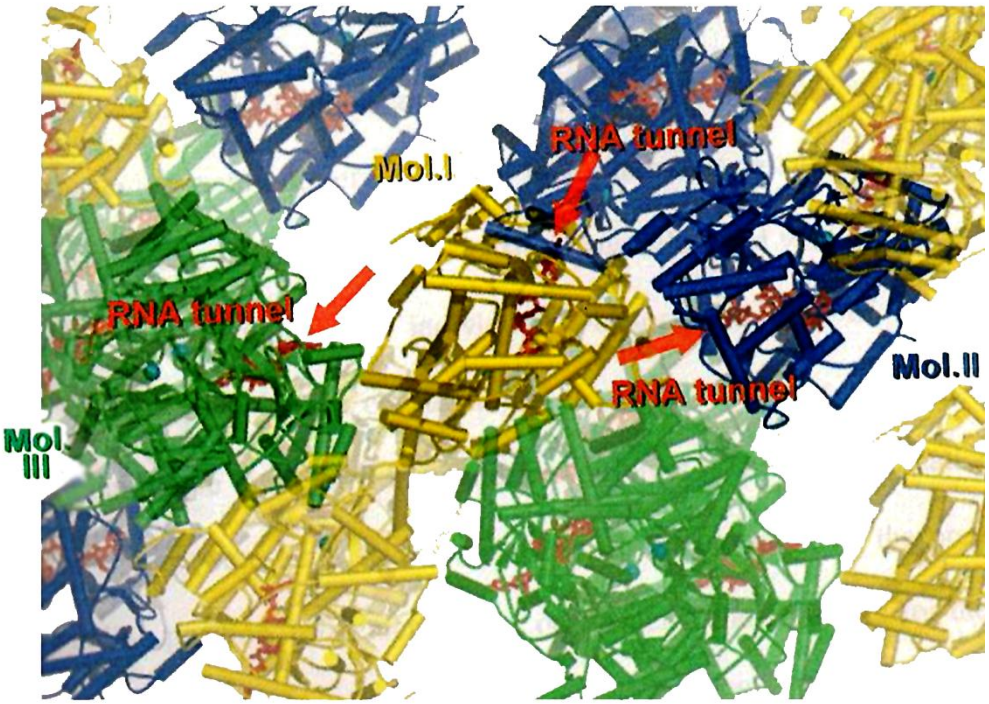
2.5.1. WT ϕ6 RdRp complexes

For WT complexes with DNA, NTPs and Mg^{2+}, initial models were obtained by rigid body refinement with PHENIX (Adams *et al*., 2002) using the WT apo $P2_1$ structure (Butcher *et al*., 2001) as a starting model (PDB accession no. 1HHS). Visual inspection of the resulting SIGMAA |Fo|-|Fc| and 2|Fo|-|Fc| difference-Fourier maps allowed identification of ligands bound in all three molecules of the ASU (see below). Manual modelling was carried out in COOT (Emsley and Cowtan, 2004) and the structures refined in PHENIX (Adams *et al*., 2002) imposing three-fold non-crystallographic symmetry (NCS) restraints for the different protein sub-domains [palm, fingers, thumb and CTD, as defined by Butcher *et al*., (2001)].

Importantly, the three molecules in the ASU of the WT ϕ6 RdRp complexes (WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}, WT-Tri4T-ATP-Mg^{2+} and WT-5'-AATCT-3'-GTP-Mg^{2+} crystals) exhibit different features to one another. In terms of ligands bound and occupancy, molecules I and II that are stabilized by more crystallographic contacts tend to exhibit similar features to each other, whereas molecule III that has fewer crystal contacts exhibits a much greater molecular flexibility. For example, the template tunnel for molecule III is less obscured by neighbouring molecules than the template tunnels of molecules I and II (see **Fig. 2.5.**). However, for the WT complexes described here molecule II exhibits different features to molecules I and III. Crystal packing about the template tunnel of molecule II enabled trapping of an incoming oligonucleotide sandwiched between the entrance to the tunnel and a nearby molecule III helix, which was not detected at other positions (see Chapter 3). Differences in crystal packing across the three molecules can thereby provide distinct molecular snapshots of ϕ6 RdRp in complex with various ligands.

Figure 2.5. Crystal contacts in *P*2$_1$ space group.

Cartoon representation of the crystallographic packing of ϕ6 RdRp co-crysallised with a 6nt RNA oligonucleotide (shown in sticks, red). The three molecules of the ASU are represented in different colours: molecules I (yellow), molecule II (blue) and molecule III (green). Crystallographically related molecules are shown in transparent mode with equivalent colours. Cyan spheres indicate Mn^{2+} ions. The entrance to the template tunnel is marked by a red arrow and labelled. Note the more accessible entrance in molecule III due to less tight crystallographic contacts surrounding these molecules. [Modified from (Salgado, 2005)].

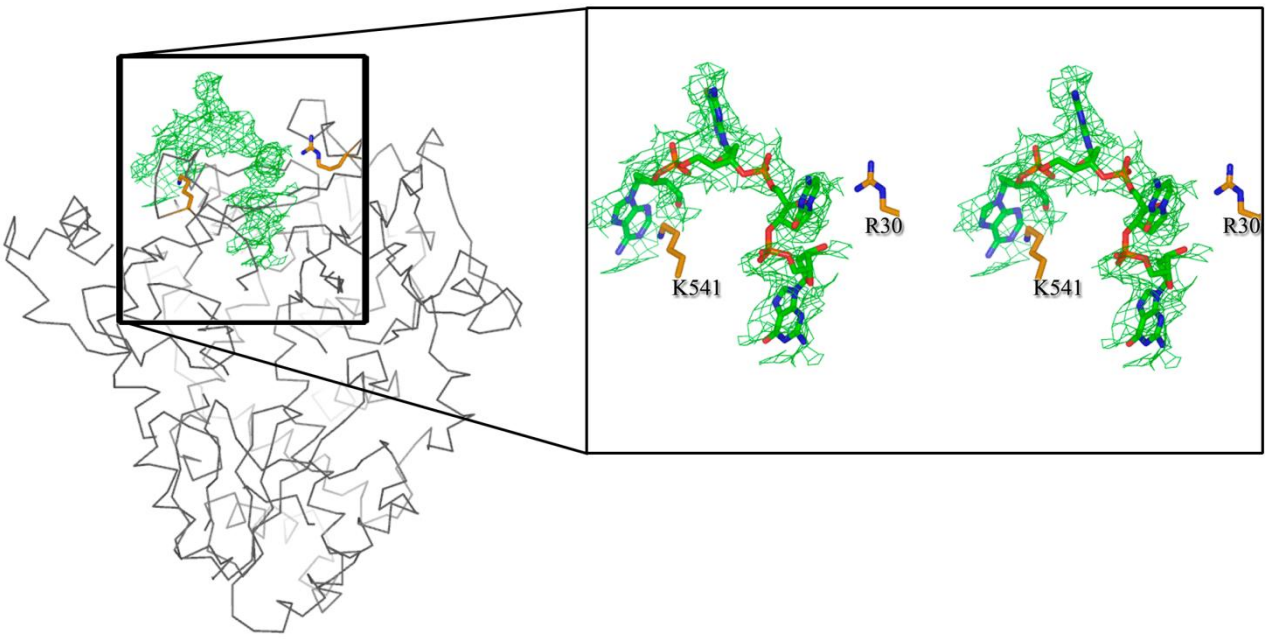


2.5.1.1. WT-Tri4T-ATP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of two ATP molecules at the active site of three molecules. In molecule I, only the triphosphates of the second ATP were built due to weak density and disorder of the sugar and base moieties. In molecules I and III, a Mg^{2+} could be built at the non-catalytic ion site. In molecule II, a Mg^{2+} could be built between the two phosphate backbones of the bound ATPs, and additionally four nucleotides of Tri4T DNA hairpin (5'-AATG-3') could be fitted at the template entrance reaching over a positively charged groove on the outer surface of the polymerase stabilised by the crystal packing of nearby molecule III α-helix. Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. In addition, simulated annealing Fo-Fc omit difference maps were generated in PHENIX (Adams *et al*., 2002) with the ligands omitted from the model in order to reduce model bias and confirm the presence and position of bound ligands. An example simulated annealing Fo-Fc omit difference map showing the difference density for bound oligonucleotide at the entrance to molecule II is shown in **Fig. 2.6**. Final refinement in PHENIX to a resolution of 3.4Å imposing NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=19.6% (R_{free}= 30.4%) and good stereochemistry [root mean square deviation (RMSD) bond length=0.008Å; RMSD bond angle=1.3^{o}].

Figure 2.6. Simulated annealing Fo-Fc omit difference map for WT-Tri4T-ATP-Mg^{2+} (molecule II)

Left panel: 3.4Å simulated annealing Fo-Fc omit difference map contoured to 1.5σ (green wire) for WT-Tri4T-ATP-Mg^{2+} molecule II (protein backbone depicted by grey ribbon) at the entrance to the template tunnel. Ligands were omitted from the model during the simulated annealing in order to reduce model bias and confirm the presence and position of oligonucleotide density at the tunnel entrance. ***Right Panel:*** Stereo close-ups of simulated annealing Fo-Fc omit difference density contoured to 1.5σ (green wire) with the oligonucleotide (5'-AATG-3', left to right) traced into the density (stick representation: carbon, green; phosphorous, orange; oxgen, red; nitrogen, blue). In addition, residues R30 and K541 that are essential for efficient template binding (Sarin *et al*., 2009) are shown for completeness (stick representation: carbon, orange; nitrogen, blue).



2.5.1.2. WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of GTP and the triphosphate backbone of a second NTP at the active site in all three molecules. A Mg^{2+} ion could be built at the non-catalytic ion site in all molecules. Part of the 5'-AATCT-3' DNA soaked into the crystal could be fitted at the entrance to the template tunnel in the same position seen for WT-Tri4T-ATP-Mg^{2+}. Due to local disorder of the more external nucleotides, only one or two nucleotides of DNA were built for molecules I and III respectively. For molecule II, four nucleotides of DNA (5'-AATC-3') could be built convincingly into the stronger positive difference density, stabilised by a nearby molecule III α-helix. Final refinement in PHENIX (Adams *et al*., 2002) to a resolution of 2.9Å imposing NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=17.9% (R_{free}=27.6%) and good stereochemistry [root mean square deviation (RMSD) bond length=0.008Å; RMSD bond angle=1.2^{o}].

2.5.1.3. WT-5'-AATCT-3'-GTP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of a single GTP in the NTP tunnel and Mg^{2+} ion at the non-catalytic ion site in all three molecules of the asymmetric unit. Part of the 5'-AATCT-3' DNA soaked into the crystal could be fitted at the entrance to the template tunnel WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}. No evidence for DNA was seen for molecule I and only two nucleotides of DNA were built for molecule III (5'-TC-3'), presumably due to local disorder of the more external nucleotides. For molecule II, four nucleotides of DNA (5'-AATC-3') could be built into the stronger positive difference density, stabilised by a nearby molecule III α-helix; density was too weak to fit the 3'-end thymidine of the 5nt template. Final

refinement in PHENIX (Adams *et al*., 2002) to a resolution of 3.7Å imposing NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=19.9% (R_{free}=29.4%) and good stereochemistry [root mean square deviation (RMSD) bond length=0.008Å; RMSD bond angle=1.3^{o}].

2.5.1.4 Trypsin-digested WT ϕ6 RdRp ($P2^{T1}$)

The $P2^{T1}$ structure was approximately isomorphous with the $P3_2$ SeMet WT structure and E491 to Q mutant structures previously determined (Butcher *et al*., 2001; Salgado *et al*., 2004). An initial model was obtained by molecular replacement in Phaser (McCoy, 2007) with the higher resolution WT apo $P2_1$ structure (Butcher *et al*., 2001) as a starting model (PDB accession no. 1HHS) followed by one round of rigid body refinement in PHENIX (Adams *et al*., 2002) defining the three molecules of the ASU. Visual inspection of the initial model and manual rebuilding was carried out in COOT (Emsley and Cowtan, 2004). Surprisingly, the resulting electron density did not support the absence of CTD residues 607-664 in the cleaved $P2^{T1}$ protein. Residues 610-664 could be built unambiguously into the available density in all three monomers of the ASU, however in all monomers very poor electron density was observed at residue 607 about the trypsin-cleaved region (residues 603-609); these residues were removed for subsequent refinement steps. Final refinement in PHENIX (Adams *et al*., 2002) imposing three-fold NCS restraints on the invariant parts of the molecules, using data to a resolution of 3.3Å resulted in a model with R_{work}=23.3% (R_{free}=30.6%) and good stereochemistry (RMSD bond length = 0.008Å and RMSD bond angle = 1.2°).

2.5.2. E634Q mutant apoenzyme and complexes from soaking experiments

For the E634Q apoenzyme and complexes from soaking in DNA hairpin, ATP analog (AMPPCP) and Mg^{2+}, initial models were obtained by rigid body refinement with PHENIX (Adams *et al*., 2002) using the WT apo $P2_1$ structure (Butcher *et al*., 2001) as a starting model (PDB accession no. 1HHS). Visual inspection of the resulting SIGMAA |Fo|-|Fc| and 2|Fo|-|Fc| difference-Fourier maps allowed identification of any ligands bound in all three molecules of the asymmetric unit. Manual modelling was carried out in COOT (Emsley and Cowtan, 2004) and the structures refined in PHENIX (Adams *et al*., 2002) imposing three-fold NCS restraints for the different protein sub-domains [palm, fingers, thumb and CTD, as defined by Butcher *et al*., (2001)]. As described in section 2.5.1 **(Fig. 2.5.)**, molecule III that has fewer crystal contacts compared to molecules I and II, and exhibits a much greater molecular flexibility. For the E634Q hairpin complexes described here molecule III does exhibit different features to molecules I and II reflected by differences in the strength and position of the electron density in the template and NTP binding regions. Ligands observed in each molecule of the ASU were therefore modelled and refined separately. Detailed refinement statistics for E634Q crystal structures can be found in **Table 2.3**.

2.5.2.1. E634Q apoenzyme

Visual inspection of the initial model allowed unambiguous modelling of Mn^{2+} into the significant positive difference density at the non-catalytic ion site for all three monomers. The structure of E634Q apo was essentially indistinguishable from the previously published structure for WT ϕ6 RdRp (Butcher *et al*., 2001), other than the lack of the salt-bridge between K144 and E634, resulting in an increased flexibility of Q634 and weaker electron density for this residue in all three monomers. Final refinement in PHENIX (Adams *et al*.,

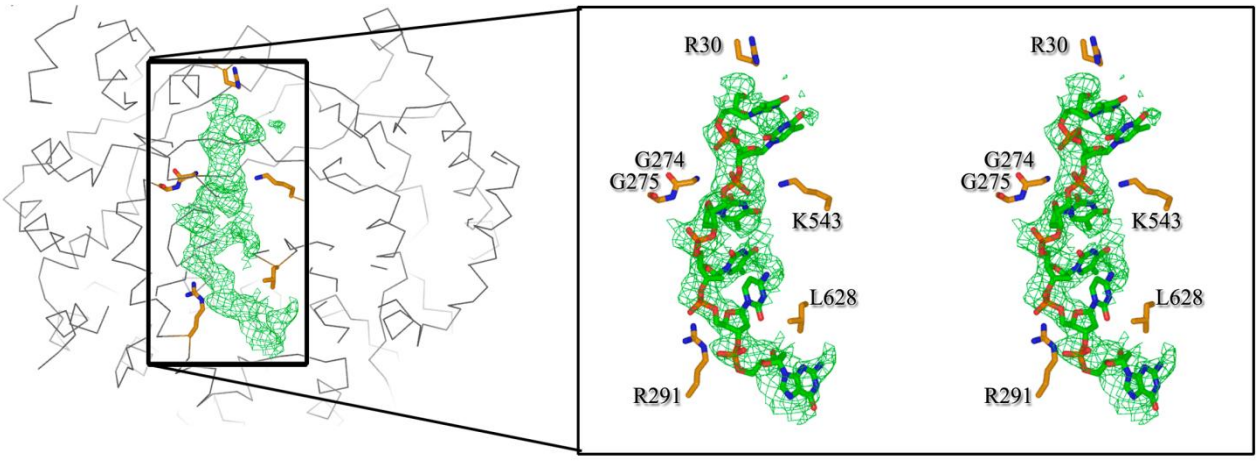
2002) using data to a resolution of 2.7Å, imposing three-fold NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=18.5% (R_{free}=25.3%) and good stereochemistry (RMSD bond length=0.008Å and RMSD bond angle=1.2°).

2.5.2.2. E634Q-Tri2T-AMPPCP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of the triphosphate backbone of AMPPCP at the NTP interrogation (site I, Chapter 1, section 1.1.3.4) bound to all three monomers of the ASU and a Mg^{2+} ion bound at the non-catalytic ion site only in molecules I and II. Part of the Tri2T hairpin (5'-TTCGCGTAGCG-3') template could be built in the template tunnel, with 5'-CG-3', 5'-GCG-3' and 5'-TTCGCG-3' modelled for molecules I, II and III respectively. Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. In molecule III, residues 606-611 and 631-641 (including mutated residue Q634) were highly disordered with no continuous peptide density for these regions in the |Fo|-|Fc| maps even at <0.6σ; these residues were removed for subsequent refinement. Simulated annealing Fo-Fc omit difference maps were generated in PHENIX (Adams *et al*., 2002) with the ligands and disordered regions omitted from the model in order to reduce model bias and confirm the presence and position of missing peptide regions and bound ligand. Simulated annealing did not improve the density in the poorly ordered regions (606-611 and 631-641), however difference density for oligonucleotide in the template tunnel was much improved. An example simulated annealing Fo-Fc omit difference map highlighting the difference density for oligonucleotide for molecule III is shown in **Fig. 2.7**. Final refinement in PHENIX using data to a resolution of 3.0Å imposing three-fold NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=19.0% (R_{free}=28.0%) and good stereochemistry (RMSD bond length=0.009Å and RMSD bond angle=1.3°).

Figure 2.7. Simulated annealing Fo-Fc omit difference map for E634Q-Tri2T-AMPPCP-Mg^{2+} (molecule III)

Left panel: 3.0Å simulated annealing Fo-Fc omit difference map contoured to 1.5σ (green wire) for E634Q-Tri2T-AMPPCP-Mg^{2+} molecule III (protein backbone depicted by grey ribbon). Ligands and questionable regions (high disordered C-terminal domain residues 606-611 and 631-641) were omitted from the model during the simulated annealing in order to reduce model bias and confirm the presence and position of oligonucleotide density within the template tunnel and beyond the active site. ***Right Panel:*** Stereo close-ups of simulated annealing Fo-Fc omit difference density for oligonucleotide contoured to 1.5σ (green wire); the oligonucleotide (5'-TTCGCG -3', top to bottom) has been traced convincingly into the density (stick representation: carbon, green; phosphorous, orange; oxgen, red; nitrogen, blue). In addition, residues R30, G274, G275, R291, K543 and L626 that line the template tunnel and have been shown to interact with bound oligonucleotide (Salgado *et al*., 2004) are shown for completeness (stick representation: carbon, orange; nitrogen, blue; oxygen, red).



2.5.2.3. E634Q-Tri4T-AMPPCP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of the triphosphate backbone of AMPPCP at the NTP interrogation site (I) bound to all three monomers of the ASU, and a Mg^{2+} ion bound at the non-catalytic ion site in molecules I and II. A second Mg^{2+} ion could be built in molecule I co-ordinated by the triphosphates of AMPPCP, the side chain of D324 and backbone carbonyl group of V325. Part of the Tri4T hairpin (5'-TTTTCGCGTAGCG-3') template could be built in the template tunnel, with 5'-GCG-3', 5'-AGCG-3' and 5'-TTTTCG-3' modelled for molecules I, II and III respectively. Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. In molecule III, residues 607-611 and 631-643 (including mutated residue Q634) were highly disordered with no continuous peptide density for these regions in the |Fo|-|Fc| maps even at <0.6σ; these residues were removed for subsequent refinements. Final refinement in PHENIX (Adams *et al*., 2002) using data to a resolution of 3.4Å imposing three-fold NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=21.0% (R_{free}=31.9%) and good stereochemistry (RMSD bond length=0.008Å and RMSD bond angle=1.3°).

2.5.2.4. E634Q-Tetra2T-AMPPCP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of the triphosphate backbone of AMPPCP at the NTP interrogation site (I) bound to all three monomers of the ASU and a Mg^{2+} ion bound at the non-catalytic ion site only in molecules I and II. A second Mg^{2+} ion could be built in molecule I co-ordinated by the triphosphates of AMPPCP, the side chain of D324 and backbone carbonyl group of V325. Part of the Tetra2T hairpin (5'-TTCGCGTAAGCG-3')

template could be built in the template tunnel, with 5'-CG-3', 5'-GCG-3' and 5'-TTCGCG-3' modelled for molecules I, II and III respectively. Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available density. In molecule III, residues 607-614 and 631-643 (including mutated residue Q634) were highly disordered with no continuous peptide density for these regions in the |Fo|-|Fc| maps even at <0.6σ; these residues were removed for subsequent refinement steps. Final refinement in PHENIX (Adams *et al*., 2002) using data to a resolution of 3.0Å imposing three-fold NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=19.7% (R_{free}=28.4%) and good stereochemistry (RMSD bond length=0.008Å and RMSD bond angle=1.3°).

2.5.2.5. E634Q-Tetra4T-AMPPCP-Mg^{2+}

Visual inspection of the SIGMAA 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of the triphosphate backbone of AMPPCP at the NTP interrogation site (I) bound to all three monomers of the ASU and a Mg^{2+} ion bound at the non-catalytic ion site only in molecules I and II. Part of the Tetra4T hairpin (5'-TTTTCGCGTAAGCG-3') could be built in the template tunnel, with 5'-CG-3', 5'-GCG-3' and 5'-TTTTCG-3' modelled for molecules I, II and III respectively. Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. In molecule III, residues 606-612 and 631-647 (including mutated residue Q634) are highly disordered with no continuous peptide density for these regions in the |Fo|-|Fc| maps even at <0.6σ; these residues were removed for subsequent refinement steps. Final refinement in PHENIX (Adams *et al*., 2002) using data to a resolution of 3.4Å imposing NCS restraints on the invariant parts of the molecules, resulted in a model with R_{work}=21.0% (R_{free}=34.0%) and good stereochemistry (RMSD bond length = 0.009Å and RMSD bond angle = 1.5°).

2.5.2.6. E634Q-UUCCCC-GTP-Mg^{2+}

Visual inspection of the SIGMAA difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of two GTPs and two Mg^{2+} ions about the active site in molecules I and II, and one GTP and one Mg^{2+} ion in molecule III. In addition, the two 3'-end nucleotides 5'-CC-3' of the RNA oligonucleotide (5'-UUCCCC-3') could be built convincingly at the entrance to the template tunnel of molecules II and III. Final refinement in PHENIX (Adams *et al*., 2002) using data to a resolution of 3.4Å resulted in a model with R_{work}=20.1% (R_{free}=31.1%) and good stereochemistry (RMSD bond length=0.009Å and RMSD bond angle=1.4°).

2.5.2.7. E634Q-TTCCCC-GTP-Mg^{2+}-Mn^{2+}

Visual inspection of the SIGMAA difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of one GTP at the interrogation site for molecules I and III, and one GTP and triphosphate backbone of a second GTP in molecule II. A Mn^{2+} ion could be built at the non-catalytic ion site of all three molecules, co-ordinated by D454, E491 and A495. In addition, part of the 3'-end of the DNA oligonucleotide (5'-TTCCCC-3') could be built convincingly at the entrance to the template tunnel in all three molecules: 5'-CC-3' for molecule II, and 5'-CCC-3' for molecules I and III. Final refinement in PHENIX (Adams *et al*., 2002) using data to a resolution of 3.8Å resulted in a model with R_{work}=21.4% (R_{free}=34.8%) and good stereochemistry (RMSD bond length=0.009Å and RMSD bond angle=1.5°).

2.5.3. E634Q co-crystallisation (E634Q-Tetra2T-coxtal)

E634Q crystal structures obtained from co-crystallisations with Tetra2T DNA hairpin were approximately isomorphous with the $P3_2$ SeMet WT structure and E491 to Q mutant

structures previously determined (Butcher *et al*., 2001; Butcher *et al*., 2000; Poranen *et al*., 2008b). For E634Q co-crystals, initial models were obtained by molecular replacement in Phaser (McCoy, 2007) with the higher resolution WT apo $P2_1$ structure (Butcher *et al*., 2001) as a starting model (PDB accession no. 1HHS) followed by one round of rigid body refinement in PHENIX (Adams *et al*., 2002) defining the three molecules of the asymmetric unit. Visual inspection of the resulting SIGMAA |Fo|-|Fc| and 2|Fo|-|Fc| difference-Fourier maps in COOT (Emsley and Cowtan, 2004) allowed modelling of part of the Tetra2T hairpin (5'-ACGC-3') into the template tunnel of all three molecules of the ASU. Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. Additional rounds of refinement in PHENIX (Adams *et al*., 2002) imposing strict three-fold NCS restraints for the different protein sub-domains [palm, fingers, thumb and CTD, as defined by Butcher *et al*., (2001)] did not significantly improve the template tunnel density. Final refinement for E634Q-Tetra2T-coxtal to a resolution of 3.1Å resulted in model with R_{work}=20.1% (R_{free}=30.8%) and good stereochemistry (RMSD bond length=0.007Å and RMSD bond angle=1.1°).

Chapter 3

ϕ6 RdRp – Analysis of Results

In this chapter, a detailed description of each structure determined is given, together with an analysis of its contribution to the understanding of the mechanism of polymerisation catalysed by ϕ6 RdRp. Thermal denaturation curves for WT, E634Q and E491Q mutants under different divalent cation conditions are also detailed. The descriptions and analyses are grouped again into subchapters according to the statement being addressed:

(i) The effect of Mn^{2+} and Mg^{2+} on the structural stability of ϕ6 RdRp

- Thermal denaturation experiments (WT, E634Q and E491Q)

(ii) Template and NTP entry during preinitiation

- WT-5'-AATCT-3'-GTP-Mg^{2+}, WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}, WT-Tri4T-ATP-Mg^{2+}
- E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} and E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+}

(iii) C-terminal domain displacement and transition to elongation

- Trypsin-digested WT RdRp ($P2^{T1}$)
- E634Q apo and E634Q-hairpin-AMPPCP-Mg^{2+} complexes

(iv) Entry of Mg^{2+} ions with incoming NTPs

- E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}, E634Q-Tetra2T-AMPPCP-Mg^{2+} and E634Q-Tri4T-AMPPCP-Mg^{2+} complexes

Unless otherwise stated, figures in this chapter were drawn using Pymol (Delano, 2002).

3.1. The effect of Mn^{2+} and Mg^{2+} on the thermal stability of ϕ6 RdRp

Both Mg^{2+} and Mn^{2+} have been shown to stimulate the polymerisation activity of ϕ6 RdRp and RdRps of several other viruses (see Chapter 1, section 1.1.3.2., C). A high affinity Mn^{2+} binding site (or non-catalytic ion site) was identified in (Butcher *et al*., 2001) at the palm domain, co-ordinated by D454, E491 and A495. A similar non-catalytic site has been identified in other vRdRps (see Chapter 1, **Table 1.2**). It was hypothesised that the Mn^{2+} ion is bound at this site in the purified 'apo' protein and could have a structural role in stabilizing the overall structure of ϕ6 RdRp. To investigate this hypothesis and the effect of Mn^{2+} and Mg^{2+} on ϕ6 RdRp stability, thermal denaturation experiments were carried out in buffered solutions in the presence or absence of EDTA and/or divalent cations. In addition, similar experiments were carried out for the CTD mutant E634Q [with a 2.5-fold higher polymerisation activity, (Sarin *et al*., 2009)] and Mn^{2+} binding site mutant E491Q [with compromised polymerisation activity (Poranen *et al*., 2008b)] to try and pinpoint the effects of the divalent cations.

3.1.1. Manganese

The melting curves obtained for WT RdRp (**Fig. 3.1, A** and statistics shown in **Table 3.1**) reveal that Mn^{2+} is bound in the purified WT apoenzyme and actually reduces its structural stability [compare melting temperature in reference buffer (45.8^{o}C) and in 5 mM EDTA (48.6^{o}C) that strips out the bound ion; **Table 3.1** and curve shift **Fig. 3.1, A**]. The higher melting temperature in the presence of 5 mM EDTA could be returned to reference buffer levels by introducing 10 mM Mn^{2+} (melting temperature for 5 mM EDTA + 10 mM $MnCl_2$ of 45.6^{o}C; **Table 3.1**). The bound Mn^{2+} is not removed from the RdRp by 0.1 or 1 mM EDTA

used in standard reaction and purification buffers (no change in T_m from reference buffer; **Table 3.1** and no curve shift in **Fig. 3.1, A**) suggesting that the high affinity site binds the Mn^{2+} from the cell during RdRp expression that remains stably bound during purification.

The E634Q mutant, which removes the E634 - K144 salt-bridge pinning the CTD to the main body of the protein (Sarin *et al*., 2009; thermal denaturation curves shown in **Fig. 3.1, C** with statistics in **Table 3.1**) revealed a significantly lower thermal stability compared to the WT RdRp in reference buffer (melting temperature in reference buffer 40.8^oC compared to 45.8^oC for WT; **Table 3.1**). This decrease in thermal stability for E634Q is presumably due to increased flexibility of the CTD. Similarly, Mn^{2+} was shown to co-purify with E634Q and structurally destabilise the polymerase [compare melting temperatures in reference buffer (40.8^oC) and in 5 mM EDTA (44.8^oC); **Table 3.1** and curve shift **Fig. 3.1, C**]. Again, melting temperatures could be reduced to reference buffer levels by the addition of 10 mM Mn^{2+} (melting temperature for 5 mM EDTA + 10 mM $MnCl_2$ of 40.3^oC; **Table 3.1**).

In contrast to WT and E634Q, Mn^{2+} is not co-purified with the Mn^{2+} binding site mutant E491Q which has an enhanced structural stability in reference buffer (49.1^oC, compared with 45.6^oC and 40.8^oC for WT and E634Q respectively; **Table 3.1** and **Fig. 3.1, B**). The absence of Mn^{2+} was confirmed by addition of 0.1 – 5 mM EDTA that had no effect on the increased thermal stability [compare E491Q in reference buffer (49.1^oC) with 0.1, 1 and 5 mM EDTA (49.1^oC, 49.1^oC and 49.6^oC respectively); **Table 3.1** and no curve shift **Fig. 3.1, B**]. Interestingly, E491Q was somewhat destabilised when saturated with 10 mM Mn^{2+}, resulting in a melting temperate comparable to WT in reference buffer [compare E491Q in reference buffer (49.1^oC) and in 10 mM $MnCl_2$ (45.8^oC), with WT in reference buffer (45.8^oC); **Table 3.1** and curve shift **Fig. 3.1, B**], confirming that Mn^{2+} ions are required to destabilize the RdRp.

Attempts were made to run the ϕ6 RdRp initiation reaction *in crystallo* using the E491Q mutant (Poranen *et al*., 2008b). The results demonstrated that in the absence of bound Mn^{2+} at the non-catalytic site [that is present in the WT initiation complex, (Butcher *et al*., 2001)], an inactive initiation complex was formed. The lack of Mn^{2+} resulted in the movement of Y630 to a position 4Å away from the active site and complete reorientation of non-catalytic ion site residue D454 to hydrogen bond with the sugar of the D1 NTP. As a result, D1 was positioned 6Å away from the phosphate backbone of D2 NTP, preventing the catalysis; a similar situation was observed for a Ca^{2+}-inhibited complex determined by Salgado *et al*., (2004), see Chapter 1, section 1.1.3.2, C. This finding suggests that in addition to the increased molecular flexibility, bound Mn^{2+} may be required to establishing the correct geometry about the active site for efficient catalysis.

Table 3.1. Thermal denaturation statistics.

Condition	WT		E491Q		E634Q	
	T_m [a] (°C)	ΔT_m [b] (°C)	T_m [a] (°C)	ΔT_m [b] (°C)	T_m [a] (°C)	ΔT_m [b] (°C)
Reference buffer [c]	45.8 (±0.05)	–	49.1 (±0.08) [d]	–	40.8 (±0.03) [e]	–
+ 0.1mM EDTA	45.6 (±0.07)	- 0.2	49.1 (±0.06)	- 0.0	40.1 (±0.08)	- 0.7
+ 1mM EDTA	45.6 (±0.07)	- 0.2	49.1 (±0.06)	- 0.0	40.3 (±0.03)	- 0.5
+ 5mM EDTA	48.6 (±0.08)	+ 2.8 [d]	49.6 (±0.05)	+ 0.5	44.8 (±0.03)	+ 4.0 [d]
+ 5mM EDTA + 10mM $MgCl_2$	48.1 (±0.06)	+ 2.3	48.1 (±0.06)	- 1.0 [e]	44.3 (±0.03)	+ 3.5
+ 10mM $MgCl_2$	48.1 (±0.06)	+ 2.3	48.1 (±0.06)	- 1.0	43.8 (±0.03)	+ 3.0
+ 5mM EDTA + 10mM $MnCl_2$	45.6 (±0.07)	- 0.2	47.8 (±0.05)	- 1.3	40.3 (±0.05)	- 0.5
+ 10mM $MnCl_2$	45.8 (±0.06)	- 0.2	45.8 (±0.05)	- 3.3	40.1 (±0.08)	- 0.7

[a] T_m values shown were calculated by applying a Boltzmann Distribution Equation about the sigmoidal melting curves to obtain the inflection point (the midpoint of the unfolding transition) of the slope (GraphPad Prism software). Values in parentheses indicate the standard error for calculated T_m values.

[b] Any shift in T_m (ΔT_m) from reference under different conditions indicates a change in protein stability. An increase in T_m ($+\Delta T_m$) indicates a stabilization of the protein by an increase in structural order and a reduction in conformational flexibility; a decrease in T_m ($-\Delta T_m$) indicates a destabilization (Geerlof *et al*., 2006).

[c] Reference buffer is 50 mM Tris-HCl pH 8.0, 50 mM NaCl.

[d] Blue box highlights conditions that result in a significant stabilisation.

[e] Red box highlights conditions that result in a significant destabilisation.

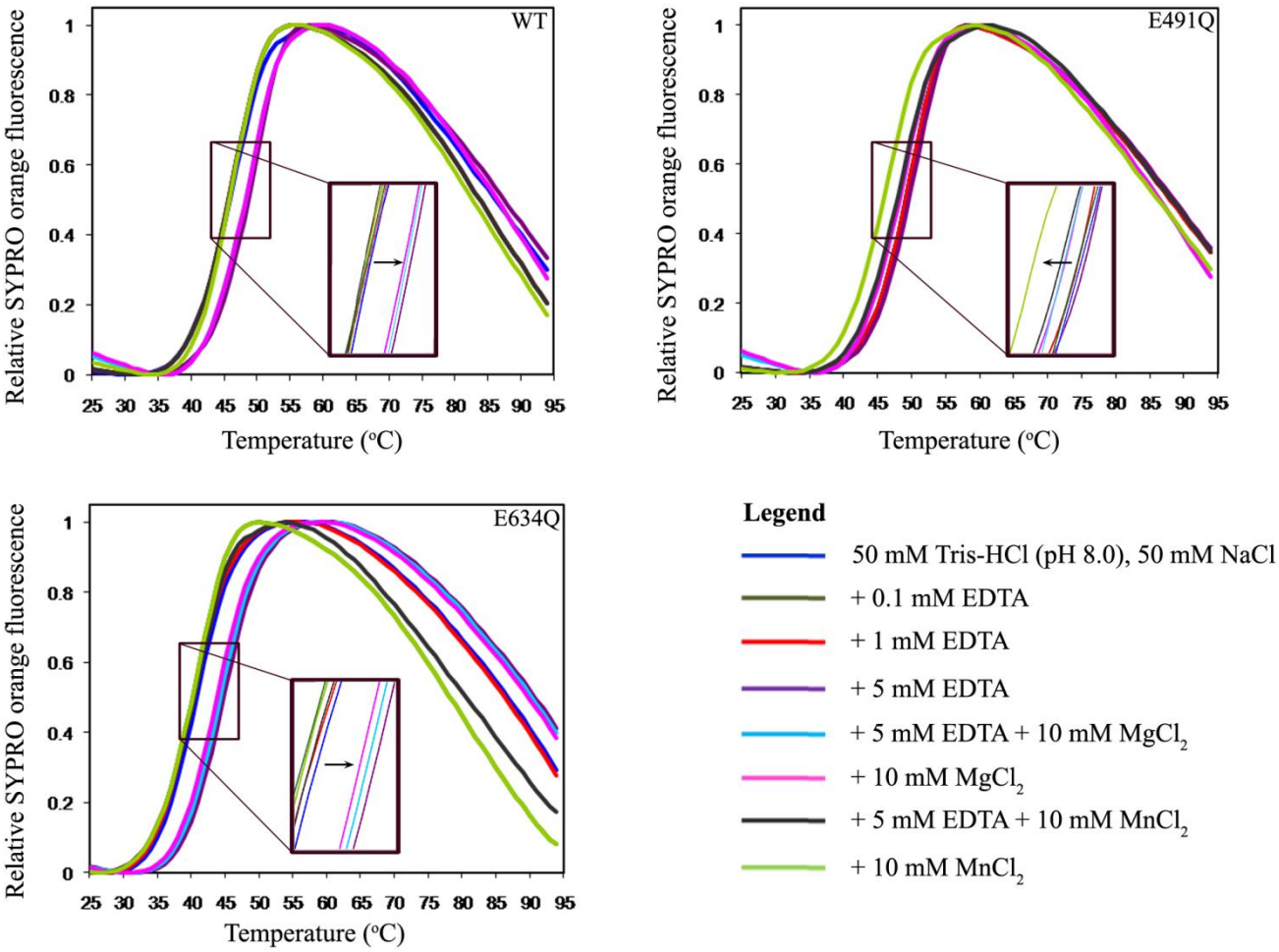


Figure 3.1. Thermal denaturation curves of ϕ6 WT, E491Q and E634Q RdRps.

The thermal shift assay (ThermoFluor) was carried out in a real-time PCR machine (BioRad DNA Engine Opticon 2) where buffered solutions of protein (WT, E491Q and E634Q RdRps) and fluorophore (SYPRO Orange; Molecular Probes, Invitrogen), with and without additives, were heated in a stepwise fashion from 25°C to 95°C. 5µL protein (1 mg/mL) and 5 µL SYPRO Orange (Molecular Probes, Invitrogen) were made up to a total assay volume of 50 µL with reference buffer (50 mM Tris–HCl, pH 8.0, 50 mM NaCl) in white low profile thin-wall PCR plates (Abgene) sealed with microseal 'B' films (BioRad). The fluorophore was excited in the range of 470–505 nm and fluorescence emission was measured in the range of 540–700 nm every 0.5°C after a 10 s hold. The additives screened were: 0.1 mM, 1 mM or 5 mM EDTA; 5 mM EDTA with 10 mM $MgCl_2$; 5 mM EDTA with 10 mM $MnCl_2$; 10 mM $MgCl_2$ without EDTA; 10 mM $MnCl_2$ without EDTA. Denaturation curves for these conditions are shown for **(A)** WT, **(B)** E491Q and **(C)** E634Q polymerases, and are coloured according to the legend. Close-ups for each condition (indicated by black boxes) are provided

to distinguish curves from one another about the melting temperature (T_m). Black arrows indicate the shift of curves to higher or lower T_m values in accordance with the text. The T_m values were calculated by applying a Boltzmann Distribution Equation about the sigmoidal melting curves to obtain the inflection point of the slope (the midpoint of the unfolding transition), using the GraphPad Prism software and are shown in **Table 3.1** with standard errors.

3.1.2. Magnesium

The melting curves obtained for WT RdRp and E634Q reveal that 10 mM Mg^{2+} can out-compete bound Mn^{2+} at the non-catalytic ion site in the purified proteins and lead to a concomitant increase in structural stability [compare melting temperatures of WT and E634Q in reference buffer (45.8^{o}C and 40.8^{o}C respectively) with WT and E634Q in 10 mM $MgCl_2$ (48.1^{o}C and 43.8^{o}C respectively); **Table 3.1** and curve shifts in **Fig. 3.1, A and C**]. However, there is little difference in the increase in structural stability provided by 10 mM Mg^{2+} compared with 5 mM EDTA that strips out the bound 'destabilizing' Mn^{2+}, leaving only the apoenzyme; compare WT and E634Q in 10 mM $MgCl_2$ (48.1^{o}C and 43.8^{o}C respectively) with WT and E634Q in 5 mM EDTA (48.6 ^{o}C and 44.8^{o}C respectively); **Table 3.1**. Furthermore, the increased melting temperature was not affected by the presence of both 10 mM Mg^{2+} and 5 mM EDTA (48.1 ^{o}C and 44.3^{o}C for WT and E634Q respectively; **Table 3.1**. These data suggest that Mg^{2+}, at a concentration of 10 mM, is sufficient to out-compete the bound destabilizing Mn^{2+} at the non-catalytic ion site, and increase the structural stability of ϕ6 RdRp to apoenzyme levels; importantly, Mg^{2+} does not functionally replace the 'destabilising' Mn^{2+} at the non-catalytic ion site. Similar experiments with E491Q mutant strengthen the idea that Mg^{2+} cannot functional replace the destabilizing Mn^{2+} at this site [compare melting temperatures of E491Q in 10 mM $MgCl_2$ (48.1^{o}C) with E491Q in 10 mM $MnCl_2$ (45.8^{o}C); **Table 3.1** and curve shift **Fig. 3.1, B**].

3.2. Template and NTP entry during preinitiation

The replacement of Mn^{2+} by 10 mM Mg^{2+} in the WT and E634Q structures allowed us to capture new preinitiation stages not observed when Mn^{2+} is bound, which probably reflects changes in the kinetics of complex formation due to stabilization of the enzyme. Each complex is described in detail in the following sections.

3.2.1. WT-5'-AATCT-3'-GTP-Mg^{2+}

In order to determine the effect of replacing the non-catalytic Mn^{2+} with Mg^{2+} (shown to increase WT RdRp thermal stability, **Table 3.1** and **Fig. 3.1**) on NTP and template entry during preinitiation of replication, WT RdRp crystals were soaked with 10 mM $MgCl_2$ together with a 5nt mimic of the 3'-end of the (+) strand of ϕ6 genomic dsRNA (5'-AATCT-3') and a GTP substrate. Mg^{2+} was found in place of the non-catalytic Mn^{2+} in all three molecules of the crystallographic asymmetric unit (I, II and III) for WT-5'-AATCT-3'-GTP-Mg^{2+} in the standard "downward" position co-ordinated by D454, E491 and A495 as observed for all published structures with Mn^{2+} at this site (Butcher *et al*., 2001; Poranen *et al*., 2008b; Salgado *et al*., 2004), **Fig. 3.2**. Replacement of the non-catalytic Mn^{2+} by Mg^{2+} was indicated by significant differences in electron density (and negative difference density if Mn^{2+} was refined into the structure instead of Mg^{2+}). The 3.2Å 2|Fo|-|Fc| SIGMAA electron density map showed a single GTP bound at the active site for all molecules with the triphosphate backbone co-ordinated by interrogation site residues R225, R268 and R270, allowing the guanine to base stack against Y630 of the *de novo* priming platform (**Fig. 3.2**); this arrangement has been seen previously for the GTP bound structure described in (Butcher *et al*., 2001).

For molecule II, the electron density map revealed DNA nts 5'-AATC-3' of the soaked in (+) strand mimic (5'-AATCT-3') bound at the entrance to the template tunnel extending out over the outer surface of the molecule (**Fig. 3.3**). The DNA is sandwiched between the entrance tunnel of molecule II and a nearby molecule III α-helix supplying K592 that may play a role in stabilising the DNA at this position. The two innermost nts 5'-TC-3' from molecule II could also be modelled at the entrance to molecule III, however there is no density to allow

positioning of the outer nts, presumably due to the lack of a nearby 'sandwiching' molecule in crystal space (see crystal packing, see Chapter 2, section 2.5.1 and **Fig. 2.5**).

Previous mutagenesis studies identified residues about the template tunnel rim as essential for efficient template binding (Sarin *et al*., 2009). Mutations R30 to A and K541 to L significantly reduce the RNA-binding affinity of the polymerase and level of RNA synthesis for both replication and transcription; it is likely that the reduced charge around the template tunnel entrance, caused by R30A and K541L mutations, interferes with the correct guidance of the template down into the tunnel. Supporting this idea, R30 and K541, which are at the centre of two positively charged patches on either side of the template tunnel entrance, make direct contacts with the phosphate backbone and/or bases of the bound DNA in molecule II (**Fig. 3.3, right panel**) and are therefore likely to be essential anchoring residues that draw the template down towards the active site.

Figure 3.2. (overleaf). NTP binding tunnel and *de novo* priming platform for molecules I-III of WT-5'-AATCT-3'-GTP-Mg^{2+}.

Secondary structure elements are shown and coloured according to the generic polymerase domain architecture: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). **(A)** Molecule I; **(B)** Molecule II; **(C)** Molecule III. For **(A)-(C)** key amino acids of the interrogation site and non-catalytic ion site are labelled and are shown in stick representation (grey, carbon; blue, nitrogen; red, oxygen; Mg^{2+} is shown as a black sphere). Y630 of the priming platform (yellow, carbon; red, oxygen) and bound GTP (green, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) are also shown by sticks. The 3.2Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are shown as blue wires. The position of Mg^{2+} at the non-catalytic ion site is listed to right of each panel.

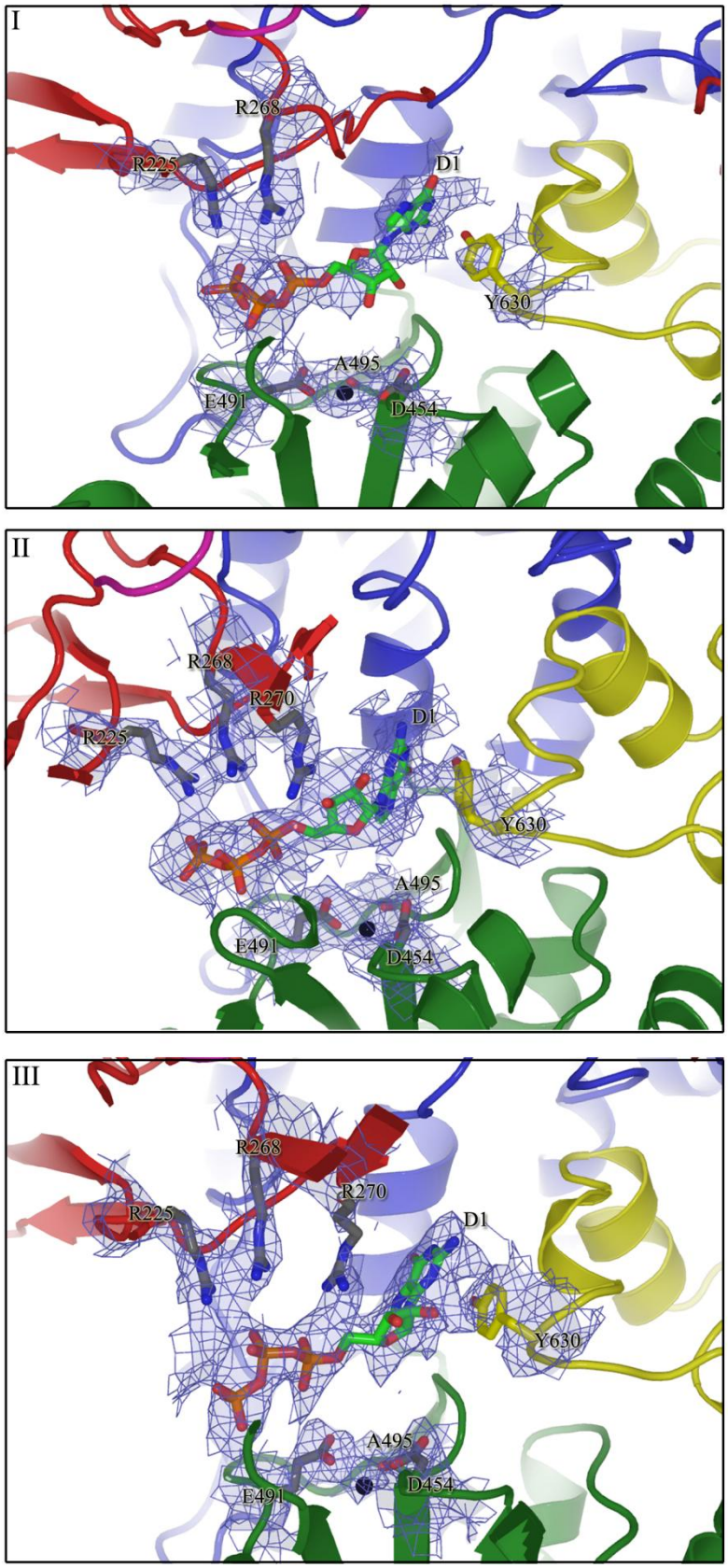
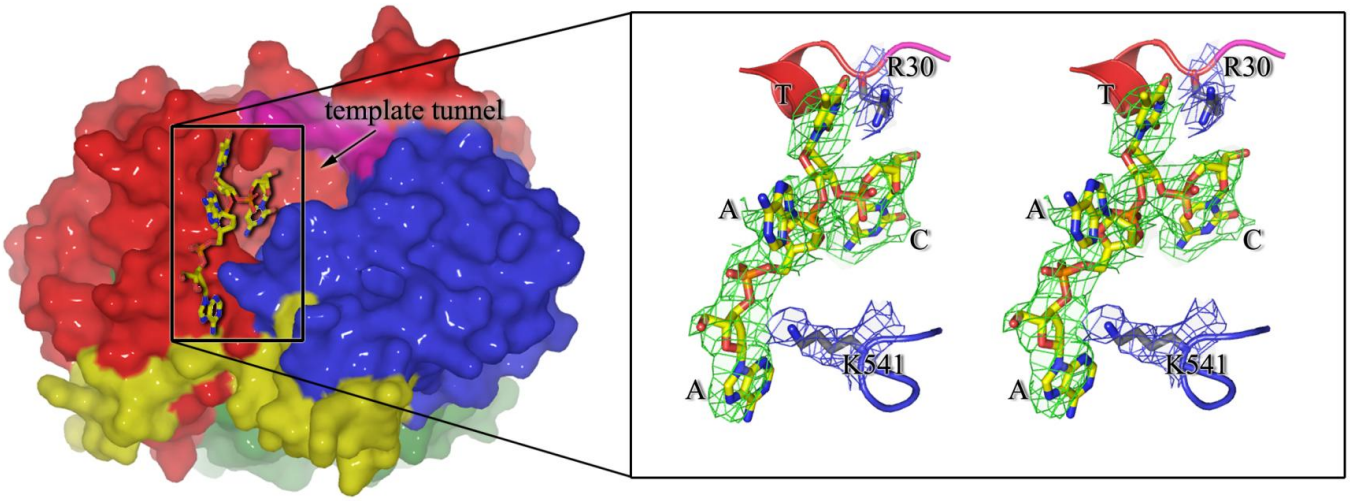

A
I
R268
R225
D1
Y630
A495
E491
D454
Downward
B
II
R268
R270
D1
R225
Y630
A495
E491
D454
Downward
C
III
R268
R270
R225
D1
Y630
A495
E491
D454
Downward

Figure 3.2.

Figure 3.3. Template binding on the surface of WT-5'-AATCT-3'-GTP-Mg^{2+} (mol. II).

Molecular surfaces (*left panel*) and fragments of polypeptide chains in the close-up stereo images (*right panel*) coloured by polymerase domain as in **Fig. 3.2**. Bound oligonucleotide is depicted using stick presentation (carbon, yellow; nitrogen, blue; oxygen, red; phosphorous, orange). The template tunnel entrance is indicated by an arrow (black). ***Right panel***, 3.2Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ for the oligonucleotide (green wire) and for amino acid residues R30 and K541 (blue wire). Residues R30 and K541 (stick presentation: carbon grey, nitrogen, blue; oxygen, red) that sit on either side of the template tunnel stabilise the incoming template. Due to local disorder in the bases, nucleotides were assigned by best-fit of sequence into the available electron density; assigned bases and positions of R30 and K541 are indicated.



3.2.2. WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}

In an attempt to capture stages leading to the initiation of replication [proposed to follow the same mechanism as for transcription as outlined in Chapter 1, section 1.1.3.4 (Butcher *et al*., 2001)], WT crystals were soaked again with a DNA minic of the 3'-end of the (+) strand of ϕ6 genomic dsRNA (5'-AATCT-3') and 10 mM $MgCl_2$, but this time with both ATP and GTP substrates. Similar to WT-5'-AATCT-3'-GTP-Mg^{2+}, the soaked in Mg^{2+} had out-competed the non-catalytic Mn^{2+} in all three molecules of the ASU. Interestingly, in this case the co-ordination of the bound Mg^{2+} about the non-catalytic ion site was sensitive to the positioning of the bound NTPs.

In molecule I, the 2.9Å 2|Fo|-|Fc| SIGMAA electron density map revealed a GTP in the active site at a D1 NTP position with an orientation suitable for Watson-Crick base pairing with the T1 (3'-end) nucleotide of the template (**Fig. 3.4**); this was demonstrated by superimposing the published initiation complex (Butcher *et al*., 2001) with WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}. Unlike WT-5'-AATCT-3'-GTP-Mg^{2+}, the D1 triphosphate backbone was not held at the interrogation site but instead angled down towards the non-catalytic ion site, stabilised by a nearby R81 (**Fig. 3.4, A**). Interrogation site residues Q206, R225 and R268 are instead involved in the co-ordination of the triphosphate backbone of a second NTP (built as triphosphates of an ATP that could base pair with the penultimate 3' cytosine nucleotide) at a position comparable to those of the D2 NTP in the published initiation complex (**Fig. 3.4, B**). In molecules II and III a similar arrangement of D1 and D2 NTPs was observed, with the triphosphate backbone of the D2 NTP co-ordinated at the interrogation site (residues Q206, R225 and R268 for molecule II; R225, R268 and R270 for molecule III), and D1 NTP base stacking with Y630; **Fig. 3.4, B and C**. As in molecule I, the D1 NTP triphosphate backbone

is angled towards the non-catalytic ion site in order to stabilise the negative charge. However, in molecules II and III the Mg^{2+} bound at the non-catalytic ion site is drawn out of its standard "downward" position and is co-ordinated instead in a "centred" (molecule II) or "upward" (molecule III) position; the naming refers to the position of Mg^{2+} either below (downward), in line with (centred), or above (upward) the D454 and E491 axis of the non-catalytic ion site. In the centred position, Mg^{2+} is co-ordinated by the γ-phosphate of the D1 GTP alongside non-catalytic ion site residues D454, E491 and A495 (**Fig. 3.4, B**). In the upward position, Mg^{2+} shifts even further out of its standard downward position and is co-ordinated by the β- and γ- phosphates of the D1 GTP and non-catalytic ion site residues D454 and E491 only; there is no co-ordination from the backbone carbonyl of A495 (**Fig. 3.4, C**). In all molecules a nearby R81 also stabilises the negative charge of the D1 NTP triphosphates about the active site.

The 2.9Å 2|Fo|-|Fc| SIGMAA electron density map for WT-5'-AATCT-3'-ATP-GTP-Mg^{2+} also revealed DNA nts 5'-AATC-3' of the soaked in (+) strand mimic (5'-AATCT-3') bound in molecule II at the entrance to the template tunnel extending out over the outer surface of the polymerase and stabilised by R30 and K541 (**Fig. 3.5**); this was also observed for WT-5'-AATCT-3'-GTP-Mg^{2+}. The DNA is sandwiched between the entrance tunnel of molecule II and a nearby molecule III α-helix as described previously. Only the inner most nucleotide (5'-C-3') could be modelled at the entrance to molecules I and III stabilised by R30, presumably due to the lack of nearby crystal contacts to stabilise the external nucleotides.

Figure 3.4. (overleaf). NTP binding tunnel and *de novo* priming platform for molecules I-III of WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}

Secondary structure elements are shown and coloured according to the generic polymerase domain architecture: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Key amino acids of the interrogation site and non-catalytic ion site are labelled and are shown in stick representation (grey, carbon; blue, nitrogen; red, oxygen; Mg^{2+} is shown as a black sphere). Downward, centred and upward Mg^{2+} positions are indicated for molecules I **(A)**, II **(B)** and III **(C)** respectively. For **(A)-(C)** bound GTP (green, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) is held at a D1 NTP position (indicated) and base stacks with Y630 of the priming platform (sticks: yellow, carbon; red, oxygen); the position of D2 ATP triphosphates (sticks: red, oxygen; orange, phosphorous) is also indicated. The 2.9Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are shown as blue wires.

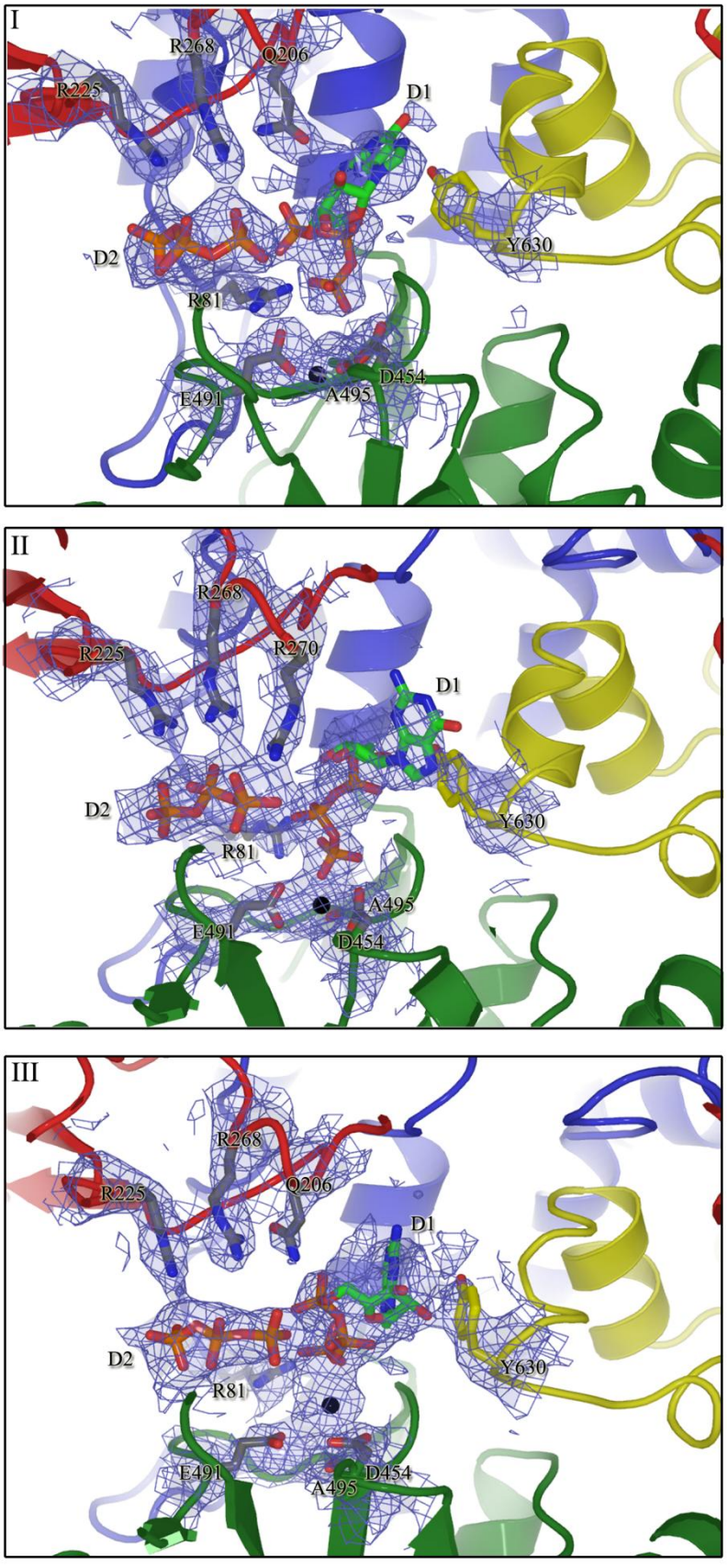
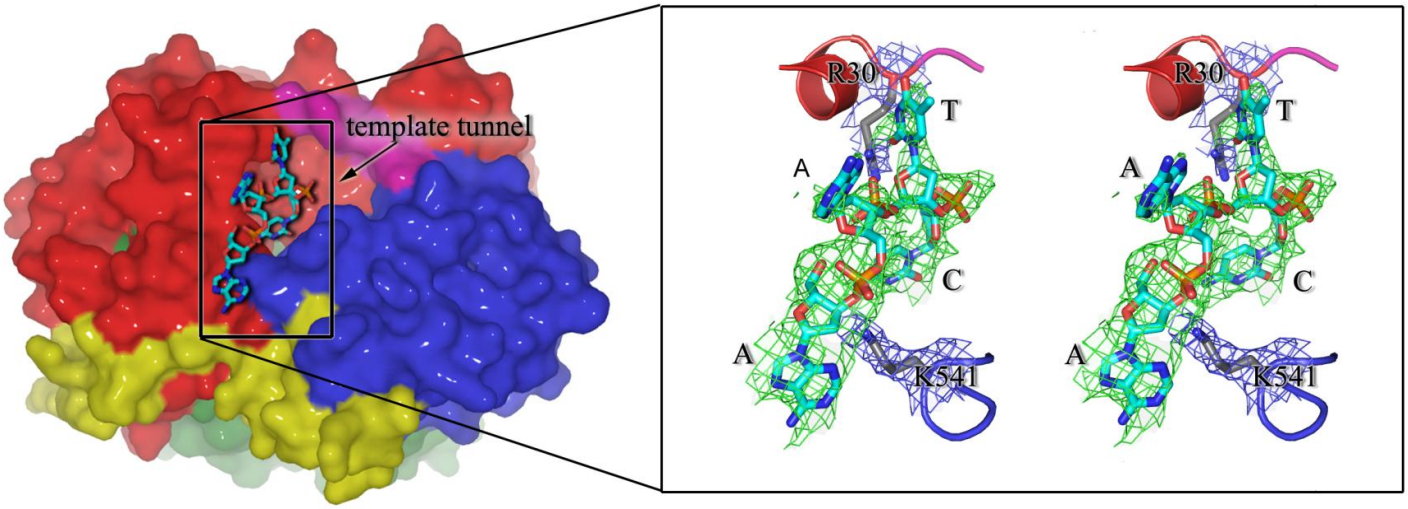

A
I
R268
Q206
R225
D1
D2
Y630
R81
D454
E491
A495
Downward
B
II
R268
R225
R270
D1
D2
Y630
R81
A495
E491
D454
Centred
C
III
R268
R225
Q206
D1
D2
Y630
R81
E491
D454
A495
Upward

Figure 3.4.

Figure 3.5. Template binding on the surface of WT-5'-AATCT-3'-ATP-GTP-Mg^{2+} (molecule II).

Molecular surfaces (*left panel*) and fragments of polypeptide chains in the close-up stereo images (*right panel*) coloured by polymerase domain as in **Fig. 3.4**. Bound oligonucleotide is depicted using stick presentation (carbon, cyan; nitrogen, blue; oxygen, red; phosphorous, orange). The template tunnel entrance is indicated by an arrow (black). ***Right panel***, 2.9Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ for the oligonucleotide (green wire) and for amino acid residues R30 and K541 (blue wire). Residues R30 and K541 (stick presentation: carbon grey, nitrogen, blue; oxygen, red) that sit on either side of the template tunnel stabilise the incoming template. Due to local disorder in the bases, nucleotides were assigned by best-fit of sequence into the available electron density; assigned bases and positions of R30 and K541 are indicated.



3.2.3. WT-Tri4T-ATP-Mg^{2+}

In attempt to capture an elongation complex of ϕ6 RdRp and assess its helicase ability, required for separating ssRNA from dsRNA about the entrance to the template tunnel (described in Chapter 1, section 1.1.3.3), WT crystals were soaked with a Tri4T DNA hairpin (5'-TTTTCGCGTAGCG-3'; underlined residues base pair to form a short hairpin with a T_m of 72.1°C), ATP and 10 mM $MgCl_2$. The hairpin was designed to mimic the synthesis of the third nucleotide of the daughter strand (D3 NTP base pairing with T3 nucleotide of template) so as to capture post-initiation rearrangements if the oligonucleotide entered the active site. The 3.3Å 2|Fo|-|Fc| SIGMAA electron density map revealed two ATPs (D1 and D2) at the active site for all three molecules, with the strongest electron density seen for molecules II and III (**Fig. 3.6**). For molecule I, the sugar and base moieties of the D2 ATP were disordered and only the triphosphate backbone was built (**Fig. 3.6, A**). Unexpectedly, four nucleotides of the Tri4T hairpin were bound at the entrance to template tunnel of molecule II (**Fig. 3.7**) in the same position and register of bases as seen for molecule II of WT-5'-AATCT-3'-GTP-Mg^{2+} and WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}. The bases 5'-AGCG-3' were assigned by best-fit into the available electron density and considering that ϕ6 RdRp catalyses the addition of NTPs onto the terminal 3'-end of bound template (Makeyev and Bamford, 2000b). These data suggest that interactions with the molecular surface of ϕ6 RdRp facilitate denaturation of the hairpin secondary structure to allow release and threading of its 3'-end into the template tunnel. Although unwound hairpin had been trapped in the entrance tunnel, an elongation complex with the hairpin loop at the active site had not formed.

Molecules I and III showed no evidence for bound oligonucleotide, however in these molecules the Mg^{2+} (at a concentration of 10 mM) had out-competed the co-purifying Mn^{2+} ion at the non-catalytic ion site and adopted different co-ordination positions, as described for

WT-5'-AATCT-3'-ATP-GTP-Mg^{2+}, to stabilise the triphosphate backbone of bound NTPs (**Fig. 3.6)**. For molecule III, the bound Mg^{2+} was co-ordinated in the "upward" position by D454, E491Q and the β- and γ- phosphates of the D1 ATP, with no co-ordination from non-catalytic ion site residue A495 (**Fig. 3.6, C**). For molecule I however, Mg^{2+} could partially occupy both an "upward" and "downward" position about the non-catalytic ion site. The upward Mg^{2+} was co-ordinated by D454, E491 and the α- and β- phosphates of D1 ATP, and the downward Mg^{2+} co-ordinated by standard non-catalytic ion site residues D454, E491 and A495 (**Fig. 3.6, A**). The occupancy was set to 0.5 for both positions. The D2 ATP triphosphates for both molecules were co-ordinated by interrogation site residues R225 and R268 (**Fig. 3.6, A and C**). Unexpectedly, for molecule II, no ion was bound at the non-catalytic ion site, however a Mg^{2+} ion could be built at the entrance to the NTP tunnel co-ordinated by the γ-phosphate of the D2 ATP (**Fig. 3.6, B**). The triphosphate backbone of the D2 ATP was co-ordinated by interrogation site residues R225, R288 and R270 on one side and by the Mg^{2+} on the other (**Fig. 3.6, B**). The position of the Mg^{2+} suggests that catalytic Mg^{2+} ions are drawn into the polymerase by the incoming NTPs.

Figure 3.6. (overleaf). NTP binding tunnel and *de novo* priming platform for molecules I-III of WT-Tri4T-ATP-Mg^{2+}.

Secondary structure elements are shown and coloured according to the generic polymerase domain architecture: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Key amino acids of the interrogation site and non-catalytic ion site are labelled and are shown in stick representation (grey, carbon; blue, nitrogen; red, oxygen; Mg^{2+} is shown as a black sphere). **(A)** Molecule I, partial occupancy of Mg^{2+} in both upward and downward positions. **(B)** Molecule II, no divalent cation bound at non-catalytic ion site, but Mg^{2+} is coordinated by the D2 ATP triphosphates; the triphosphate of the D1 ATP are stabilised by a nearby R80 (sticks: grey, carbon; blue, nitrogen). **(C)** Molecule III, upward positioning of Mg^{2+}. For **(A)-(C)** positions of bound D1 ATP (sticks: green, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) and D2 ATP (sticks: cyan, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) are indicated. The D1 ATP base stacks with Y630 of the priming platform (sticks: yellow, carbon; red, oxygen). The 3.3Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are shown as blue wires. The Mg^{2+} positions are indicated to the right of each panel.

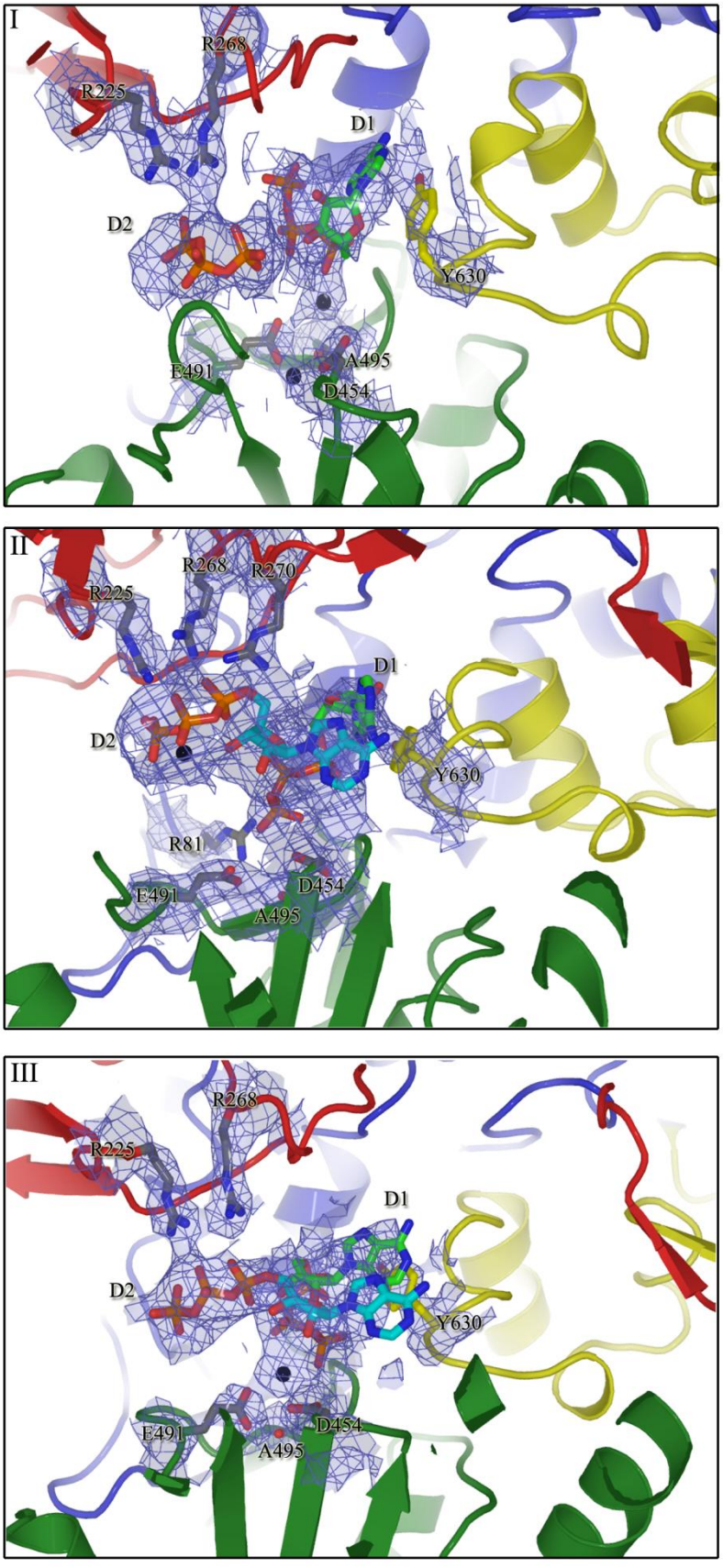
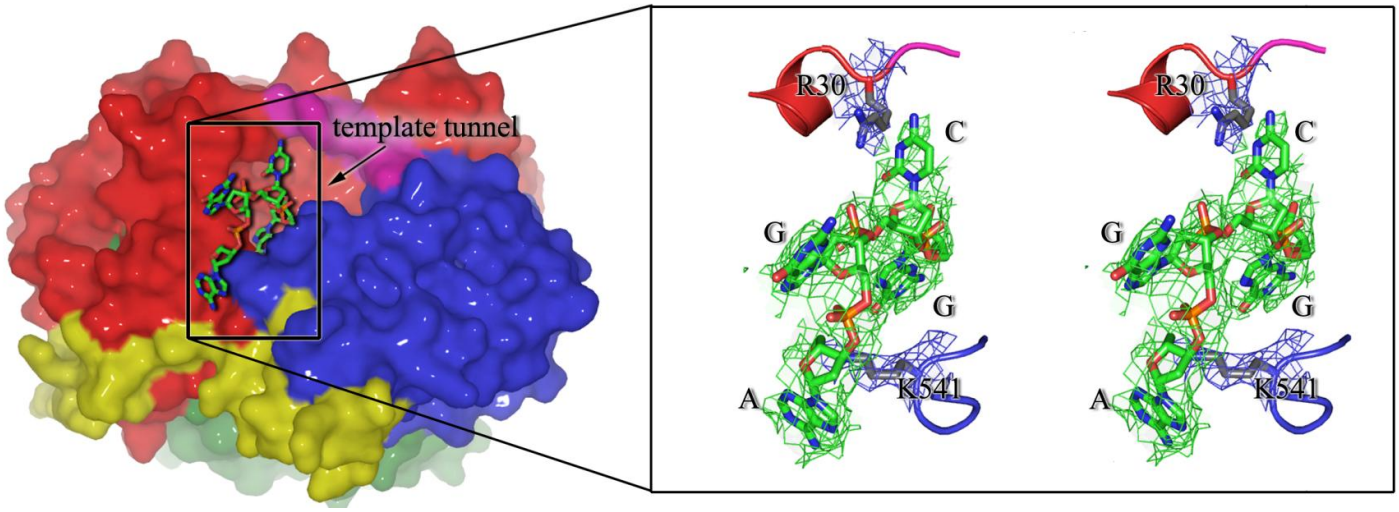

A
I
R268
R225
D1
D2
Y630
E491
A495
D454
Upward &
Downward
B
II
R268
R270
R225
D1
D2
Y630
R81
E491
D454
A495
No ion at
non-catalytic
ion site
C
III
R268
R225
D1
D2
Y630
E491
D454
A495
Upward

Figure 3.6.

Figure 3.7. Template binding on the surface of WT-Tri4T-Mg^{2+} (molecule II).

Molecular surfaces (*left panel*) and fragments of polypeptide chains in the close-up stereo images (*right panel*) coloured by polymerase domain as in **Fig. 3.6**. Bound oligonucleotide is depicted using stick presentation (carbon, green; nitrogen, blue; oxygen, red; phosphorous, orange). The template tunnel entrance is indicated by an arrow (black). ***Right panel***, 3.3Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ for the oligonucleotide (green wire) and for amino acid residues R30 and K541 (blue wire). Residues R30 and K541 (stick presentation: carbon grey, nitrogen, blue; oxygen, red) that sit on either side of the template tunnel stabilise the incoming template. Due to local disorder in the bases, nucleotides were assigned by best-fit of sequence into the available electron density; assigned bases and positions of R30 and K541 are indicated.



3.2.4. E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}

The previously discussed WT complexes suggest that in the presence of Mg^{2+} at the non-catalytic ion site, template cannot enter the active site. To test this hypothesis, crystals of the E634Q mutant, with a higher polymerisation activity compared to WT, were soaked with a GTP and Mg^{2+} as for the WT complexes, and a short RNA oligonucleotide equivalent to the consensus sequence of the 3'-ends (-) strands of the ϕ6 genomic dsRNA [containing the most preferred 3'-end sequence for initiation, (Makeyev and Grimes, 2004)].

In the crystal structure, the 3.4Å 2|Fo|-|Fc| SIGMAA electron density map revealed that Mg^{2+} had replaced the co-purifying Mn^{2+} at the non-catalytic ion site in all three molecules of the ASU as expected (**Fig. 3.8**). Two nucleotides of the RNA template (5'-CC-3') could be built at the entrance to the template tunnel for molecules II and III in the R30-K541 region, but there was no evidence of RNA in the active site. In molecule I the density at the entrance to the template tunnel was too weak to accurately fit any bound RNA. Previously published structures have demonstrated binding of RNA oligonucleotide mimics of the ϕ6 (-) strand (5'...-CC-3') at the active site, spanning the length of the template tunnel when Mn^{2+} is present at the non-catalytic ion site (Salgado *et al.*, 2004). This suggests that Mg^{2+} binding at this site prevents the movement of both RNA and DNA templates into the tunnel.

For molecule II and III, a single GTP was bound in a position and orientation similar to WT-5'-AATCT-3'-GTP-Mg^{2+} and the previously published GTP bound structure (Butcher *et al.*, 2001), with the triphosphate backbone co-ordinated at the interrogation site by R225, R678 and Q206, and guanine base moiety base stacking with Y630 of the *de novo* priming platform (**Fig. 3.8, B and C**). In these molecules, Mg^{2+} was bound in the standard "downward" position co-ordinated by D454, E491 and A495, as expected for the single NTP bound state.

A second Mg^{2+} was bound in molecule II co-ordinated by the α- and β- triphosphate backbone of the bound GTP, side chain of D327 and backbone carbonyl of S326; these residues lie on the opposite side to the interrogating arginines (**Fig. 3.8, B**). Interestingly, Mg^{2+} was built in a similar position for WT-Tri4T-ATP-Mg^{2+} molecule II co-ordinated only by the triphosphate backbone of the ATP.

Surprisingly, in molecule I, two GTPs were bound at the active site with well-defined electron density for the triphosphate backbone, sugar and base moieties for both D1 and D2 NTPs (**Fig. 3.8, A**). The triphosphates of the D1 GTP were positioned near the non-catalytic ion site in a similar fashion to that described for WT-5'-AATCT-3'-ATP-GTP-Mg^{2+} and WT-Tri4T-ATP-Mg^{2+} structures, however the D1 GTP had penetrated deeper into the active site causing Y630 to rotate away from its standard NTP-tunnel facing positioning so as to base stack with the D1 GTP (**Fig. 3.8, A**). The bound Mg^{2+} is drawn out to an "upward" position co-ordinated by non-catalytic ion site residues D454 and E491, and the α- and β- phosphates of the D1 GTP, as previously observed for WT complexes. The triphosphate backbone of the D2 GTP is held on one side by interrogation site residues R225 and R268, and on the other side is co-ordinating a second Mg^{2+} ion with nearby residues S326 and D327, as observed for molecule II (**Fig. 3.8, A and B**). Neither the D1 or D2 GTP base stack with Y630 resulting in an unusual arrangement of NTPs about the active site. The bases of the bound D1 and D2 GTPs are angled away from an initiation position suitable for base pairing with template by approximately 45^{o} and greater than 90^{o} respectively, determined by superimposing the initiation complex (Butcher *et al*., 2001) with E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} (**Fig. 3.11, A**). These data demonstrate that two NTPs can be stabilised at the active site in the absence of Watson-crick base pairing with template.

Figure 3.8. (overleaf). NTP binding tunnel and *de novo* priming platform for molecules I-III of E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}.

Secondary structure elements are shown and coloured according to the generic polymerase domain architecture: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Key amino acids of the interrogation site, non-catalytic ion site and second Mg^{2+} ion co-ordination site are labelled and are shown in stick representation (grey, carbon; blue, nitrogen; red, oxygen; Mg^{2+} is shown as a black sphere). **(A)** Molecule I, D1 and D2 GTPs bound at active site; Mg^{2+} co-ordinated in an upward position at non-catalytic ion site and second Mg^{2+} co-ordinated at D327/S326 pocket. **(B)** Molecule II, D1 GTP bound at active site; Mg^{2+} co-ordinated in a standard downward position at non-catalytic ion site and second Mg^{2+} co-ordinated at D327/S326 pocket. **(C)** Molecule III, D1 GTP bound at active site; Mg^{2+} co-ordinated in a standard downward position at non-catalytic ion site. For **(A)-(C)** the positions of bound D1 GTP (sticks: green, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) and for molecule I, D2 GTP (sticks: cyan, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) are indicated. Y630 of the priming platform is also shown (sticks: yellow, carbon; red, oxygen). The 3.4Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are depicted as blue wires. The Mg^{2+} positions about the non-catalytic site are indicated to the right of each panel.

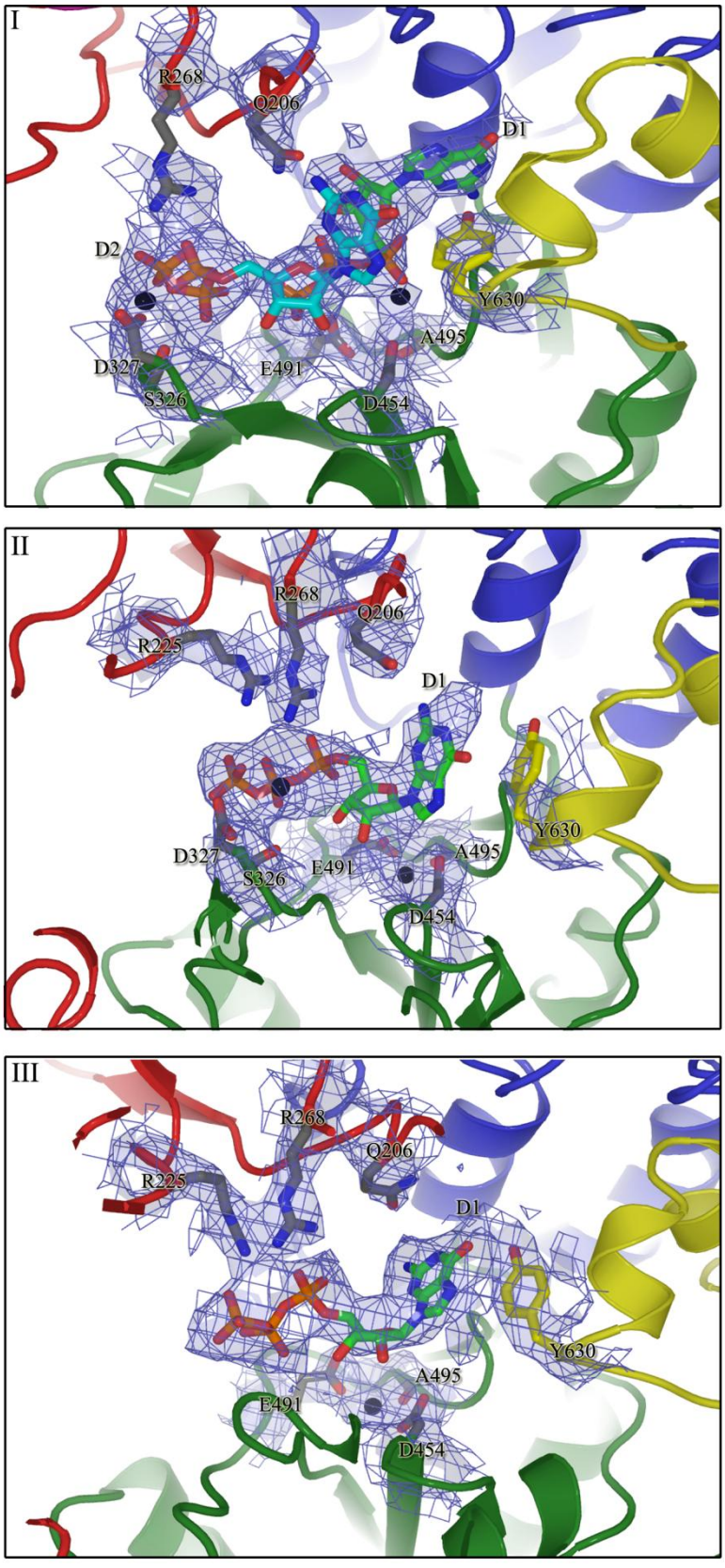

A
I
R268
Q206
D1
D2
Y630
A495
D327
E491
S326
D454
Upward
B
II
R268
Q206
R225
D1
Y630
D327
A495
E491
S326
D454
Downward
C
III
R268
Q206
R225
D1
Y630
A495
E491
D454
Downward

Figure 3.8.

3.2.5. E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+}

E634Q mutant crystals were soaked with both 10 mM $MgCl_2$ and 5 mM $MnCl_2$ to establish whether the co-purifying Mn^{2+} would remain bound at the high affinity Mn^{2+} non-catalytic ion site, as seen for previously published structures in the presence of both 2 – 5 mM $MnCl_2$ and 5 – 25 mM $MgCl_2$ (Butcher *et al*., 2001; Poranen *et al*., 2008b; Salgado *et al*., 2004) . In addition, the crystals were soaked with GTP and DNA under similar conditions described for WT complexes. The soaked in oligonucleotide (5'-TTCCCC-3') was a DNA version of the RNA oligonucleotide used for E634Q-5'-UUCCCC-3'-GTP-Mg^{2+}, designed to mimic the most preferred 3'-end entry sequence for the polymerase. A similar short DNA mimic (5'-TTTCC-3') has been shown to enter the active site with the 3'-end (T1 nucleotide) held in a specificity pocket (site S, see Chapter 1, section 1.1.3.4) when Mn^{2+} was bound at the non-catalytic ion site under similar divalent cation conditions (Butcher *et al*., 2001).

The crystal structure revealed a single GTP at the active site and unambiguous identification of a bound Mn^{2+} ion at the non-catalytic ion site in all three molecules of the ASU (**Fig. 3.9**). In addition, for all molecules, oligonucleotide was bound in the template tunnel reaching into the active site confirming that Mn^{2+} was required for complete template entry (**Fig. 3.10**).

For molecule I, three nucleotides of template (5'-CCC-3') could be modelled into the available electron density of the 3.8Å 2|Fo|-|Fc| SIGMAA map. The 3'-end (T1) cytidine had passed the active site and bound at site S placing the T2 cytidine at the active site (**Fig. 3.10**); this is in agreement with previously published structures (Butcher *et al*., 2001; Salgado *et al*., 2004) and supports the mechanism of initiation described in Chapter 1, section 1.1.3.4, where T2 is involved in the first base pair interaction with incoming NTP. The GTP in molecule I is well ordered and co-ordinated at the interrogation site by residues R268 and Q209. It has not proceeded to a base stack position with Y630 that would allow it to base pair with the T2

cytidine (**Fig. 3.9, A**).

For molecule III, 5'-CCC-3' could be built in the template tunnel but had not penetrated into the specificity pocket (**Fig. 3.10**). In this case the T1 nucleotide is held at the active site and is offset by one nucleotide with respect to molecule I, placing T3 at a position near the entrance to the template tunnel co-ordinated by R30. The GTP is well ordered with triphosphates near the interrogation site residues R225 and R268 (although density in this protein region is quite poor) and guanine moiety ready to base stack with Y630 (**Fig. 3.9, C**). Like in molecule I, the GTP has not proceeded to a base pairing position with the T1 cytidine at the active site.

Figure 3.9. (overleaf). NTP binding tunnel and *de novo* priming platform for molecules I-III of E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+}.

Secondary structure elements are shown and coloured according to the generic polymerase domain architecture: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Key amino acids of the interrogation site and non-catalytic ion site are labelled and shown in stick representation (grey, carbon; blue, nitrogen; red, oxygen; Mn^{2+} is shown as a cyan sphere). D1 GTP is bound at active site for **(A)** molecule I, **(B)** molecule II and **(C)** molecule III, near Y630 (sticks: yellow, carbon; red, oxygen) of the priming platform. For **(A)-(C)** Mn^{2+} is co-ordinated in the standard downward position at non-catalytic ion site and the positions of bound D1 GTP (sticks: green, carbon; blue, nitrogen; red, oxygen; orange, phosphorous) are indicated. The 3.8Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are depicted as blue wires. The Mn^{2+} positions about the non-catalytic site are indicated to the right of each panel.

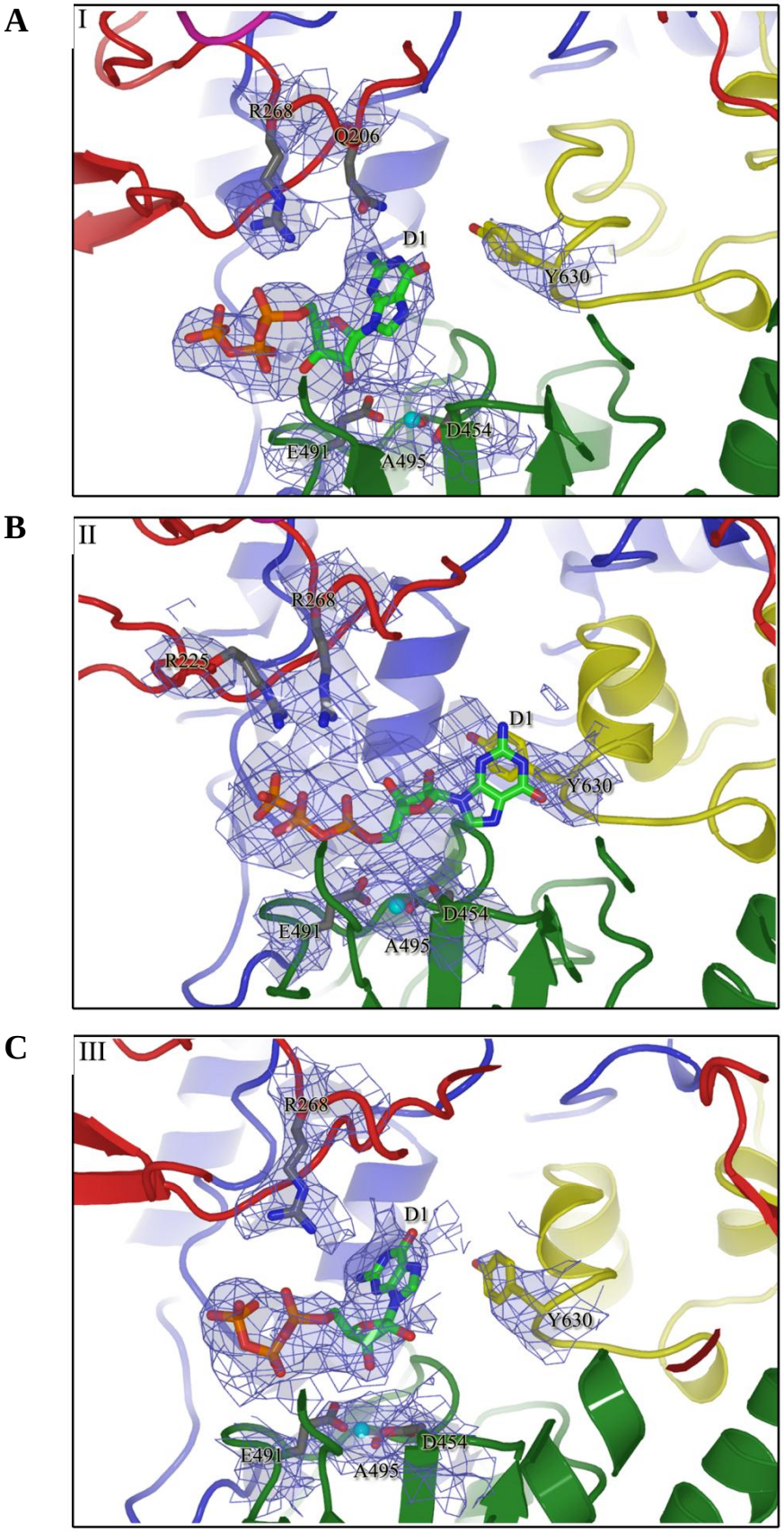
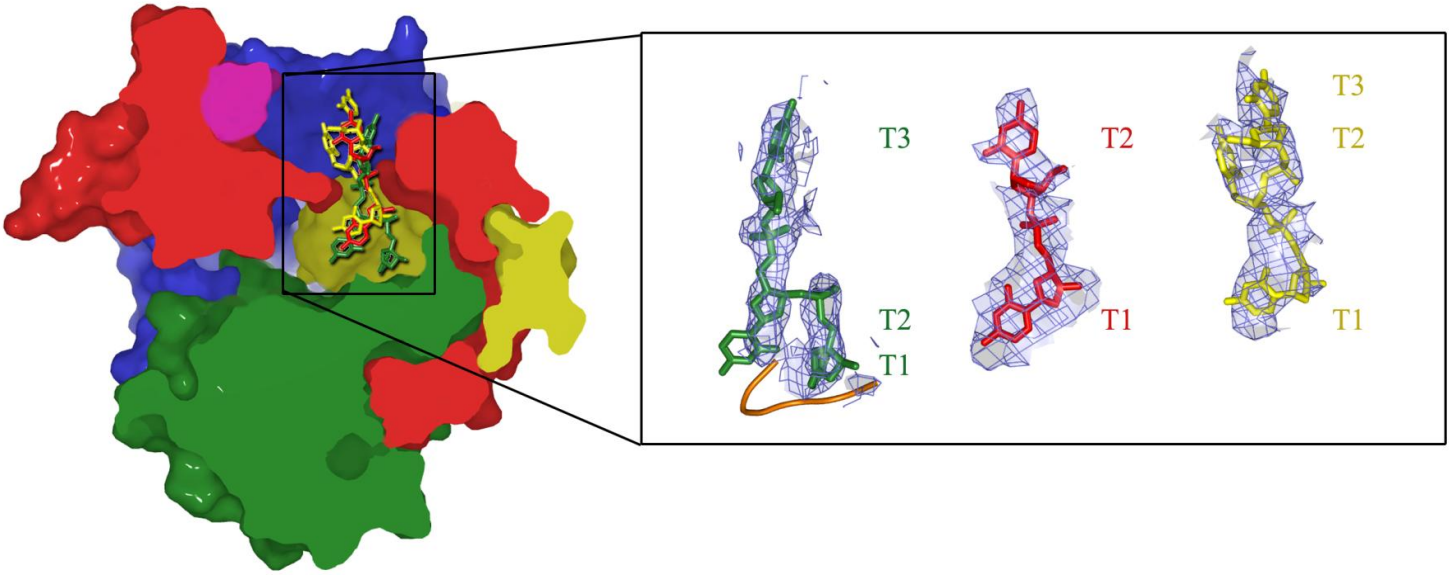

A
I
R268
Q206
D1
Y630
D454
E491
A495
Downward
B
II
R268
R225
D1
Y630
D454
E491
A495
Downward
C
III
R268
D1
Y630
D454
E491
A495
Downward

Figure 3.9.

Figure 3.10. Template binding in the template tunnel in the presence of Mn^{2+} for E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+} (molecules I-III).

Left panel, cross-section of molecular surfaces coloured by polymerase domain as in **Fig. 3.9**. Bound oligonucleotides from molecules I, II and III are superimposed and depicted using stick presentation (green, molecule I; red, molecule II; yellow, molecule III). ***Right panel***, 3.8Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ for the separate oligonucleotides (blue wire). The polypeptide chain for 629QYKW632 of the priming platform that define the specificity pocket, is shown in orange for molecule I that enters this pocket. Oligonucleotides are labelled in accordance with (Butcher *et al*., 2001) with T1 (3'-end), T2 (penultimate) and so on. Assigned bases are 5'-CCC-3', 5'-CC-3' and 5'-CCC-3' for molecules I, II and III respectively.



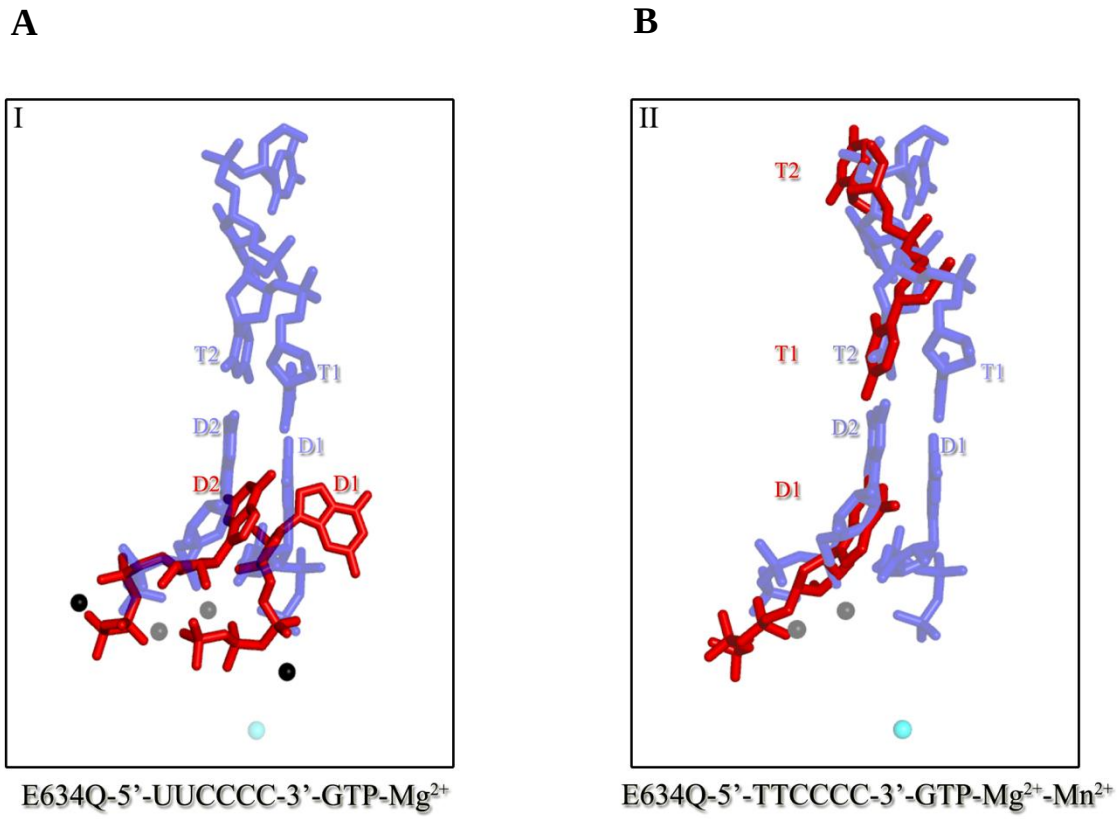
Finally, for molecule II, two nucleotides of DNA oligonucleotide (5'-CC-3') were bound at in the template tunnel in the same register observed for molecule III (T1 at active site and not in the specificity pocket (site S); **Fig. 3.10**). Interestingly, the GTP in this molecule has shifted to a base stack position with Y630 placing the guanine base near to the T1 cytidine at the active site (**Fig. 3.9, B; Fig. 3.11, B**). The guanine base is disordered (represented by weak electron density) and the ordered T1 cytidine is not oriented in a suitable position for a base pair interaction. When the initiation complex structure (Butcher *et al*., 2001) was superimposed with molecule II, it became clear that only small movements in the GTP position would be required to establish a base pair interaction with the T1 cytidine (**Fig. 3.11, B**).

Figure 3.11. NTP and template of E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} and E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+} molecules superimposed on the initiation complex of WT ϕ6 RdRp (Butcher *et al.*, 2001).

For both panels, the initiation complex is shown in semi-transparent light blue. D1 and D2 NTPs that base pair with T1 (3'-end) and T2 (penultimate) nucleotides of bound template are indicated with blue text. The two catalytic Mg^{2+} ions (semi-transparent black) and non-catalytic Mn^{2+} ion (semi-transparent cyan) are also shown.

(A) E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} molecule I. D1 and D2 GTPs (red and indicated with red text) are bound at active site with two Mg^{2+} ions (black), one at the D327/S326 pocket co-ordinated by D2 GTP and the other in an upward position at the non-catalytic ion site co-ordinated by the D1 GTP. The bound NTPs are angled away from a suitable base pairing position in the absence of bound template.

(B) E634Q-5'-TTCCCC-3'-GTP-Mg^{2+}-Mn^{2+} molecule II. D1 GTP and template 5'-CC-3' (both red, nucleotides indicated with red text) are bound at active site. Mn^{2+} (cyan) occupies the non-catalytic ion site in the standard downward position overlapping the Mn^{2+} in the initiation complex. The T1 (3'-end) of the template aligns well with the T2 nucleotide position of the initiation complex, but bound D1 GTP is somewhat displaced from the required base pairing position and aligns with D2 GTP of the initiation complex.



3.2.6. Preinitiation stages – Summary

Replacement of the non-catalytic Mn^{2+} by Mg^{2+} clearly affected the ability of the WT and E634Q RdRps to bind template. None of the three tested oligonucleotides (5'-AATCT-3', Tri4T hairpin and 5'-UUCCCC-3') could enter the template tunnel (**Fig. 3.3, Fig. 3.5 and Fig. 3.7**) when Mg^{2+} was bound. Conversely, strong density for 5'-TTCCCC-3' was seen in the template tunnel (active site and specificity pocket) for E634Q RdRp with bound Mn^{2+} (**Fig. 3.10**) which is in agreement with the previously published Mn^{2+}-bound WT structures with DNA or RNA template bound (Butcher *et al*., 2001; Salgado *et al*., 2004) and demonstrates that the non-catalytic Mn^{2+} is not lost when both Mn^{2+} and Mg^{2+} are soaked into the polymerase. In addition, it is worth noting that the E491Q mutant that does not co-purify with Mn^{2+} (section 3.1 and **Table 3.1**), has a weakened template binding (Poranen *et al*., 2008b). In line with these observations, the thermal denaturation experiments (described in section 3.1) show that both WT and E634Q RdRps have an enhanced molecular flexibility when Mn^{2+} is bound at the non-catalytic ion site (**Fig. 3.1, A and C; Table 3.1**) and increased structural stability when out-competed by Mg^{2+}. The structural destabilisation inferred by bound Mn^{2+} may be required for the efficient movement of template down towards the active site.

Although template could not be detected inside the binding tunnel with bound Mg^{2+}, four nucleotides of DNA could be fitted into electron density at the outermost region of the tunnel of molecule II for all the three WT complexes (**Fig. 3.3, Fig. 3.5 and Fig. 3.7**); two RNA nucleotides were bound at this site for the E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} complex. A closer look at this region revealed that the template enters the tunnel over a positively charged groove on the outer surface of the RdRp. Residues R30 and K541 that lie in the centre of the two positively charged patches on either sides of the template tunnel, make

direct contacts with the phosphate backbone and bases of the incoming template (**Fig. 3.3, Fig. 3.5 and Fig. 3.7**) drawing the template towards the active site; this is in direct agreement with previous mutagenesis data that showed significant reduction in template binding and polymerisation activity in both R30A and K541L mutants (Sarin *et al*., 2009). Interestingly, opening of the Tri4T hairpin was evident on the surface of the RdRp (**Fig. 3.7**) indicating that interactions with the molecular surface of the RdRp can facilitate denaturation of oligonucleotide secondary structures before the template enters into the template tunnel as suggested by Bruenn (2003), Butcher *et al*., (2001) and Sarin *et al*., (2009).

Although formation of a Watson-Crick base-paired initiation complex was hindered by the reduced ability of template to enter the tunnel with bound Mg^{2+}, two NTPs could still be stabilised at the active site for both WT and E634Q RdRps, in an initiation-like position about the *de novo* priming platform, which has not been observed previously (**Fig. 3.4, Fig. 3.6 and Fig. 3.8, A**). When only one GTP is bound (**Fig. 3.2; Fig. 3.8, B and C**), the Mg^{2+} at the non-catalytic site is coordinated by the side-chains of D454 and E491, and the carbonyl backbone of A495, as seen in the published Mn^{2+}-bound state ("downward position"). However, two additional Mg^{2+}-bound states ("centred" or "upward") were captured in structures with two bound NTPs. In the centred position (**Fig. 3.4, B**), Mg^{2+} sits directly between the side-chains of D454 and E491, coordinated additionally by A495 and one of the phosphate groups of the D1 NTP. In the upward position (**Fig. 3.4, C; Fig. 3.6, C; Fig. 3.8, A**), Mg^{2+} sits above the side-chains of D454 and E491 and is also coordinated by two phosphate groups of the D1 NTP (no interaction with A495). It is likely that Mg^{2+} continually flips between a downward and upward position about the D454-E491 axis of the non-catalytic ion site, stabilising the nearby negative charge of the D1 NTP when two NTPs are bound during preinitiation; in agreement with this the electron density in WT-Tri4T-ATP-

Mg^{2+} molecule I suggests partial occupancy of Mg^{2+} both in upward and downward states (**Fig. 3.6, A**).

3.3. C-terminal domain and transition to elongation

After formation of the initiation complex (dinucleotide state), the CTD (residues 604-664), which blocks the exit of template from the polymerase, must be displaced for chain elongation to proceed (shifting down of template and addition of the third nucleotide). In attempt to capture this CTD displacement and transition to elongation, a series of soaking experiments were carried out with short DNA hairpins designed to mimic a synthesised trinucleotide: Tri2T (5'-TTCGCGTAGCG-3'), Tri4T (5'-TTTTCGCGTAGCG-3'), Tetra2T (5'-TTCGCGTAAGCG-3') and Tetra4T (5'-TTTTCGCGTAAGCG-3'); nucleotides that can hybridize to form a trinucleotide base-paired stem are underlined. These hairpins were designed to have a high melting temperature so as to capture a stalled elongation complex at the active site with the CTD displaced (see Chapter 2, section 2.3.2.1 for full details). We were unable to obtain direct crystallographic evidence for the CTD movement during the switch to elongation with WT enzyme soaked with hairpins (see WT-Tri4T-ATP-Mg^{2+}, section 3.2.3), as the unwound hairpin did not enter the polymerase in the presence of bound Mg^{2+}. Thermal denaturation experiments suggested that the E634Q mutant may provide an alternative route to visualising movement of the CTD, given its enhanced molecular flexibility compared to WT (**Table 3.1** and **Fig. 3.1, C**) and weakened CTD attachment to the main body of the polymerase (no E634-K144 salt bridge, **Fig. 3.13**). To test this hypothesis, E634Q crystals were soaked with one of the four DNA hairpins, Mg^{2+} and AMPPCP in a similar fashion described for WT-Tri4T-ATP-Mg^{2+}; the ATP analog (AMPPCP) was used to stall the polymerisation reaction at the 5'-thymidines of the hairpin. In addition, the crystal structure of purified E634Q was obtained and used as a reference for structural changes.

3.3.1. E634Q apoenzyme

The 2.7Å crystal structure of E634Q RdRp revealed Mn^{2+} bound at the non-catalytic sites of all three molecules in the ASU (**Fig. 3.12**); this is in agreement with the thermal denaturation results (**Table 3.1** and **Fig. 3.1, C**) that show that Mn^{2+} co-purifies with E634Q RdRp. The structure of Mn^{2+}-bound E634Q was essentially indistinguishable from that of the WT enzyme (Butcher *et al.*, 2001), apart from the lack of the salt-bridge between K144 and E634 (**Fig. 3.13**), conferring increased flexibility on residue Q634. The loss of this salt bridge did not cause a noticeable disorder at the C-terminal domain, which remained well-defined.

Figure 3.12. E634Q apoenzyme non-catalytic ion site.

The non-catalytic ion site is depicted within the palm domain (green, showing secondary structure elements) for molecules I, II and III of the ASU. Side-chains of residues D454 and E491, and the backbone carbonyl group of A495 are depicted by sticks (grey, carbon; red, oxygen). Mn^{2+} is bound at all sites and is represented by a large cyan sphere. The 2.7Å 2|Fo|-|Fc| SIGMAA electron density maps for non-catalytic ion site residues and Mn^{2+} are shown, contoured at 1.2σ (blue wire).

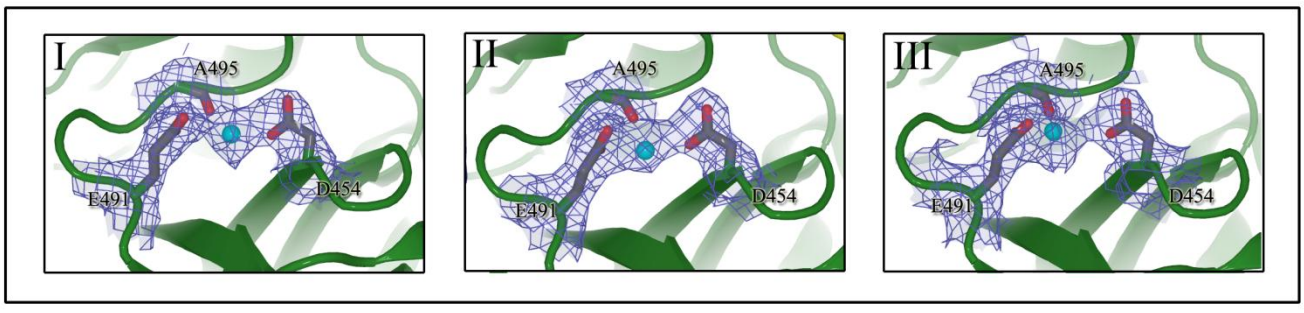
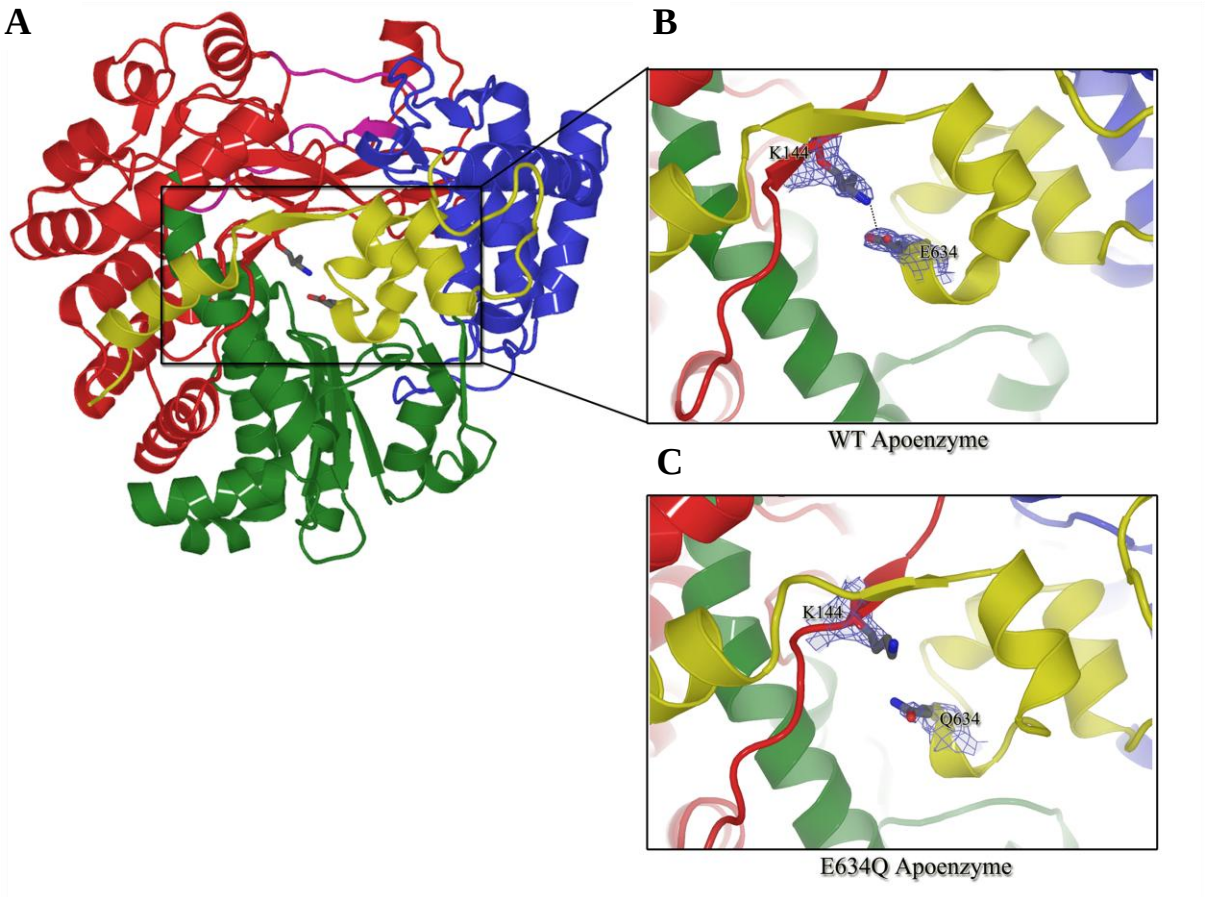


Figure 3.13. Position of E634Q – K144 salt bridge.

(A) WT ϕ6 RdRp [(Butcher *et al*., 2001), PDB accession no. 1HHS, 2.0Å crystal structure] showing secondary structure elements coloured by domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Residues E634 (CTD) and K144 (fingers domain) are depicted with sticks (grey, carbon; blue, nitrogen; red, oxygen).

(B) Close-up of the E634 – K144 salt bridge (dotted line) in WT RdRp coloured and oriented in accordance with (A).

(C) Close-up of the salt-bridge region for the E634Q apo structure, coloured and oriented in accordance with (A). The 2.7Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ (blue wire) are shown for the depicted residues in (B) and (C). The density in (C) for the salt bridge in E634Q mutant is much weaker and Q634 is not co-ordinated with K144 indicating the loss of the salt-bridge.



3.3.2. E634Q-Tri2T-AMPPCP-Mg^{2+}

The 3.0Å crystal structure of E634Q-Tri2T-AMPPCP-Mg^{2+} revealed differences in molecule III compared to molecules I and II of the ASU in terms of ions bound, ligand position and protein disorder. However, in all molecules the 3'-end of the Tri2T hairpin (5'-TTCGCGTAGCG-3') had been unwound and threaded into the template tunnel, as observed previously for the WT-Tri4T-ATP-Mg^{2+} complex.

In molecule I, the 3.0Å 2|Fo|-|Fc| SIGMAA electron density map revealed two DNA nucleotides (5'-CG-3') of the Tri2T hairpin bound in the template tunnel, with the 3'-end guanidine held at the active site in a position equivalent to the T2 nucleotide of the initiation complex (Butcher *et al*., 2001). In molecule II, DNA (5'-GCG-3') had penetrated deeper in the template tunnel with the two 3'-end nucleotides in this case bound at the active site in a position equivalent to the T1 and T2 nucleotides of the initiation complex (Butcher *et al*., 2001). In both these molecules Mg^{2+} had replaced Mn^{2+} at the non-catalytic ion site (**Fig. 3.14, A**) as previously observed for WT and E634Q complexes. The unwound hairpin however had not displaced the CTD region blocking the exit to the template tunnel, resulting in non-complementary 3'-end nucleotides held at the active site that do not base pair with AMPPCP; in fact AMPPCP remained at the interrogation site of all three molecules, co-ordinated by residues K223, R225 and R678, all with disordered base and sugar moieties (see **Fig. 3.14, D**). In terms of protein structure, molecules I and II are essentially indistinguishable from the E634Q apoenzyme.

In molecule III, there were a number of structural changes. Residues 606-611 and 631-641 (including mutated residue Q634) of the CTD were highly disordered with no continuous peptide density for these regions. Instead this area was occupied by the unwound Tri2T

hairpin that had penetrated passed the active site and specificity pocket, and had exited the template tunnel at the displaced CTD (**Fig. 3.14, B and C)**. Six nucleotides of unwound hairpin (5'-TTCGCG-3') were bound, spanning the length of the polymerase from entrance to exit, with the middle two nucleotides (5'-CG-3') held at the active site (**Fig. 3.14, B**). Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density (**Fig. 3.14, C)**. Unlike molecules I and II, no divalent cation was bound at the non-catalytic ion site in molecule III (**Fig. 3.14, A**) suggesting that it had been lost upon CTD displacement. Although molecule III had a higher RMSD (0.6Å) with the E634Q apo structure compared to molecules I and II (both 0.3Å), the main body of the protein had retained its structural integrity in the presence of a disordered CTD. The structural integrity of the polymerase core structure is further highlighted when the structure of ϕ6 RdRp with template bound at the entrance to the template tunnel (e.g. molecule II of WT-Tri4T-ATP-Mg^{2+}) was compared to molecule III of E634Q-Tri2T-AMPPCP-Mg^{2+} (**Fig. 3.15**). Although E634Q-Tri2T-AMPPCP-Mg^{2+} has a disordered CTD and template accommodated deep in the template tunnel (unlike molecule II of WT-Tri4T-ATP-Mg^{2+}) the RMSD between structures is still relatively small at 0.9Å, suggesting that the core structure of the protein is maintained under different ligand binding conditions.

Figure 3.14. (overleaf). E634Q-Tri2T-AMPPCP-Mg^{2+}: CTD disordering and loss of the non-catalytic ion.

(A) The non-catalytic ion site for molecules I, II and III. The non-catalytic ion site is depicted within the palm domain (green, showing secondary structure elements). Side-chains of residues D454 and E491, and backbone carbonyl of A495 are depicted by sticks (grey, carbon; red, oxygen). Mg^{2+} is bound in molecules I and II (large black sphere). In molecule III, water is bound at the site, represented by a small red sphere. The 3.0Å 2|Fo|-|Fc| SIGMAA electron density maps for non-catalytic ion sites are shown, contoured at 1.2σ (blue wire).

(B) Cross-section of molecular surfaces for E634Q-Tri2T-AMPPCP-Mg^{2+} molecule III. Surfaces are coloured by polymerase domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain, CTD (604–664); pink, connecting chains (31–36, 92–103). C-terminal domain residues 606-611 and 631-641 are disordered opening a path for the exit of bound template. Molecular surfaces of disordered residues have been modelled onto the structure (semi-transparent yellow) and depict the CTD blocking the exit path. Six nucleotides of unwound Tri2T hairpin can be seen spanning the length of the template tunnel from entrance to exit with middle two nucleotides held at the active site; the unwound hairpin is depicted by sticks (yellow, carbon; blue, nitrogen; red, oxygen; orange, phosphorous). Triphosphates of bound AMPPCP can be seen at the entrance to the NTP tunnel (sticks: red, oxygen; orange, phosphorous).

(C) Stereo close-ups of unwound Tri2T hairpin. Close-ups coloured and oriented in accordance with (B), showing the 3.0Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ. Due to local disorder in the bases, nucleotides were assigned by best-fit of the unwound sequence into the available electron density; one letter codes mark the assigned bases (right).

(D) Close-up of triphosphates bound at interrogation site. Close-up coloured and oriented in accordance with (B) and showing secondary structure elements coloured by domain as in (B). Interrogation site residues R225 and R678 are shown by sticks (grey, carbon; blue, nitrogen). The 3.0Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ are also shown.

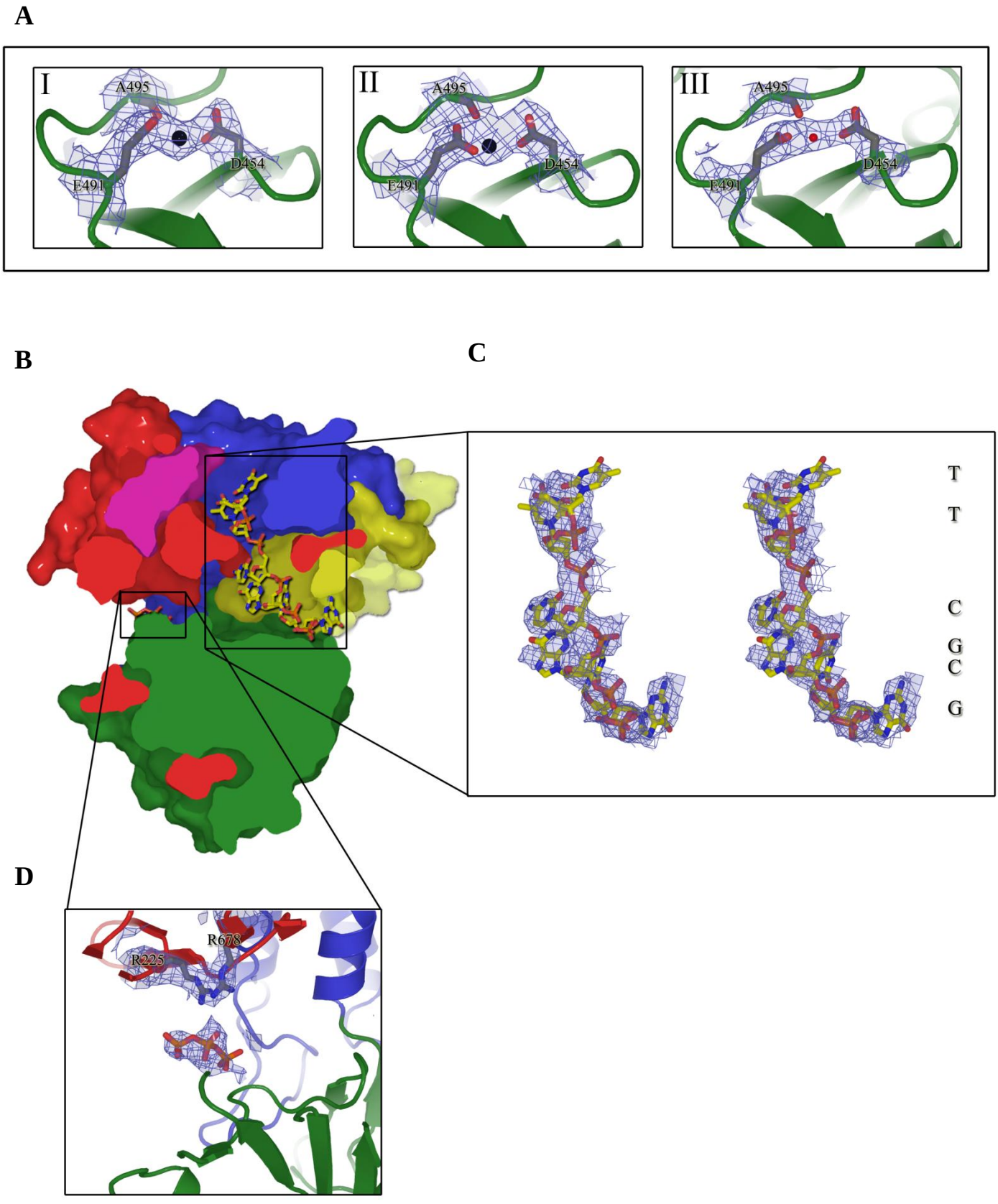
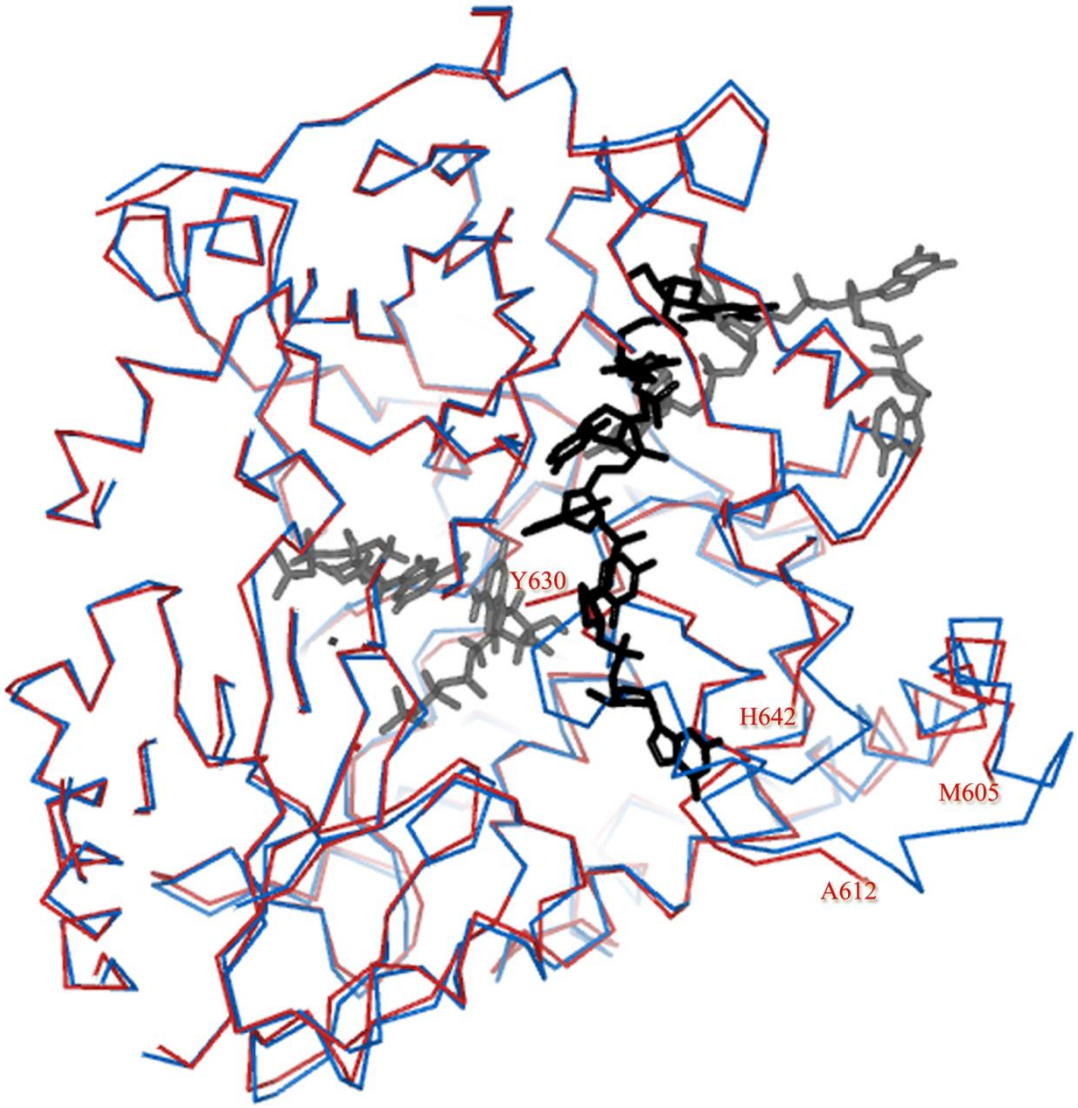

A
I
A495
D454
E491
II
A495
D454
E491
III
A495
D454
E491
B
C
T
T
C
G
C
G
D
R678
R225

Figure 3.14.

Figure 3.15. Structural integrity of ϕ6 RdRp core is maintained under different ligand binding conditions.

Molecule III of E634Q-Tri2T-AMPPCP-Mg^{2+}, with a disordered CTD and template bound within the template tunnel and active site, was superimposed on molecule II of WT-Tri4T-ATP-Mg^{2+} that has template bound at the entrance to the template tunnel on the surface of the polymerase and two bound ATPs. The protein backbones of the superimposed structures are depicted as ribbons (E634Q-Tri2T-AMPPCP-Mg^{2+}, red; WT-Tri4T-ATP-Mg^{2+}, blue). The RMSD between structures is relatively low at 0.9Å, highlighting that ϕ6 RdRp structural integrity is maintained under different ligand bound states. Missing residues 606-611 and 631-641 (see red labels on the figure) from the E634Q-Tri2T-AMPPCP-Mg^{2+} structure were not included in the superposition. The ligands for both structures are depicted by sticks (E634Q-Tri2T-AMPPCP-Mg^{2+}, black; WT-Tri4T-ATP-Mg^{2+}, grey). A black sphere indicates the position of a bound Mg^{2+} in WT-Tri4T-ATP-Mg^{2+} and a red sphere for bound H_2O at the non-catalytic ion site in E634Q-Tri2T-AMPPCP-Mg^{2+}.



3.3.3. E634Q-Tri4T-AMPPCP-Mg^{2+}

The crystal structure of E634Q-Tri4T-AMPPCP-Mg^{2+} was very similar to E634Q-Tri2T-AMPPCP-Mg^{2+}, although the evidence for bound hairpin template in the 2.9Å 2|Fo|-|Fc| SIGMAA electron density map of E634Q-Tri4T-AMPPCP-Mg^{2+} was much stronger. In molecule I, 5'-GCG-3' of the Tri4T hairpin (5'-TTTTCGCGTAGCG-3') was bound in the template tunnel with the 3'-end nucleotides in a position equivalent to the T1 and T2 nucleotides of the initiation complex (Butcher *et al*., 2001). In molecule II, the hairpin had penetrated one nucleotide deeper into the template tunnel, with 5'-AGCG-3' bound and the 3'-end guanidine positioned at the specificity pocket with the middle two nucleotides (5'-GC-3') at the active site; binding of template in the specificity pocket has been observed previously (Butcher *et al*., 2001; Salgado *et al*., 2004). In both these molecules Mg^{2+} had replaced Mn^{2+} at the non-catalytic ion site (**Fig. 3.16, A**) and there was no evidence for CTD disorder. In molecule III, six nucleotides of hairpin were bound (5'-TTTTCG-3') in position equivalent to that observed for E634Q-Tri2T-AMPPCP-Mg^{2+}, spanning the full length of the tunnel from entrance to exit (**Fig. 3.16, B**). Due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. Similarly, CTD residues 607-611 and 631-643 (including mutated residue Q634) were disordered in molecule III allowing the template to exit the polymerase at the C-terminus. In addition, no divalent cation was observed at the non-catalytic ion site (**Fig. 3.16, A**), suggesting again that it is lost during the CTD displacement.

A single AMPPCP was bound in each of the three molecules and remained associated with interrogation site residues K223, R225 and R268 (see **Fig. 3.16, D**); the base and sugar moieties were disordered as in E634Q-Tri2T-AMPPCP-Mg^{2+}. In molecule I however, the triphosphates also co-ordinate a second Mg^{2+} ion with the side chain of D324 and backbone

carbonyl group of V325 that sit on the opposite side of the NTP tunnel to the interrogation site (**Fig. 3.17, B**). Co-ordination of a second Mg^{2+} ion in this region has been observed previously in molecules I and II of the E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} complex (see section 3.3.8).

Figure 3.16. (overleaf). E634Q-Tri4T-AMPPCP-Mg^{2+}: CTD disordering and loss of the non-catalytic ion.

(A) The non-catalytic ion site for molecules I, II and III. The non-catalytic ion site is depicted within the palm domain (green, showing secondary structure elements). Side-chains of residues D454 and E491, and the backbone carbonyl group of A495 are depicted by sticks (grey, carbon; red, oxygen). Mg^{2+} is bound in molecules I and II represented by a large black sphere. In molecule III, water is bound at the site, represented by a small red sphere. The 2.9Å 2|Fo|-|Fc| SIGMAA electron density maps for non-catalytic ion sites are shown, contoured at 1.2σ (blue wire).

(B) Cross-section of molecular surfaces for E634Q-Tri4T-AMPPCP-Mg^{2+} molecule III. Surfaces are coloured by polymerase domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). C-terminal domain residues 607-611 and 631-643 are disordered opening a path for the exit of bound template. Molecular surfaces of disordered residues have been modelled onto the structure (semi-transparent yellow) and depict the CTD blocking the exit path. Six nucleotides of unwound Tri4T hairpin can be seen spanning the length of the template tunnel from entrance to exit (through the disordered region) with middle two nucleotides held at the active site; the unwound hairpin is depicted by sticks (cyan, carbon; blue, nitrogen; red, oxygen; orange, phosphorous). Triphosphates of bound AMPPCP can be seen at the entrance to the NTP tunnel (sticks: red, oxygen; orange, phosphorous).

(C) Stereo close-ups of unwound Tri4T hairpin. Close-ups coloured and oriented in accordance with (B), showing the 2.9Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ. Due to local disorder in the bases, nucleotides were assigned by best-fit of the unwound sequence into the available electron density; one letter codes mark the assigned bases (right).

(D) Close-up of triphosphates bound at interrogation site. Close-up coloured and oriented in accordance with (B) and showing secondary structure elements coloured by domain as in (B). Interrogation site residues R225 and R678 are shown by sticks (grey, carbon; blue, nitrogen). The 2.9Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ are also shown.

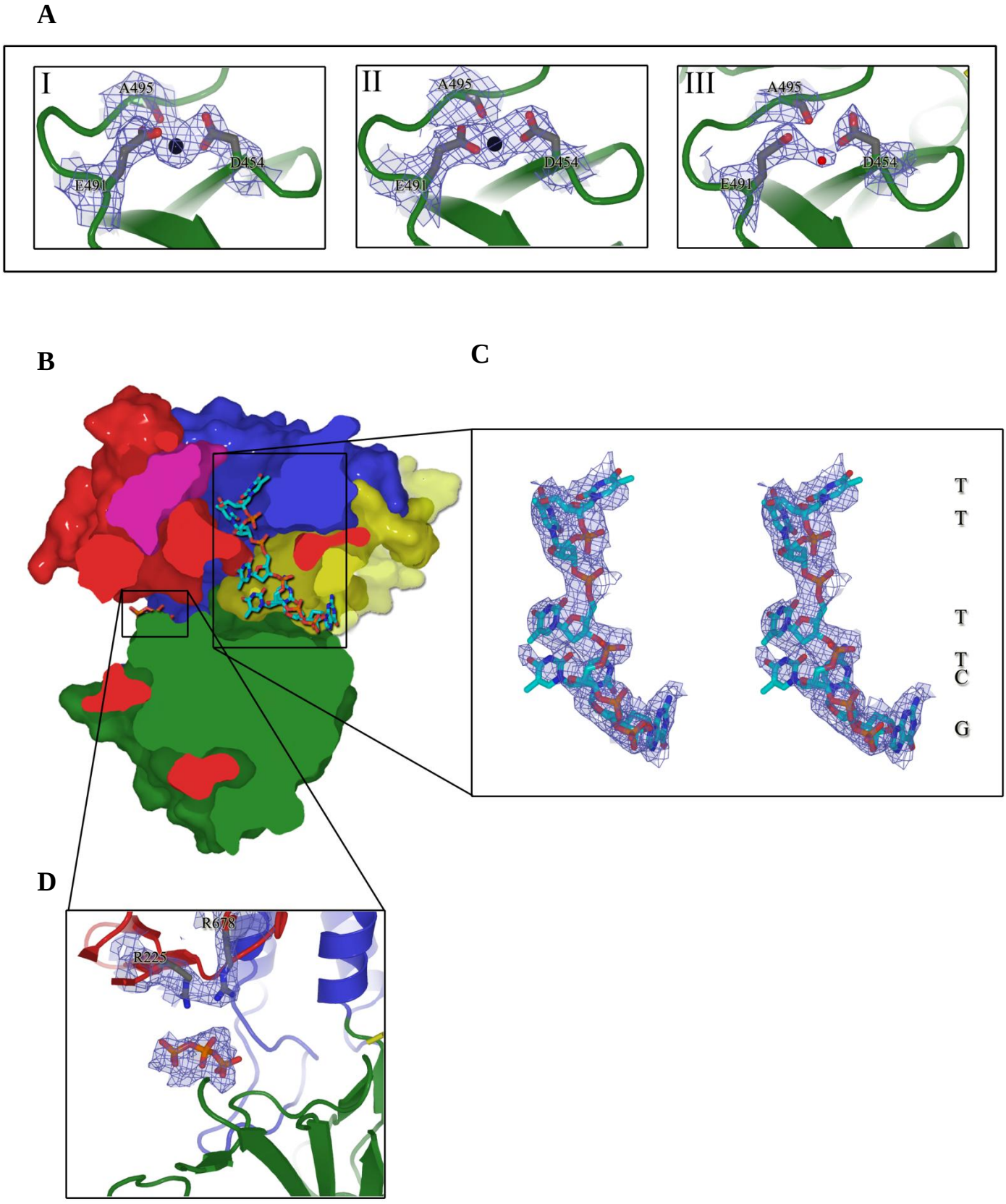

A
I
A495
D454
E491
II
A495
D454
E491
III
A495
D454
E491
B
C
T
T
T
T
C
G
D
R678
R225

Figure 3.16.

3.3.4. E634Q-Tetra2T-AMPPCP-Mg^{2+}

The crystal structure of E634Q-Tetra2T-AMPPCP-Mg^{2+} was essentially indistinguishable from E634Q-Tri2T-AMPPCP-Mg^{2+} other than the stronger evidence for bound hairpin and AMPPCP in the 3.0Å 2|Fo|-|Fc| SIGMAA electron density map (**Fig. 3.18, C and D**) and presence of a second Mg^{2+} bound in molecule I. The second Mg^{2+} ion is co-ordinated by the triphosphates of AMPPCP, side chain of D324 and backbone carbonyl group of V326 (**Fig. 3.17, A**). Interestingly, a second Mg^{2+} was also observed at this position for molecule I of E634Q-Tri4T-AMPPCP-Mg^{2+} (**Fig. 3.17, B**), and also molecules I and II of the E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} complex (**Fig. 3.8, A and B**). Two (5'-CG-3'), three (5'-GCG-3') and six (5'-TTCGCG-3') nucleotides of Tetra2T hairpin (5'-TTCGCGTAAGCG-3') were bound in the template tunnel for molecules I, II and III respectively, in the same position and orientation as E634Q-Tri2T-AMPPCP-Mg^{2+}; due to local disorder of DNA bases, nucleotides were assigned by best-fit of sequence into the available electron density. CTD residues 607-614 and 631-643 were disordered in molecule III allowing the exit of unwound hairpin from the active site (**Fig. 3.18, B**). Mg^{2+} had replaced Mn^{2+} at the non-catalytic ion site for molecules I and II, but no divalent cation was bound in the CTD-displaced molecule III (**Fig. 3.18, A**).

Figure 3.17. Second Mg^{2+} ion site for molecule I of E634Q-hairpin-AMPPCP-Mg^{2+} complexes.

(A) E634Q-Tetra2T-AMPPCP-Mg^{2+} (molecule I), labelled as Tetra2T (left). **(B)** E634Q-Tri4T-AMPPCP-Mg^{2+} (molecule I), labelled as Tri4T (left). For (A) and (B) secondary structure elements are shown and coloured according to the generic polymerase domain architecture: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Key amino acids of the interrogation site and non-catalytic ion site are labelled and shown in stick representation (grey, carbon; blue, nitrogen; red, oxygen; Mg^{2+} in the standard downward position is shown as a black sphere). Triphosphates of AMPPCP (sticks: red, oxygen; orange, phosphorous) and priming platform residue Y630 (sticks: yellow, carbon; red, oxygen) are also shown. The second Mg^{2+} (black sphere) is co-ordinated by the side-chain of D324 and backbone carbonyl of V324 (sticks: grey, carbon; blue, nitrogen; red, oxygen). The 3.0Å (Tetra2T) and 2.9Å (Tri4T) 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are shown for completeness.

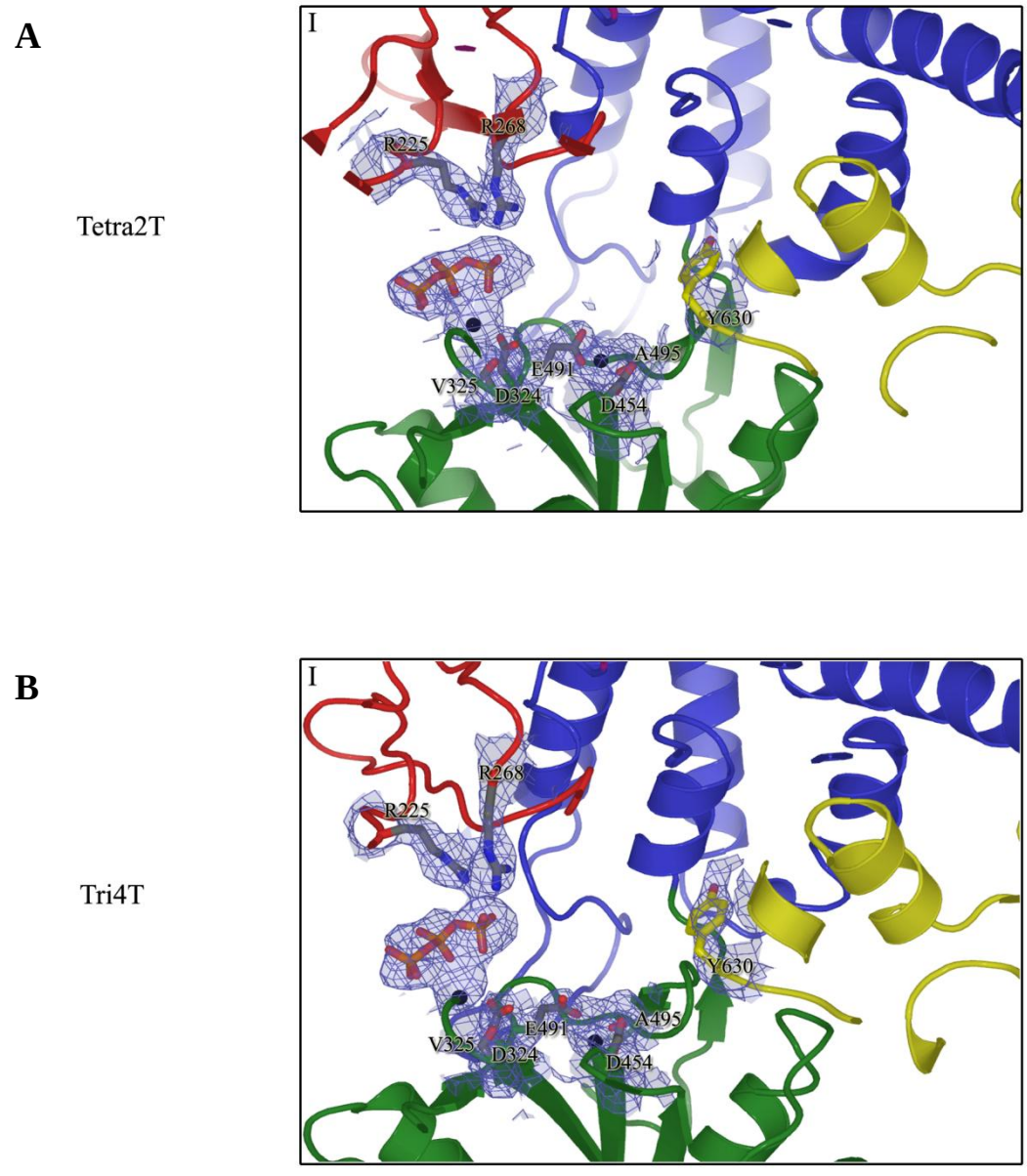


Figure 3.18. (overleaf). E634Q-Tetra2T-AMPPCP-Mg^{2+}: CTD disordering and loss of the non-catalytic ion.

(A) The non-catalytic ion site for molecules I, II and III. The non-catalytic ion site is depicted within the palm domain (green, showing secondary structure elements). Side-chains of residues D454 and E491, and the backbone carbonyl group of A495 are depicted by sticks (grey, carbon; red, oxygen). Mg^{2+} is bound in molecules I and II represented by a large black sphere. In molecule III, no ion is bound. The 3.0Å 2|Fo|-|Fc| SIGMAA electron density maps for non-catalytic ion sites are shown, contoured at 1.2σ (blue wire).

(B) Cross-section of molecular surfaces for E634Q-Tetra2T-AMPPCP-Mg^{2+} molecule III. Surfaces are coloured by polymerase domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). C-terminal domain residues 607-614 and 631-643 are disordered opening a path for the exit of bound template. Molecular surfaces of disordered residues have been modelled onto the structure (semi-transparent yellow) and depict the CTD blocking the exit path. Six nucleotides of unwound Tetra2T hairpin can be seen spanning the length of the template tunnel from entrance to exit (through the disordered region) with middle two nucleotides held at the active site; the unwound hairpin is depicted by sticks (green, carbon; blue, nitrogen; red, oxygen; orange, phosphorous). Triphosphates of bound AMPPCP can be seen at the entrance to the NTP tunnel (sticks: red, oxygen; orange, phosphorous).

(C) Stereo close-ups of unwound Tetra2T hairpin. Close-ups coloured and oriented in accordance with (B), showing the 3.0Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ. Due to local disorder in the bases, nucleotides were assigned by best-fit of the unwound sequence into the available electron density; one letter codes mark the assigned bases (right).

(D) Close-up of triphosphates bound at interrogation site. Close-up coloured and oriented in accordance with (B) and showing secondary structure elements coloured by domain as in (B). Interrogation site residues R225 and R678 are shown by sticks (grey, carbon; blue, nitrogen). The 3.0Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ are also shown.

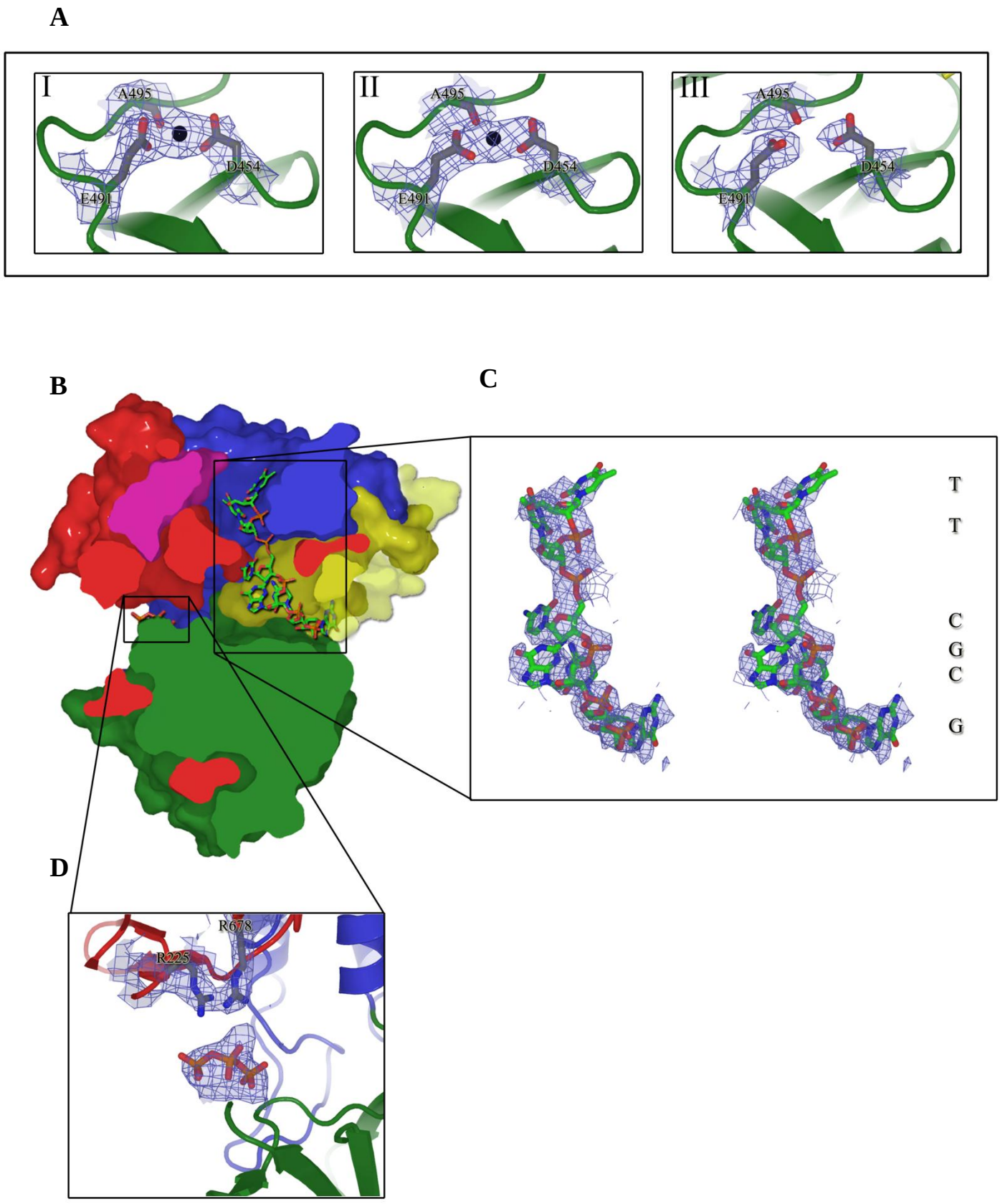

A
I
A495
D454
E491
II
A495
D454
E491
III
A495
D454
E491
B
C
T
T
C
G
C
G
D
R678
R225

Figure 3.18.

3.3.5. E634Q-Tetra4T-AMPPCP-Mg^{2+}

The crystal structure of E634Q-Tetra4T-AMPPCP-Mg^{2+} was again very similar to the three previously reported E634Q-hairpin-AMPPCP-Mg^{2+} structures. The 3.4Å 2|Fo|-|Fc| SIGMAA electron density map revealed AMPPCP bound at K223, R225 and R268 for all three molecules (base and sugar disordered, see **Fig. 3.19, D**) and that Mg^{2+} had replaced Mn^{2+} at the non-catalytic ion site in molecules I and II, with no divalent cation bound in molecule III (**Fig. 3.19, A**). CTD residues 606-612 and 631-647 (including mutated residue Q634) were disordered in molecule III and six nucleotides (5'-TTTTCG-3') of unwound Tetra4T hairpin (5'-TTTTCGCGTAAGCG-3') were bound in the polymerase spanning the length of the template tunnel from entrance to exit at the 'opened' C-terminus (**Fig. 3.19, B and C**). In molecule I, 5'-CG-3' was bound with the 3'-end held at the active site in same position and orientation described for molecule I of E634Q-Tri2T-AMPPCP-Mg^{2+} and E634Q-Tetra2T-AMPPCP-Mg^{2+}. However, in molecule II, 5'-GCG-3' was bound in an orientation not observed in any of the other hairpin soaked structures. The 3'-end was held in the specificity pocket, but there was no evidence of electron density for the base component of the penultimate cytidine facing the active site and Y630 of the priming platform. The base component was instead built in the available weak electron density at a position 180^{o} in the opposite direction to its standard base pairing position, where there was evidence for a stabilising interaction with the N4 of the cytidine and backbone carbonyl group of a nearby residue S150; however, density for this base was particularly poor and precise positioning could not be determined from this dataset. Given the disorder observed for the base components of bound hairpin in each molecule, especially at the active site in the absence of Watson-Crick base pairing with incoming NTPs, nucleotides were assigned by best-fit into the available electron density.

Figure 3.19. (overleaf). E634Q-Tetra4T-AMPPCP-Mg^{2+}: CTD disordering and loss of the non-catalytic ion.

(A) The non-catalytic ion site for molecules I, II and III. The non-catalytic ion site is depicted within the palm domain (green, showing secondary structure elements). Side-chains of residues D454 and E491, and the backbone carbonyl group of A495 are depicted by sticks (grey, carbon; red, oxygen). Mg^{2+} is bound in molecules I and II represented by a large black sphere. In molecule III, no ion is bound. The 3.4Å 2|Fo|-|Fc| SIGMAA electron density maps for non-catalytic ion sites are shown, contoured at 1.2σ (blue wire).

(B) Cross-section of molecular surfaces for E634Q-Tetra4T-AMPPCP-Mg^{2+} (molecule III). Surfaces are coloured by polymerase domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). C-terminal domain residues 606-612 and 631-647 are disordered opening a path for the exit of bound template. Molecular surfaces of disordered residues have been modelled onto the structure (semi-transparent yellow) and depict the CTD blocking the exit path. Six nucleotides of unwound Tetra4T hairpin can be seen spanning the length of the template tunnel from entrance to exit (through the disordered region) with middle two nucleotides held at the active site; the unwound hairpin is depicted by sticks (magenta, carbon; blue, nitrogen; red, oxygen; orange, phosphorous). Triphosphates of bound AMPPCP can be seen at the entrance to the NTP tunnel (sticks: red, oxygen; orange, phosphorous).

(C) Stereo close-ups of unwound Tetra4T hairpin. Close-ups coloured and oriented in accordance with (B), showing the 3.4Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ. Due to local disorder in the bases, nucleotides were assigned by best-fit of the unwound sequence into the available electron density; one letter codes mark the assigned bases (right).

(D) Close-up of triphosphates bound at interrogation site. Close-up coloured and oriented in accordance with (B) and showing secondary structure elements coloured by domain as in (B). Interrogation site residues R225 and R678 are shown by sticks (grey, carbon; blue, nitrogen). The 3.4Å 2|Fo|-|Fc| SIGMAA electron density maps (blue wire) contoured at 1.2σ are also shown.

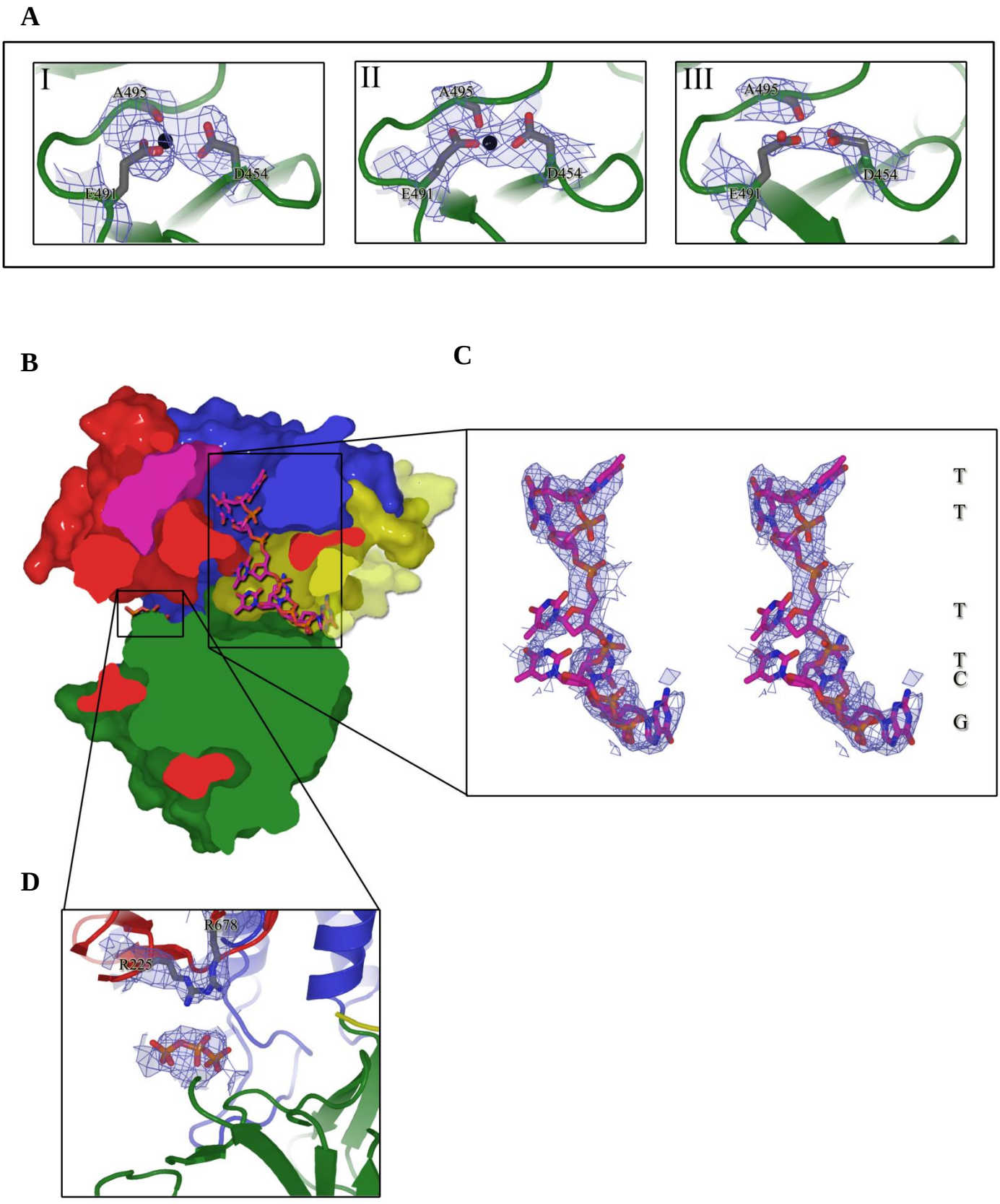

A
I
A495
D454
E491
II
A495
D454
E491
III
A495
D454
E491
B
C
T
T
T
T
C
G
D
R678
R225

Figure 3.19.

3.3.5. E634Q-Tetra2T-coxtal

In order to determine whether presence of divalent cations at the non-catalytic ion site were preventing CTD displacement in molecule I and II of the E634Q-hairpin-AMPPCP-Mg^{2+} complexes, co-crystals of E634Q with the Tetra2T hairpin (5'-TTCGCGTAAGCG-3') were grown in the presence of 5 mM EDTA (shown to strip out bound co-purifying Mn^{2+}, see thermal denaturation experiments, section 3.1 and **Table 3.1**) and were soaked with additional Tetra2T hairpin prior to data collection. The 3.1Å 2|Fo|-|Fc| SIGMAA electron density map for E634Q-Tetra2T-coxtal revealed four nucleotides of hairpin (5'-AGCG-3') bound in an all three molecules of the ASU. This density was not very well defined suggesting that bound template in each molecule was disordered and/or not fully occupied, especially with regard to molecule III (**Fig. 3.20**). In molecules I and II, the 3'-end of the bound nucleotide sits in the specificity pocket with the next two nucleotides held at the active site (**Fig. 3.20**). In molecule III, the 3'-end and penultimate nucleotides sit in a position equivalent to the T1 and T2 nucleotides of the published initiation complex (Butcher *et al*., 2001) and has not reached the specificity pocket (**Fig. 3.20**).

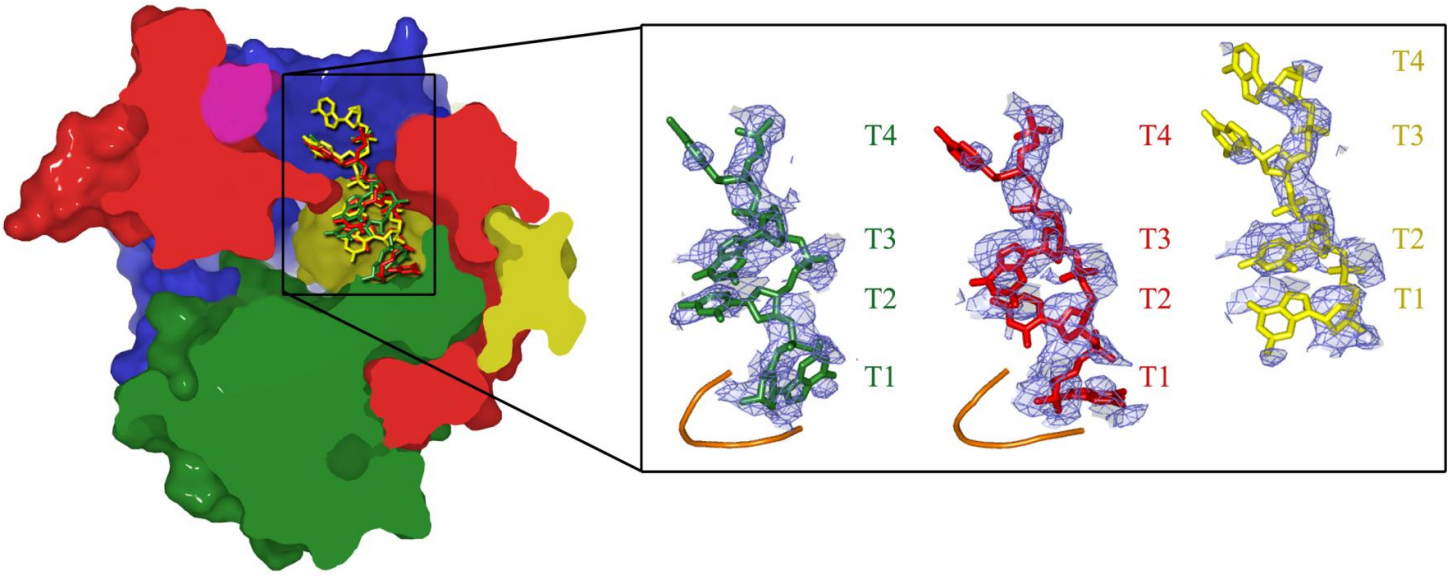
No divalent cations were observed at the non-catalytic ion site in any of the three molecules as expected, and there was no evidence for CTD displacement by the bound hairpins. This suggests that the CTD remains ordered when the non-catalytic ion site is empty or occupied by Mg^{2+} (as seen for molecules I and II of all E634Q-hairpin-AMPPCP-Mg^{2+} complexes). In fact, the thermal denaturation experiments (**Table 3.1** and **Fig. 3.1**) show that removal of co-purifying Mn^{2+} (by 5 mM EDTA) or replacement of the Mn^{2+} by Mg^{2+} (at a concentration of 10 mM), structurally stabilises E634Q RdRp. These findings suggest that binding of Mn^{2+} at the non-catalytic ion site is required for CTD displacement, given that the template is stalled about the active site in its absence. If this is true, lack of a divalent cation in molecule III of

all E634Q-hairpin-AMPPCP-Mg^{2+} complexes would reflect the loss of the bound 'destabilising' Mn^{2+} during the structural rearrangement of the CTD and shifting down of template.

This considered, CTD displacement may be less energetically favourable in the $P3_2$ crystal form observed for the E634Q-Tetra2T-coxtal structure given the greater number of crystal contacts about each molecule and thereby reduced molecular flexibility within the crystal. CTD displacement was not observed for example with molecules I and II of the $P2_1$ form for all E634Q-hairpin-AMPPCP-Mg^{2+} complexes, possibly for the same reason. Unfortunately, E634Q co-crystals with hairpin could only be obtain in the $P3_2$ crystal form and grew only in the presence of 5 mM EDTA, suggesting that Mn^{2+}-driven destabilisation and opening of the CTD with concomitant threading of hairpin through the polymerase, were preventing crystal growth in the presence of hairpin. Co-crystals of WT and E491Q mutant polymerases with short linear DNA and RNA oligonucleotides in the absence of 5 mM EDTA have been reported previously (Butcher *et al*., 2001; Salgado *et al*., 2004).

Figure 3.20. Template bound in template tunnel for E634Q-Tetra2T-Coxtal (mol I-III).

Left panel, cross-section of molecular surfaces coloured by polymerase domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). Bound oligonucleotides from molecules I, II and III are superimposed and depicted using stick presentation (green, molecule I; red, molecule II; yellow, molecule III). ***Right panel***, close-up of the disordered oligonucleotides coloured and oriented in accordance with left panel. The 3.1Å 2|Fo|-|Fc| SIGMAA electron density maps contoured at 1.2σ are shown for the separate oligonucleotides (blue wire). The polypeptide chain for 629QYKW632 of the priming platform (defining the specificity pocket), is shown in orange for molecules I and II that enter this pocket. Oligonucleotides are labelled in accordance with (Butcher *et al*., 2001) with T1 (3'-end), T2 (penultimate) and so on. Assigned bases are 5'-ACGC-3' for all molecules.



3.3.6. Trypsin-digested WT RdRp ($P2^{T1}$)

To obtain further information on the C-terminal domain, a set of CTD deletion constructs were designed by Peter Sarin as part of our collaboration with the University of Helsinki (see Chapter 2, section 2.1.6). These constructs remove the entire CTD, terminating at or near the putative "hinge" region (residues 595-608), whereas some extend further into the main body of the enzyme, also removing parts of the thumb domain. None of the deletion constructs yielded a viable polymerase suggesting that the CTD is crucial for the correct folding of the protein (Sarin *et al*., 2010; to be submitted). However, a biochemically active C-terminal truncation, $P2^{T1}$ (cleaved at R607) was obtained by protease digestion of the WT ϕ6 RdRp with trypsin. Surprisingly, crystallization of $P2^{T1}$ revealed that the protein was indistinguishable from the previously published structure for the full-length WT ϕ6 RdRp (Butcher *et al*., 2001) except for a nick in the polypeptide chain at position 607, where no evidence for electron density is seen for residues 603-609 in all three molecules of the ASU (**Fig. 3.21**). Superimposing the 3.3Å $P2^{T1}$ structure with the published full-length 2.0Å WT apo structure demonstrated significant differences in electron density about this surface-exposed loop region (**Fig. 3.21**). Interestingly, the E634 – K144 salt bridge (that is lost in the E634Q mutant, **Fig. 3.13**) is still very well-defined in the $P2^{T1}$ structure suggesting that it plays a key role in anchoring the CTD to the main body of the polymerase.

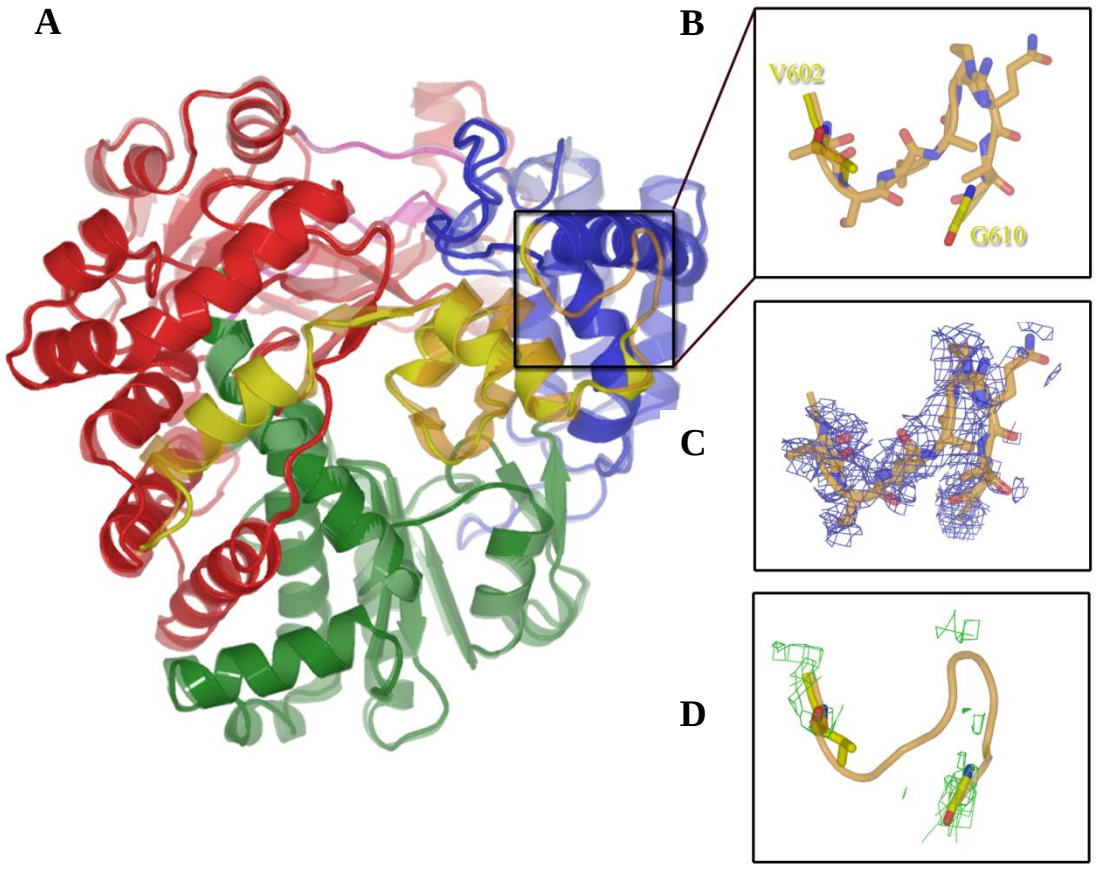
Figure 3.21. $P2^{T1}$ superimposed with WT RdRp.

(A) $P2^{T1}$ has been superimposed with WT RdRp [(Butcher *et al*., 2001); PDB accession no. 1HHS] and both are represented by secondary structure elements. P2T1 is coloured by domain: red, fingers domain (amino acids 1–30, 104–276, 333–397); green, palm domain (277–332, 398–517); blue, thumb domain (37–91, 518–600); yellow, C-terminal domain (604–664); pink, connecting chains (31–36, 92–103). WT RdRp is coloured as follows: light red, fingers domain; light green, palm domain; light blue, thumb domain; orange, C-terminal domain; light pink, connecting chains.

(B) Close-up of surface-exposed loop (603-609) coloured and oriented in accordance with (A). For $P2^{T1}$, amino acids V602 and G610 either side of the nicked region are labelled and depicted with sticks (yellow, carbon; blue, nitrogen; red, oxygen). For WT RdRp, V602 and G610 are depicted with sticks alongside residues A603, S604, M605, A606, R607, Q608 and A609, that are missing in $P2^{T1}$ (orange, carbon; blue, nitrogen; red, oxygen).

(C) Close-up of surface-exposed loop for WT RdRp coloured and oriented as in (B) and showing the 2.0Å 2|Fo|-|Fc| SIGMAA electron density map (contoured at 1.2σ) for WT RdRp in this region (blue wire).

(D) Close-up of surface-exposed loop for $P2^{T1}$ coloured and oriented as in (B) and showing the 3.3Å 2|Fo|-|Fc| SIGMAA electron density map (contoured at 1.2σ) for this region (green wire). WT RdRp is shown as a backbone ribbon (orange) to indicate the position of the loop as depicted in (B) and (C). $P2^{T1}$ shows little to no density for this loop region.



3.3.7. Transition to Elongation – Summary

In order for the template to transition to a position suitable for elongation (with the T3 nucleotide held at the active site ready to base pair with the D3 NTP), the CTD (residues 604-664) that blocks the exit of the polymerase must be displaced. The crystal structure of trypsin-digested $P2^{T1}$ (cleaved at residue 607) revealed that the protein was indistinguishable from the previously published structure for full-length WT ϕ6 RdRp (Butcher *et al*., 2001) except for a nick in the polypeptide chain at position 607, where no evidence for electron density is seen for residues 603-609 in all three molecules of the ASU (**Fig. 3.21, C**). Additional biochemical studies carried out by Peter Sarin at the University of Helsinki demonstrated that the C-terminal domain of $P2^{T1}$ remains attached to the main body of the polymerase during affinity column purification under native conditions and could only be removed when stringent denaturing conditions were applied, such as 0.05 % sodium dodecyl sulphate and 6 M urea (Sarin *et al*., 2010; to be submitted). Upon removal, the main body of the polymerase rapidly degraded, whereas the CTD remained stable in solution. In addition, a set of C-terminal deletions were designed in an attempt to express ϕ6 RdRp without the CTD. None of the deletion constructs yielded a viable polymerase suggesting that the C-terminal domain is crucial for the correct folding of the protein (Sarin *et al*., 2010; to be submitted). These data highlight the importance of the CTD and demonstrate that there are strong interactions that fix it to the main body of the polymerase; these strong interactions must be destabilised during the transition to elongation.

The crystal structure of E634Q RdRp revealed Mn^{2+} bound at the non-catalytic sites of all three molecules in the ASU (**Fig. 3.12**) confirming that Mn^{2+} co-purifies with the polymerase as suggested by the thermal denaturation experiments (section 3.1, **Table 3.1** and **Fig. 3.1**). The E634Q apo structure also confirmed the loss of a salt bridge (**Fig. 3.13**) connecting the

CTD to the main body of the polymerase (E634 – K144) that provides the mutant with a 2.5-fold higher polymerisation activity (Sarin *et al*., 2009), presumably due to a lower energy barrier for CTD displacement. Furthermore, consistent with the WT complex structures, the Mn^{2+} could be replaced by Mg^{2+} by soaking with 10 mM $MgCl_2$. Although the E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} complex demonstrated a reduced template entry with bound Mg^{2+} (like WT complexes under similar conditions), the E634Q-hairpin-AMPPCP-Mg^{2+} complexes described here have retained their ability to bind template in the tunnel even in the presence of Mg^{2+} at the non-catalytic ion site for molecules I and II (**Fig. 3.13, A; Fig. 3.16, A; Fig. 3.18, A; Fig. 3.19, A**). Thus, the more flexible E634Q RdRp can bind template with either Mg^{2+} or Mn^{2+} bound at the non-catalytic ion site, unlike WT RdRp that requires the presence of Mn^{2+} for sufficient template entry into the active site.

For all E634Q-hairpin-AMPPCP-Mg^{2+} complexes, the DNA hairpins were unwound and threaded through the template tunnel so that up to three, four and six DNA bases could be modelled into molecules I, II and III, respectively. Thus, the best ordered DNA density was seen in the least constrained molecule, III (**Fig. 3.13, C; Fig. 3.16, C; Fig. 3.18, C; Fig. 3.19, C**). In the presence of template this molecule also showed evidence for CTD movement: in all E634Q-hairpin-AMPPCP-Mg^{2+} complexes the region spanning residues 606-614 and 631-647 were highly disordered, with little or no evidence of electron density. The disordering is apparently induced by the oligonucleotide which has penetrated deeper down into the template tunnel, passing the active site and mimicking the switch to elongation that places the T3 nucleotide in the active site (**Fig. 3.13, B; Fig. 3.16, B; Fig. 3.18, B; Fig. 3.19, B**). In molecules I and II which possess more rigid crystal contacts (see Chapter 2, section 2.5.1) and do not show CTD disorder, a Mg^{2+} ion is seen at the non-catalytic ion site coordinated by D454 and E491 in the standard 'downward' position (**Fig. 3.13, A; Fig. 3.16, A; Fig. 3.18,**

A; Fig. 3.19, A). However, there is little electron density at this site in molecule III, suggesting that disordering of the CTD, and concomitant movement of the template, has released the bound divalent ion.

Co-crystallising E634Q RdRp with the Tetra2T hairpin in the presence of 5 mM EDTA demonstrated that in the absence of Mn^{2+} bound at the non-catalytic ion site, the unwound hairpin could only reach the specificity pocket and not displace the CTD to allow its exit from the polymerase (**Fig. 3.20**). Given that Mn^{2+} was shown to lower the structural integrity of E634Q RdRp, providing it with a much greater structural fluidity (see thermal denaturation results in section 3.1, **Table 3.1** and **Fig. 3.1**), it is likely that Mn^{2+} binding at the non-catalytic ion site is required to lower the energy barrier for the CTD displacement and overcome the strong interactions of the CTD with the main body of the protein. If this is true, the absence of a bound divalent cation at the non-catalytic ion site in molecule III of all E634Q-hairpin-AMPPCP-Mg^{2+} complexes, where CTD displacement has occurred, would indeed reflect the loss of the 'destabilising' Mn^{2+} by a transient weakening of the high affinity site during the Mn^{2+}-driven CTD structural rearrangement and concomitant movement of bound template (**Fig. 3.13; Fig. 3.16; Fig. 3.18; Fig. 3.19**).

In order to confirm this inferred role of Mn^{2+} in the transition to elongation during ϕ6 RdRp-catalysed polymerisation, a series of biochemical experiments were carried out by Minna Poranen at the University of Helsinki and were published alongside crystal structures of WT and E634Q complexes and the thermal denaturation experiments reported herein (Wright *et al*., 2010; submitted). A brief outline of the biochemical results is given below.

Previous studies have indicated that the initial phase of the polymerization reaction is sensitive to heparin, which competes with template at the template tunnel; the elongation reaction however is heparin resistant (Ackermann and Padmanabhan, 2001; Yang *et al*., 2003). To determine the precise stage in the polymerization reaction that accounts for this sensitivity, standard polymerization reactions were carried out using a set of template–NTP combinations designed to stall the reactions at different stages. These experiments reveal that the transition from heparin sensitivity to heparin resistance occurs when the third nucleotide (D3) is added to the dinucleotide initiation complex (Wright *et al*., 2010; submitted), presumably because the template becomes locked into the polymerase when the CTD is displaced and cannot be released. Evidence for template release after catalysis of only the dinucleotide (before the transition to elongation) has been presented previously [(Salgado *et al*., 2004), PDB accession no. 1UVK]. The heparin trap assay therefore provides a way to dissect the initiation (dinucleotide state) and elongation phases (trinucleotide state and beyond) and was used to identify which stages are stimulated by external Mn^{2+} ions.

In the absence of external $MnCl_2$ the polymerisation reaction was unable to reach the heparin resistant elongation state (**Fig. 3.22, A: lanes 4 and 6**; Wright *et al*., 2010; submitted). This implies that Mn^{2+} is essential for the initial phase of polymerization, prior to the formation of the nascent trinucleotide product. The role of Mn^{2+} ions in the early phase of the polymerization reaction was analyzed in more detail by separating the products of the initiation reaction (template and [γ-P^{32}]-labelled GTP substrate) on a high resolution urea containing polyacrylamide gel (**Fig. 3.22, B**). Under optimal conditions (2 mM $MnCl_2$, 5 mM $MgCl_2$) a trinucleotide product (pppGpGpG) was synthesized as expected (**Fig. 3.22, B: lane 1**; Wright *et al*., 2010; submitted). However, if Mn^{2+} ions were omitted from the mixture, only the initiation complex was produced (**Fig. 3.22, B: lane 2**). In the absence of

both Mn^{2+} and Mg^{2+}, no catalysis occurred (**Fig. 3.22, B: lane 4**; Wright *et al*., 2010; submitted). Furthermore, Mn^{2+} ions could support the formation of trinucleotides in the absence of Mg^{2+} ions (**Fig. 3.22, B: lane 3**; Wright *et al*., 2010; submitted). This implies that the addition of Mn^{2+} ions is not required for the formation of the initiation complex (dinucleotide), but is essential for the subsequent nucleotide addition (trinucleotide).

Furthermore, the effect of Mg^{2+} and Mn^{2+} on RNA chain elongation after the formation of a trinucleotide product was analysed (Wright *et al*., 2010; submitted). The results revealed that Mn^{2+} was not required to extend the trinucleotide product if Mg^{2+} was present, however the rate of elongation was reduced to approximately one forth (23 nt/s versus 6 nt/s) if the elongation mixture did not contain Mn^{2+} (Wright *et al*., 2010; submitted).

Taken together, these biochemical observations are in direct agreement with the crystallographic and thermal denaturation data reported here and confirm that binding of Mn^{2+} at the non-catalytic ion site is essential to structurally destabilise the polymerase and lower the energy barrier for CTD displacement.

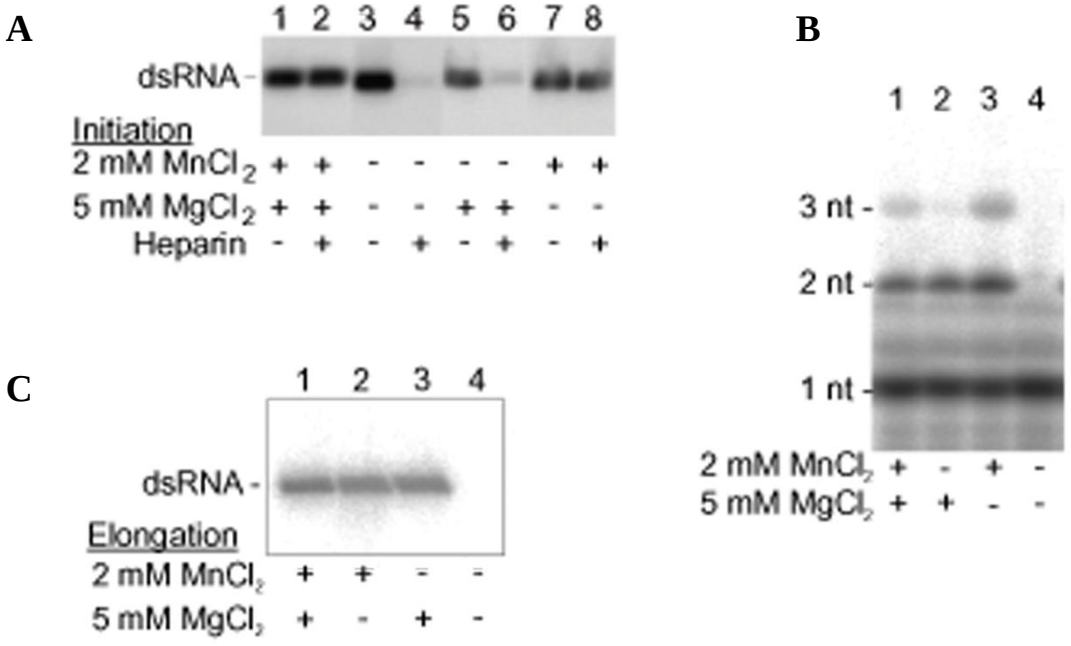
Figure 3.22. Summary of biochemical results for the role of Mn^{2+} in ϕ6 RdRp catalysed polymerisation.

(A) and (B) Mn^{2+}-dependence of the initial phase of polymerization reaction.

(A) Heparin trap assay. ϕ6 RdRp was incubated with ssRNA template $s\Delta^{+}_{CCC}$ [....UCUCUCCC-3', lacking internal genomic S segment region 593–2830; (Makeyev and Bamford, 2000a)] and GTP substrate (for initiation and catalysis of 3 nucleotides: pppGpGpG) under varying divalent cation conditions for 20 min. After the 20min period, missing NTP substrates were added with or without heparin (9.1 mg/mL, Sigma; present in lanes 2, 4, 6 and 8), that blocks formation of new initiation complexes. Divalent cations were then added up to optimal concentrations (2 mM $MnCl_2$ and 5 mM $MgCl_2$) into the new 'elongation reaction' mixture. Absence (-) or presence (+) of divalent cations during the initiation phase and/or heparin (after 20min) are indicated beneath the gels.

(B) Phosphoimaging of initiation products. ϕ6 RdRp was incubated with ssRNA template $s\Delta^{+}_{CCC}$ [....UCUCUCCC-3', lacking internal genomic S segment region 593–2830; (Makeyev and Bamford, 2000a)] and [γ-P^{32}]GTP substrate (instead of GTP) for initiation and catalysis of 3 nucleotides (pppGpGpG) under varying divalent cation conditions for 20 min. The reaction products of the initiation mixture were analyzed by phosphoimaging of the resulting 20% polyacrylamide-urea gel. The mobility of mono-, di- and trinucleotides is indicated on the right. Absence (-) or presence (+) of divalent cations are indicated beneath the gel

(C) Mn^{2+}-dependence after transition to elongation. ϕ6 RdRp was incubated with ssRNA template $s\Delta^{+}_{CCC}$ (as in **A** and **B**) and GTP under optimal divalent cation (2 mM $MnCl_2$, 5 mM $MgCl_2$) conditions for 20 min. Subsequently reactions were purified through desalting spin columns (Pierce), and the formation of new initiations was suppressed by heparin. Elongation was then allowed to proceed under a variety of divalent cation conditions (indicated below). The reaction products were analyzed by standard agarose gel electrophoresis and autoradiography. [Adapted from Wright *et al*., 2010; submitted].



3.3.8. NTPs carry Mg^{2+} to the active site – Summary

In addition to providing direct evidence for CTD displacement and template exit, the electron density maps for the different E634Q-hairpin-AMPPCP-Mg^{2+} complexes also demonstrated that incoming NTPs can bring Mg^{2+} ions with them to the active site for use in catalysis. AMPPCP was bound at the interrogation site for all E634Q-hairpin-AMPPCP-Mg^{2+} structures, with the backbone triphosphates coordinated by positively charged residues K223, R225 and R268. In molecule I of two of the E634-haipin-AMPPCP-Mg^{2+} complexes (E634Q-Tri4T-AMPPCP-Mg^{2+} and E634Q-Tetra2T-AMPPCP-Mg^{2+}) a second Mg^{2+} ion was bound (in addition to the Mg^{2+} at the non-catalytic ion site), coordinated by the triphosphates of AMPPCP, the side chain of D324 and the backbone carbonyl group of V325 (**Fig. 3.17, A and B**). Interestingly, a second Mg^{2+} was also co-ordinated in this region by the backbone carbonyl of S326 and side chain of D327 in molecules I and II of the E634Q-5'-UUCCCC-3'-GTP-Mg^{2+} complex (**Fig. 3.8, A and B**). A closer look at this area revealed a negatively charged pocket that sits on the opposite side of the NTP tunnel to the positively charged interrogation site, and is formed by D324, D327 and D329, and the backbone carbonyl groups of D324, V325 and S326. The second Mg^{2+} ions co-ordinated at this site are positioned ~5Å away from one of the catalytic Mg^{2+}-sites previously identified in the ϕ6 RdRp initiation complex structure (Butcher *et al*., 2001). These data suggest that Mg^{2+} is drawn into the polymerase by NTPs and can be stabilised about this newly identified negatively charged pocket during preinitiation stages before being drawn to a catalytic position upon NTP movement from the interrogation site to base pair with template.

3.3.9. Non-catalytic ion sites of other vRdRps

Previous analyses of viral polymerases (Poranen *et al*., 2008b) demonstrated that the non-catalytic ion site is conserved across a large number of viral RNA polymerases (Chapter 1, section 1.1.3.2., C and **Table 3.2**). Our new data demonstrate that Mg^{2+} can replace the non-catalytic Mn^{2+} if ϕ6 RdRp crystals are soaked with 10 mM $MgCl_2$. An analysis of the crystallization conditions for other RdRp structures revealed that high concentrations of different Mg^{2+} salts (up to 300 mM) were present during crystallization (**Table 3.2**) and in many cases, where no divalent cations were present, no non-catalytic ion could be modelled in the structure (e.g. WNV, FMDV, NV and RV). This questions the authenticity of the Mg^{2+} ion in the previously published structures and suggests that Mn^{2+} could in fact be the natural ion for this location; removal or replacement of Mn^{2+} could have occurred during purification or crystallisation, in the presence of EDTA or high concentration of divalent salts. In this work we demonstrate that Mn^{2+} bound at this site in ϕ6 RdRp plays an important role in controlling the switch to elongation and although Mg^{2+} can occupy this site, it cannot functionally replace the natural ion (Mn^{2+}). It is likely that the function of this non-catalytic ion site in controlling the switch to elongation is broadly conserved in viral RdRps.

Table 3.2. Purification and crystallization conditions for viral RdRps with non-catalytic divalent ions. The table is an extended version of that reported in (Poranen *et al*., 2008b) and was published alongside structural and biochemical data in (Wright *et al*., 2010; submitted). The new columns are highlighted by semi-transparent blue shading. West Nile Virus (WNV), Foot-and-Mouth Disease Virus (FMDV), Rabbit Haemorrhagic Disease Virus (RHDV), Norwalk Virus (NV), Dengue Virus (DENV), Poliovirus (PV) and Reovirus (RV).

Virus	PDB ID	Published ion	Purification and storage conditions	Divalent ions in crystallization conditions	Comments
WNV	2HCS	No ion	Protease inhibitor tablet (EDTA)	-	-
	2HCN	Ca^{2+}		200 mM $Ca(C_2H_3O_2)_2$	From map and coordination geometry it is unlikely to be Ca^{2+}
	2HFZ	Mg^{2+}		300 mM $MgCl_2$	Map and coordination geometry consistent with Mg^{2+}
FMDV	1U09	No ion	1 mM EDTA	-	-
	1WNE	Mg^{2+}		2 mM $MgCl_2$ 200 mM Mg-acetate	Structure factors not available
RHDV	1KHV	(H_2O)	0.25 mM EDTA + 10 mM $MgCl_2$	5 mM $LuCl_3$	Lu^{3+} ion 6Å from Mn^{2+} site; evidence for Mg^{2+} in place of published H_2O at Mn^{2+} site
	1KHW	Density difficult to interpret		5 mM $LuCl_3$ + 50 mM $Mg(NO_3)_2$, 18 mM $MnCl_2$	Poor density in region of active site makes interpretation difficult
NV	1SH0, 1SH2	No ion	0.25 mM EDTA	-	-
	1SH3	Mg^{2+}		20 mM $MgSO_4$	Map and coordination geometry consistent with Mg^{2+}
DENV	2J7U	Mg^{2+}	1 mM EDTA	1 mM EDTA, up to 200 mM $MgSO_4$ or $Mg(C_2H_3O_2)_2$	Map and coordination geometry consistent with Mg^{2+}
PV	1RDR	Ca^{2+}	0.1 mM EDTA	0.2 mM EDTA, 600-1200 mM $CaCl_2$	From map and coordination geometry it is unlikely to be Ca^{2+}
RV	1N35	No ion	1 mM EDTA	-	-
	1MWH	Mn^{2+}		1.5 mM Mn^{2+}	Cap bound structure with Mn^{2+} ion at equivalent site

3.4. Conclusions and a Model for Polymerisation

The non-catalytic ion site is conserved across a large number of viral RNA polymerases (Poranen *et al*., 2008b). We have demonstrated that at high concentration (10 mM) Mg^{2+} can replace Mn^{2+} at this position in the ϕ6 RdRp (**Table 3.1**; **Fig. 3.1**). Since several of the crystallization conditions for other RdRps had high concentrations of Mg^{2+} salts (up to 300 mM; **Table 3.2**) it is plausible that Mn^{2+} could be the natural ion in all these polymerases. In the cases where no divalent ions were seen this might reflect their loss during purification when chelating agents were included. Thus the important role of Mn^{2+} in controlling the switch to elongation (**Fig. 3.22**) may be broadly conserved in viral RdRps. The general dependence of viral RdRps on Mn^{2+} also supports this hypothesis.

In this section a cartoon is presented illustrating the likely sequence of events for the ϕ6 polymerase model system from the early stages through initiation to elongation, using structural and biochemical data described herein and previously published data (**Fig. 3.23**). Each stage is outlined below, focussing on the role of Mn^{2+} at the non-catalytic ion site in *de novo* RNA synthesis.

3.4.1. Apoenzyme and Mn^{2+}-bound state (Step I)

Our data reveal that the non catalytic divalent cation binding site has sufficiently high affinity for Mn^{2+} that it co-purifies with the polymerase (see **Fig. 3.23, step I**). High selectivity for Mn^{2+} is essential given its low physiological concentration (10 μM) compared to Mg^{2+} (Finney and O'Halloran, 2003). The stable binding of Mn^{2+} increases the structural fluidity of the RdRp (**Table 3.1**) which is critical for its activity (see below). However, the replacement of Mn^{2+} by Mg^{2+} (**Table 3.1**) results in stabilization of the enzyme.

3.4.2. Preinitiation (Steps II - III)

NTPs enter the polymerase at the NTP tunnel and can be stabilised at the interrogation site (see **Fig. 3.23, step II**). Our crystallographic data reveal that Mg^{2+} ions are likely to be drawn into the polymerase by these NTPs and can have a number of geometries about the non-catalytic ion site (e.g. **Fig. 3.4.**; see **Fig. 3.23, step II***). Additionally, we demonstrated that the template enters over a positively charged groove on the outside of the polymerase (see **Fig. 3.23, step II***), and is drawn into the tunnel by two positively charged patches on either side of the template tunnel [e.g. **Fig, 3.5**; for genetic evidence see Sarin *et al*., (2009)].

Entry of template into the tunnel with Mn^{2+} at the non-catalytic site has been observed previously [(Butcher *et al*., 2001; Salgado *et al*., 2004); see **Fig. 3.23, step III**] and is essential for initiation complex assembly. Replacement of the non-catalytic Mn^{2+} by Mg^{2+} reduced the ability for template to enter the tunnel (e.g. **Fig, 3.5**; also mutation E491 to Q with a lower affinity for Mn^{2+} reduces template binding, see Poranen *et al*., (2008) and consequently the assembly of the initiation complex. Nevertheless our structural data suggest that Mg^{2+} can be drawn from the canonical non-catalytic Mn^{2+} site to either a central or upwards position by the D1 and D2 NTPs present at the active site in the absence of Watson-Crick base pairing with the template (e.g. **Fig. 3.4**; see **Fig. 3.23, step II***).

3.4.3. Initiation (Step IV)

At the point of initiation the D1 and D2 NTPs (coordinated at the non-catalytic site) shift 6Å from the preinitiation position to base-pair with the ultimate (T1) and penultimate (T2) nucleotides of the template, respectively (see **Fig. 3.23, step IV**). As the NTPs move, the bound cations shift to their catalytic position. The strong affinity of the non-catalytic site for Mn^{2+} is critical for the formation of the functional initiation complex. In line with this, the

E491Q mutant formed an inactive initiation complex when Mn^{2+} was not present at the non-catalytic site (Poranen *et al*., 2008). The positions of divalent cations and NTP in a productive initiation complex is depicted in **Fig. 3.23**, **step IV** (see also Butcher *et al*., 2001).

3.4.4. Catalysis (Step V)

After linkage of the first two NTPs (which is dependent on the non-catalytic Mn^{2+} site being occupied to correctly structure the active site), the released pyrophosphate and one bound Mg^{2+} ion diffuse out of the active site whilst the second catalytic Mg^{2+} ion stays associated with 5'-triphosphate group of the initiated daughter strand (see **Fig. 3.23, step V**) . A dead-end structure showing the products of this reaction has been published previously (PDB accession no.: 1UVK; Salgado *et al*., 2004). Here the oligonucleotide template has been lost from the binding tunnel, demonstrating that at the dinucleotide stage the template is not locked into the polymerase (Salgado *et al*., 2004). In line with this, the heparin trap assay (**Fig. 3.22**, Wright *et al*., 2010; submitted) demonstrates how RNA template and heparin continually compete for entry at the template binding tunnel during initiation, until a 'locked in' heparin resistant elongation state (trinucleotide daughter strand product) is reached (**Fig. 3.22**).

3.4.5. Template movement and displacement of the C-terminal domain (Step VI)

The E634Q mutant (that binds Mn^{2+} strongly, like WT) has significantly less thermal stability than WT with stability being impaired further by Mn^{2+}-binding (**Table 3.1**). Having previously inferred the displacement of the CTD during the transition to elongation to allow the exit of the product (Butcher *et al*., 2001), we now have direct evidence for displacement from the destabilized E634Q mutant where the template can shift past the active site,

displacing part of the CTD (**Fig. 3.13**; **Fig. 3.16**; **Fig. 3.18**; **Fig. 3.19**). This transition, which is facilitated by the bound Mn^{2+}, transiently weakens the high affinity Mn^{2+} binding site so that the ion is released; this is only observed in those molecules where the transition to elongation has occurred (e.g. **Fig. 3.13, C**; see **Fig. 3.23, step VI**). In the absence of external Mn^{2+} the reaction stalls at this point.

3.4.6. Mn^{2+} dependent assembly of the elongation complex (Step VII)

Following displacement of the CTD, the high affinity binding site for Mn^{2+} is reformed, exogenous Mn^{2+} binds and establishes the correct geometry for catalysis of the third nucleotide (D3 NTP base-pairing with T3 nucleotide of template), see **Fig. 3.23, step VII**. After this transition and synthesis of the trinucleotide daughter strand, Mn^{2+} is stably bound and external ions are not required for elongation (**Fig. 3.22**). Thus the transition from initiation to elongation is the only step in the ϕ6 RdRp catalyzed polymerization reaction that is entirely dependent on the presence of *external* Mn^{2+}.

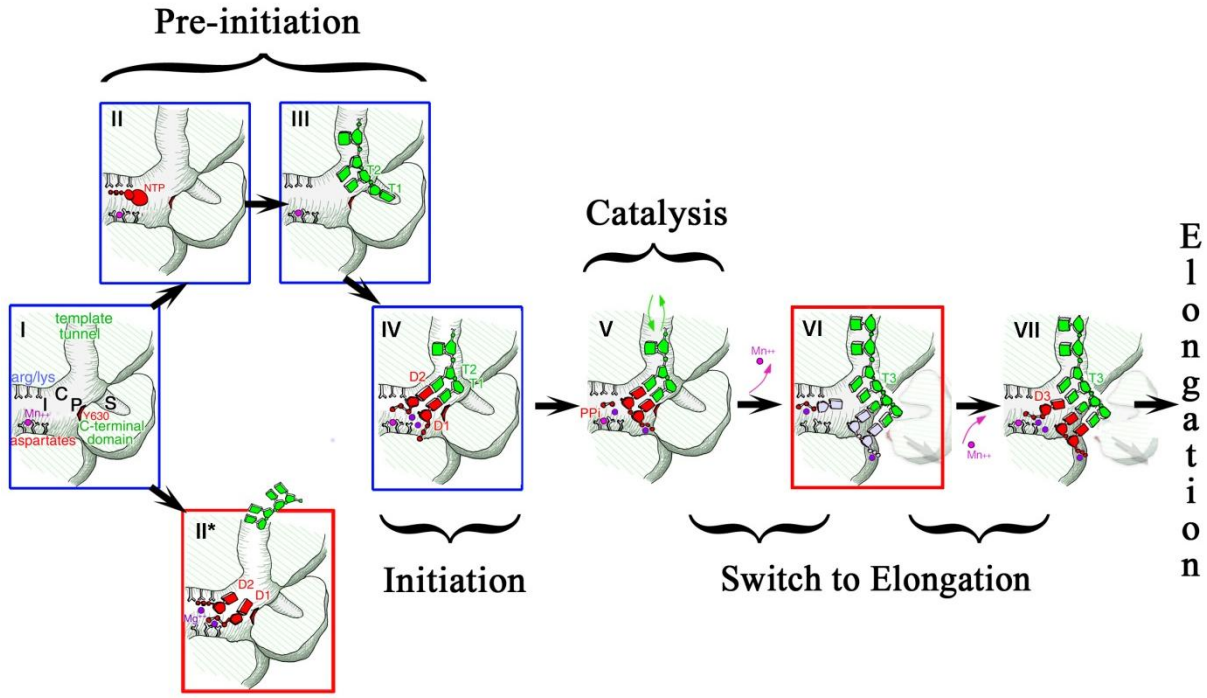


Figure 3.23. Cartoon Model for ϕ6 RdRp catalysed polymerisation.

Step I, polymerase with Mn^{2+} (magenta sphere) bound at non-catalytic ion site. The positions of the interrogation site (arg/lys, site I), catalytically relevant NTP binding site (site C), priming platform (Y630, site P) and specificity pocket (site S) are indicated; nomenclature according to Butcher *et al*., 2001. The template tunnel, C-terminal domain (CTD) and position of catalytic aspartates are also indicated. **Steps II-III** (preinitiation), continuous entry and exit of template (green) and/or D1 and D2 NTPs (red) with Mg^{2+} ions (purple spheres). The template 3'-end nucleotide (T1) enters the tunnel in the presence of Mn^{2+} and binds at site S (**Step III**, Butcher *et al*., 2001) so that the penultimate nucleotide (T2) is held at active site (site P). NTPs in the absence of Watson-Crick base pairing with template can be stabilised about the non-catalytic ion site if occupied by Mg^{2+} (**Step II***); arrows (black) indicate routes from apoenzyme to the different preinitiation states that are sensitive to divalent cation conditions (i.e. high concentrations of Mg^{2+} reduces the ability for template to enter the template tunnel and stabilises NTPs in various positions at the active site). **Step IV** (assembly of initiation complex), after binding of template at site S, the D2 NTP shifts to base pair with the T2 nucleotide at which point the template ratchets back to allow base pairing of the D1 NTP with the T1 nucleotide of the template (see Butcher *et al*., 2001 for more detail); Mg^{2+} ions shift into catalytic position and Mn^{2+} occupies the non-catalytic ion

site. **Step V** (catalysis), the released pyrophosphate (PPi) diffuses out from the polymerase with one bound Mg^{2+}, the other catalytic Mg^{2+} ion stays associated with 5'-triphosphate group of the daughter strand; template can still be lost from the tunnel at this stage (green arrows) allowing competition with heparin for entry (see Salgado, *et al*., 2004 for more detail and PDB accession no. 1UVK). **Step VI-VII**, (switch to elongation), two step mechanism involving CTD displacement (loss of Mn^{2+}) and elongation complex assembly (recapture of Mn^{2+}). **Step VI**, firstly CTD displacement is facilitated by the bound 'destabilising' Mn^{2+}. The di-nucleotide initiation complex shifts down placing the template T3 nucleotide in the active site allowing it to base pair with an incoming D3 NTP. Upon displacement of the CTD, there is transient weakening in the high affinity non-catalytic ion site and concomitant loss of the 'destabilising' Mn^{2+}; the reaction stalls at this point in the absence of *external* Mn^{2+} ions [in the cartoon, catalytic Mg^{2+} ions (purple spheres) and the positions of the daughter strand and incoming D3 NTP (light grey nucleotides) are shown for completeness but were not captured in the E634Q-hairpin-AMPPCP-Mg^{2+} structures]. **Step VII**, the CTD is at its new position and the high affinity Mn^{2+} binding site is reformed. The second step in the two step transition mechanism involves the D3 NTP moving from site I to base pair with T3 nucleotide, now positioned at the active site. *External* Mn^{2+} ions promote efficient rebinding of Mn^{2+} at the non-catalytic ion site and establish the correct geometry for elongation complex assembly and catalysis of the third nucleotide with catalytic Mg^{2+} ions. After synthesis of the trinucleotide daughter strand, Mn^{2+} is stably bound and there is no requirement for *external* Mn^{2+} (however the presence of *external* Mn^{2+} greatly increases the rate of elongation). Red boxes highlight structures described in this thesis and blue boxes indicate previously published structures (Butcher *et al*., 2001; Salgado *et al*., 2004).