Structural Studies of RNA-dependent RNA polymerases

Thesis

Structural Studies of RNA-dependent RNA polymerases

Maria Paula Santos Cordeiro Salgado

Open University

The Weatherall Institute for Molecular Medicine

Division of Structural Biology

The Henry Wellcome Building of Genomic Medicine

Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

October 2005

PAGES 234-276 NOT SCANNED AT REQUEST OF UNIVERSITY

Abstract

Most RNA viruses possess an RNA-dependent RNA polymerase (vRdRP), responsible for viral genome replication and transcription. Furthermore, a number of eukaryotic organisms, including plants, fungi, protozoa and some metazoans, produce cellular RdRPs (cRdRPs) involved in RNA silencing mechanisms.

One of the best studied vRdRPs is that of the dsRNA bacteriophage Φ6. Structures of Φ6 RdRP (Φ6pol) in complex with RNA oligonucleotides revealed the basis for template specificity: the extra hydroxyl group leads to additional RNA-protein interactions, further stabilizing the template. Structures of "manganese-free" Φ6pol and a mutated form of the protein with lower affinity for the ion (E491Q mutant) provided some hints to the role of manganese. The structure of a complex of Φ6pol with RNA oligonucleotides and GTP with the catalytic magnesium ions substituted by calcium ions shows a distorted geometry of the initiation competent state, providing a molecular explanation of the calcium inhibitory effect. Finally, the structure of a mutated form of the protein (628QYKW632-SG mutant) prone to back-priming initiation revealed a set of contacts important for *de novo* initiation. Considering the high structural homology of Φ6pol with other vRdRPs, particularly from (+)ssRNA Hepatitis C Virus (HCV), insights into the mechanistic and structural details of Φ6pol are thought to be relevant to the general understanding of vRdRPs.

The dimeric structure of QDE-1, an RdRP from the fungus *Neurospora crassa* involved in RNA silencing, revealed a surprising similarity at the active site level to multisubunit DNA-dependent RNA polymerases (DdRPs). This implies a close evolutionary relationship between these enzymes and a possible connection

between RNA silencing pathways and primordial RNA polymerisation mechanisms. Furthermore, an analysis based on the structures of several template dependent polymerases suggests that they have emerged more than once over the course of evolution.

Acknowledgments

First of all, I would like to thank Jon Grimes for his wonderful and friendly supervision, for always being willing to answer my questions and guiding me through the good and bad times. A special thanks to Dave Stuart, for his wisdom and always helpful and mind-opening insight. To both, a huge thanks for introducing me to the world of virus and polymerases and allowing me to learn so much in the last four years.

I am grateful for the funding of this work provided by the Human Frontier Research Program.

I would like to thank Melanie Globe and Dr. Kathryn Robson at the Weatherall Institute of Molecular Medicine for welcoming my registration with the Open University and their help.

I would also like to thank Dennis Bamford for a great collaboration, for all the useful discussions and his support. A special thanks to Minni L. Koivunen, my "protein production facility", for a great collaboration relationship, for always promptly answering my requests and also for her friendship. Also thanks to Sarah Butcher for her insightful comments and help. To Riita, Ninnu, Nelli and everybody in the Bamford Laboratory at the University of Helsinki, for all their help and making me feel welcome during my visit. I would also like to thank Eugene Makeyev for all his support, brilliant suggestions and useful discussions, particularly helpful during the writing of this thesis.

A special thanks goes to the staff at BM14, ESRF, particularly Martin Walsh for his help with data collection.

A lot of people at the Division of Structural Biology and the Oxford Protein Production Facility have contributed to this work and the pleasant time I spent at the Henry Wellcome Building for Genomic Medicine (HWBGM). Special thanks go to Erika Mancini, Geoff Sutton, Nicola Abrescia and Christian Siebold for all their help, support and very useful discussions. This work would have taken much longer and been much harder without your friendship and help around the lab and particularly during the synchrotron data collection trips. Thanks to Robert Esnouf and Jon Diprose for all their patient help with computing. I am grateful to Karl Harlos and Jun Dong for maintaining the in-house X-ray facility, to Margaret Jones for maintaining the wet lab running smoothly and to Linda Vincent for her help in all administrative issues. I would also like to thank Tom Walter for his help with the Cartesian Robot and his friendship. To Bal, Chris, Gijs, Kamel, James, Jo, Joe,

Maria, Mike, Robert, Ross and all other members of Strubi and the OPPF for helping to make the last four years so much fun. A special thanks to my friends at the HWBGM, past and present, for your friendship and wonderful moments together – Boni, Branwen, Phil, Silvia, Thil, Valentina.

I would like to thank everyone with whom I shared wonderful moments at Portholes and also for your support through the less happy ones – Elena, Emmanuelle, Laurence. To Gaia, my dearest friend, for reminding me to have the courage to pursue my dreams.

Thanks to everyone at the Pegasus Theatre and all the cast of "Volpone", for such a good time. A very special thanks to Alice, Ana, Jessica, Maria, Masayo and Steve for creating GAIA and keeping it alive. For your support, friendship and a breath of fresh air every week throughout the writing of this thesis.

Um agradecimento especial à Prof. Ana Margarida Damas e Dr. Paula Sebastião por me revelarem o mundo da cristalografia. À Prof. Maria João Saraiva pelo seu apoio.

Aos meus amigos em Oxford e outras paragens do Reino Unido –Rita ("Obrigada por teres ido comigo à Grécia!"), Paulo, Sandra, Marisa, Duarte, Borges, Joana, Luís, Artur, André, Susana – pelo convívio, pela amizade, pelo apoio. A vida em Inglaterra seria muito mais cinzenta e tristonha sem vocês.

A todos os amigos que me acompanharam no meu percurso até aqui, nomeadamente à Isabel, Filipa, Raquel, Óscar, Xana, Sandra, Nuno, Kalina, Filipe, Carla, Paula, Toni, Pedro, Tita, André, Jorge, Carlos, obrigada pela vossa amizade e apoio.

À Berta, Patrícia e Alexandre pela vossa amizade e carinho. À Máxima, Rui e D. Isaura pelo carinho com que sempre me receberam. Aos meus primos Telmo, Ricardo, Nuninho e Sara e a toda a minha família, pelo apoio e amizade.

Aos meus pais, um abraço enorme pelo vosso amor e carinho, pelo apoio, pelo entusiasmo e encorajamento constantes. Obrigada por me terem ensinado a sonhar, a confiar em mim e acima de tudo a ser verdadeira comigo própria em cada momento. Ao meu irmão, por ser o meu melhor amigo, sempre. Pelas risadas, pelas brincadeiras, pelos disparates, pelo apoio e carinho. Aos meus avós, um abraço grande e obrigada pelo vosso amor e carinho.

Ao Rui, por tudo. Por me aturar, acompanhar, ajudar, incentivar todos os dias durante a escrita desta tese. Por estar presente, sempre, nos bons e maus momentos. Pelo amor e carinho, sempre.

The work presented in this thesis is part of a collaborative project with Prof. Dennis Bamford group at the University of Helsinki, within the framework of the Human Frontier Research Program. Part of this work has been published or is being prepared for publication:

- **Salgado, P.S.**, Makeyev, E.V., Butcher, S.J., Bamford, D.H., Stuart, D.I. and Grimes, J.M. (2004) The Structural Basis for Rna Specificity and Ca2+ Inhibition of an Rna-Dependent Rna Polymerase. *Structure (Camb)*, **12**, 307-316.

- **Salgado, P.S.**, Walsh, M.A., Laurila, M.R., Stuart, D.I. and Grimes, J.M. (2005) Going Soft and Sad with Manganese. *Acta Crystallogr D Biol Crystallogr*, **61**, 108-111.

- Laurila, M.R., **Salgado, P.S.**, Stuart, D.I., Grimes, J.M. and Bamford, D.H. (2005b) Back-Priming Mode of Phi6 RNA-Dependent RNA Polymerase. *J Gen Virol*, **86**, 521-526.

- Laurila, M.R., **Salgado, P.S.**, Makeyev, E.V., Nettelship, J., Stuart, D.I., Grimes, J.M. and Bamford, D.H. (2005a) Gene Silencing Pathway RNA-Dependent RNA Polymerase of Neurospora Crassa: Yeast Expression and Crystallization of Selenomethionated Qde-1 Protein. *J Struct Biol*, **149**, 111-115.

- **Salgado, P.S.**, Koivunen, M.R.L., Makeyev, E.V., Bamford, D.H., Stuart, D.I. and Grimes, J.M. (2005) An evolutionary link between RNA silencing and transcription revealed by the structure of an RNAi polymerase (*submitted*)

- Koivunen, M.R.L., Poranen, M., **Salgado, P.S.**, Stuart, D.I., Grimes, J.M. and Bamford, D.H. Φ6 RNA-dependent RNA polymerase: The role of manganese (*in preparation*)

- Makeyev, E.V., **Salgado, P.S.** Structure, function and evolution of virus and cell-encoded RNA-dependent RNA polymerases. In *Recent Advances in RNA virus replication* (K. Hefferon, Ed.), Research Signpost/Transworld Research Network (*in preparation*)

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Polymerases

Polymerisation of nucleoside triphosphates (NTPs) is a central mechanism of several important and vital biological processes that include transcription, primer synthesis during deoxyribonucleic acid (DNA) replication, addition of polyadenylate tails to messenger ribonucleic acid (RNA), uridylation in RNA editing, viral RNA replication, amongst many others. The polymerisation reaction is catalysed by a wide variety of enzymes that can be classified into two major mechanistic categories: template dependent and template independent nucleotidyl transferases. The second group includes CCA-adding enzymes, poly-alanine polymerases, uridyl transferases and oligoA-synthetases. All template independent RNA transferases share the same fold of the catalytical domain. A detailed description of this group of nucleotidyl transferases is beyond the scope if this thesis, since all studied polymerases function in a template dependent manner.

The group of template dependent polymerases refers to enzymes that use a 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 (RT)

(iv) RNA-dependent RNA polymerases (RdRPs)

Despite the differences in template and product preferences between classes, it was logical to consider the possibility that they might share common mechanisms of catalysis and, consequently, sequence and structural similarities. The task of identifying those common features is difficult due to the diversity of origins of the enzymes, spanning from virus to humans. Early efforts to establish evolutionary relationships between viral RdRPs and RTs by Poch and co-workers allowed identification of at least four conserved sequence motifs A, B, C and D (Poch *et al.*, 1989). Parallel efforts to unify all classes of single subunit polymerases based on sequence comparisons were done by Delarue *et al.* (1990). Their studies revealed that two of the motifs identified by Poch *et al.* are present in all polymerase classes: motifs A and C, containing strictly conserved aspartate residues (Fig. 1.1) which were therefore postulated to be critical for polymerisation. Site-directed mutagenesis studies (reviewed by O'Reilly and Kao, 1998) subsequently lent support to this hypothesis. Moreover, structural information from a number of different nucleic acid polymerases has provided further insights to the mechanism of polymerisation and the evolutionary links between different polymerase classes.

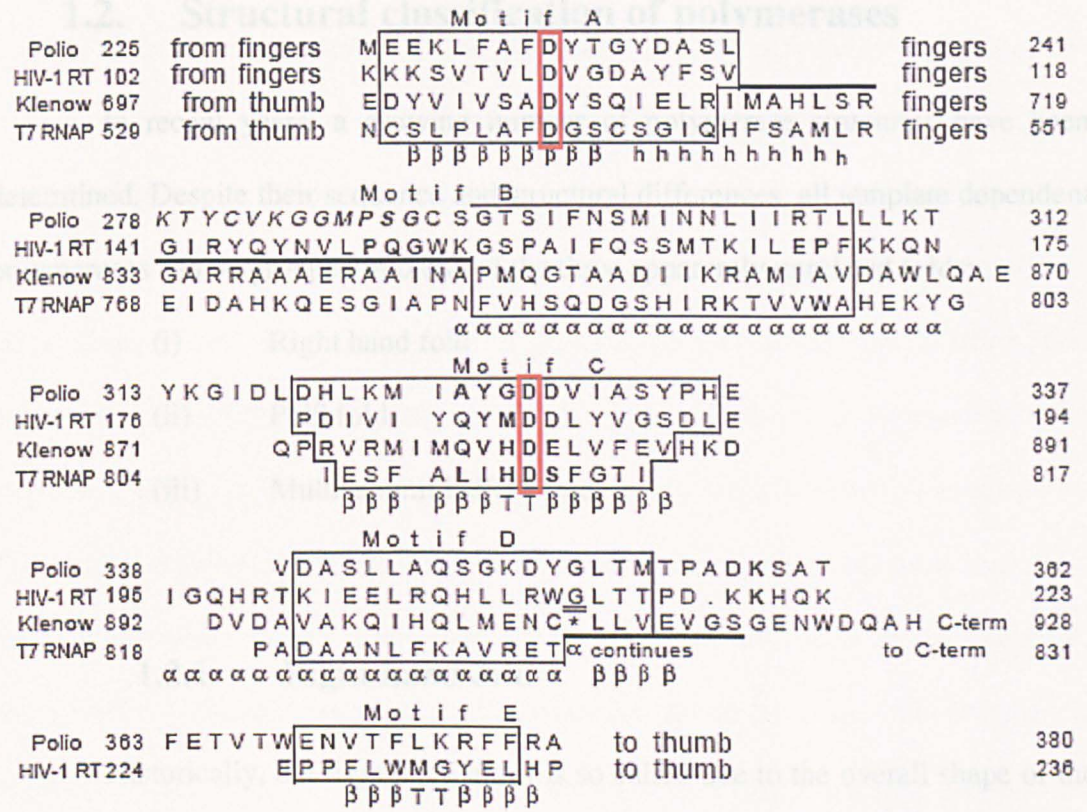


Figure 1.1. Sequence alignment of representative members of the four classes of polymerases

Structure based sequence alignments for poliovirus polymerase (RdRP), HIV-1 reverse transcriptase (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 the RNA-dependent polymerases is also shown. The strictly conserved aspartates in motif A and C are outline (red boxes).

[Modified from (Hansen *et al.*, 1997)]

1.2. Structural classification of polymerases

In recent years, a growing number of polymerase structures have been determined. Despite their sequence and structural differences, all template dependent polymerases can be grouped into one of the three apparently unrelated folds:

(i) Right hand fold

(ii) Polβ fold

(iii) Multisubunit DdRPs fold

1.2.1. Right hand fold

Historically, the right hand motif is so called due to the overall shape of the *Klenow* fragment (KF) of DNA-dependent DNA polymerase I from *Escherichia coli* (Ollis *et al.*, 1985), the first polymerase structure to be determined. It has three different sub-domains, named "palm", "fingers" and "thumb" (Fig. 1.2.A). As structures of other polymerases emerged, clear structural similarities became evident. The first reverse transcriptase to be structurally determined was that of the human immunodeficiency virus type 1 (HIV-1) by Kohlstaedt *et al.* (1992) (Fig. 1.2.C). The structure of the RNA polymerase from bacteriophage T7 (T7pol) was the first reported DNA-dependent RNA polymerase (Sousa *et al.*, 1993) (Fig. 1.2.B). RNA-dependent RNA polymerase structures were more elusive, with poliovirus RNA polymerase (PV 3Dpol) not determined until 1997 (Hansen *et al.*, 1997) (Fig. 1.2.D). All four polymerases exhibited the right hand motif. The palm sub-domain is the most well conserved structural feature between all classes (Fig. 1.2). It consists largely of a β-sheet that forms the base of the polymerase cleft and two long α-

helices packed against it, which are mostly superimposable in all four polymerase classes. This conserved structural arrangement of the palm sub-domain positions the strictly conserved aspartates (one from sequence motif A and one from sequence motif C; Fig. 1.1 & 1.2) in an equivalent three-dimensional organisation, postulated to be essential for catalysis. Based on these structural observations, a "two metal ion" mechanism, where the aspartates serve the key role of anchoring the ions, was proposed (Steitz, 1993; Joyce and Steitz, 1995; Steitz, 1998). In this mechanism, the incoming NTP is accompanied by two divalent metal ions (normally magnesium). The ions bind to the phosphates of the nucleotide and the carboxylate moieties of two of the conserved aspartates, one from motif A (Asp_A) and the other from motif C (Asp_C) (Fig. 1.3). In the reaction catalysed by these enzymes, the polymer is produced by sequentially adding a (d)NTP monomer via its α-phosphate group to the 3' hydroxyl group of the "daughter" chain, establishing a new bond and releasing the β and γ-phosphate groups as a molecule of inorganic pyrophosphate (PPi). In this reaction, the metal ions act both as facilitators and stabilizing elements. One of the metal ions (A in Fig. 1.3) lowers the affinity of the 3' hydroxyl for the hydrogen atom, facilitating the nucleophilic attack on the α-phosphate group. The second metal ion (B in Fig. 1.3) coordinates the β and γ-phosphate groups, hence assisting their release as PPi. Furthermore, both ions contribute to the stabilisation of the transient pentacovalent state of the α-phosphate group. Structural determination of the first complexes of polymerases, namely T7pol (Doublie *et al.*, 1998) and *Klenow* fragment from *T. aquaticus* (Li *et al.*, 1998), and NTPs (or their analogues) confirmed that this was the mechanism utilized by DdRPs and DdDPs. It also provided insight to the role of the conserved aspartates: Asp_A and Asp_C are involved

in the coordination of the metal ions, ensuring their correct positioning for catalysis to occur. The third conserved aspartate belonging to motif C is thought to be important to stabilize the overall geometry of the active site. More recent structural studies with viral RdRPs have further extended the understanding of catalysis at the molecular level, as discussed below. These observations highlight the significance of the palm sub-domain for catalysis, which is reflected in the high structural similarity between palm sub-domains across polymerase classes.

Conversely, the fingers and thumb sub-domains show very little structural similarity across classes of polymerases. DNA-dependent polymerases (represented by T7pol and KF) and RNA-dependent polymerases (represented by RT and PV 3Dpol) have largely α-helical thumb domains, although with only superficial similarities in structure and little or no equivalence in sequence and connectivity (Fig. 1.1 & 1.2). The fingers sub-domains are strikingly different, with DNA-dependent enzymes having predominantly α-helical "fingers" and RNA-dependent polymerases containing both α-helices and β-sheets (Fig. 1.2). Considering that both sub-domains are important in template and substrate binding (Joyce and Steitz, 1995; O'Reilly and Kao, 1998), differences in structure are likely to reflect variation in template preference.

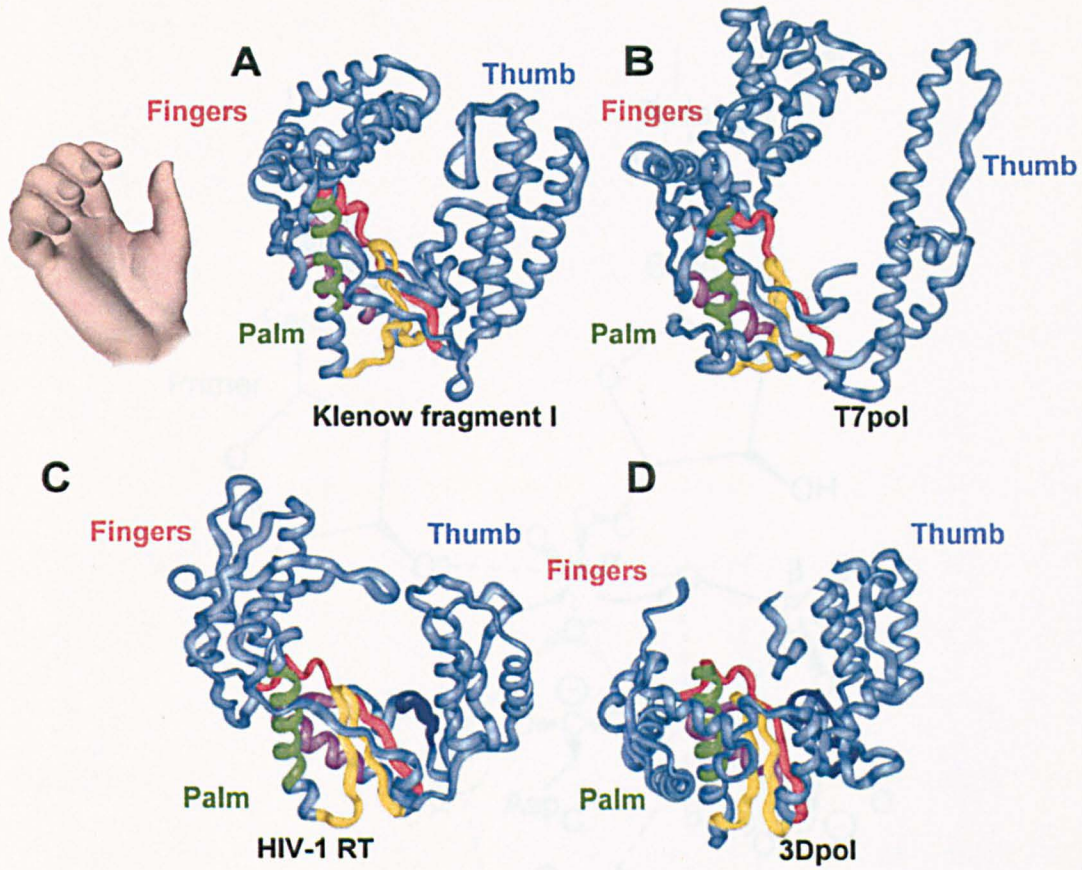


Figure 1.2. Structures of right hand fold polymerases from each class of template dependent polymerases

A. Structure of the polymerase domains of Klenow fragment of *E. coli* DNA polymerase I – DdDP (Ollis *et al.*, 1985)

B. T7 RNA polymerase structural model – DdRP (Sousa *et al.*, 1993)

C. Structure of the polymerase domain HIV-1 reverse transcriptase (HIV-1 RT) – RdDP (Kohlstaedt *et al.*, 1992)

D. Structural model of poliovirus 3Dpol - RdRP (Hansen *et al.*, 1997)

All polymerases are positioned with the thumb sub-domain to the right and the fingers sub-domain to the left. The conserved sequence motifs are colour coded: A in red, B in green, C in yellow, D in purple, and E in dark purple.

[Modified from (Hansen *et al.*, 1997)]

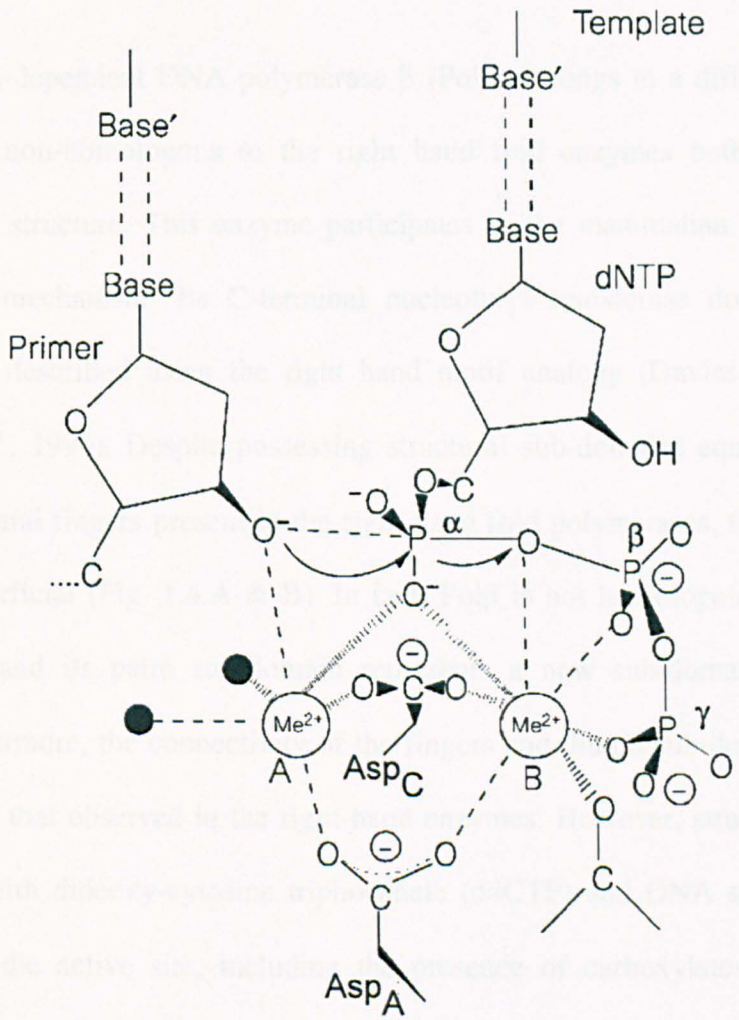


Figure 1.3. **Schematic representation of "two-metal ion" polymerisation mechanism**

Two divalent metal ions, A and B, are ligated to enzymes of the polymerase by strictly conserved aspartates residues Asp_A and Asp_C. The black circles are water molecules bound to metal ion A. [Modified from (Steitz, 1998)]

1.2.2. Polβ fold

DNA-dependent DNA polymerase β (Polβ) belongs to a different class of polymerases, non-homologous to the right hand fold enzymes both in terms of sequence and structure. This enzyme participates in the mammalian base excision DNA repair mechanism. Its C-terminal nucleotidyl transferase domain (31kDa domain) was described using the right hand motif analogy (Davies *et al.*, 1994; Pelletier *et al.*, 1994). Despite possessing structural sub-domains equivalent to the palm, thumb and fingers present in the right hand fold polymerases, the similarities are only superficial (Fig. 1.4.A & B). In fact, Polβ is not homologous to the other polymerases and its palm sub-domain represents a new sub-domain class (Fig. 1.4.B). Furthermore, the connectivity of the fingers and thumb sub-domains in Polβ is opposite to that observed in the right-hand enzymes. However, structures of Polβ in complex with dideoxy-cytosine triphosphate (ddCTP) and DNA show a similar geometry of the active site, including the presence of carboxylates coordinating metal ions to activate the 3' hydroxyl of the primer strand (Sawaya *et al.*, 1994) (Fig. 1.4.B,C) These observations indicate that a similar two metal ion mechanism is used for catalysis.

The palm sub-domain needs to generate an appropriate surface to accommodate the substrate molecules and to present the catalytic ions in the suitable geometrical arrangement relative to the NTP α-phosphate and primer 3' OH. However, it appears clear that quite distinct structures can meet these requirements. The fact that Polβ and right hand polymerases are non-homologous but share a catalysis mechanism and active site geometry (Fig. 1.4.C), indicate that they must have achieved it by convergent evolution from unrelated ancestors.

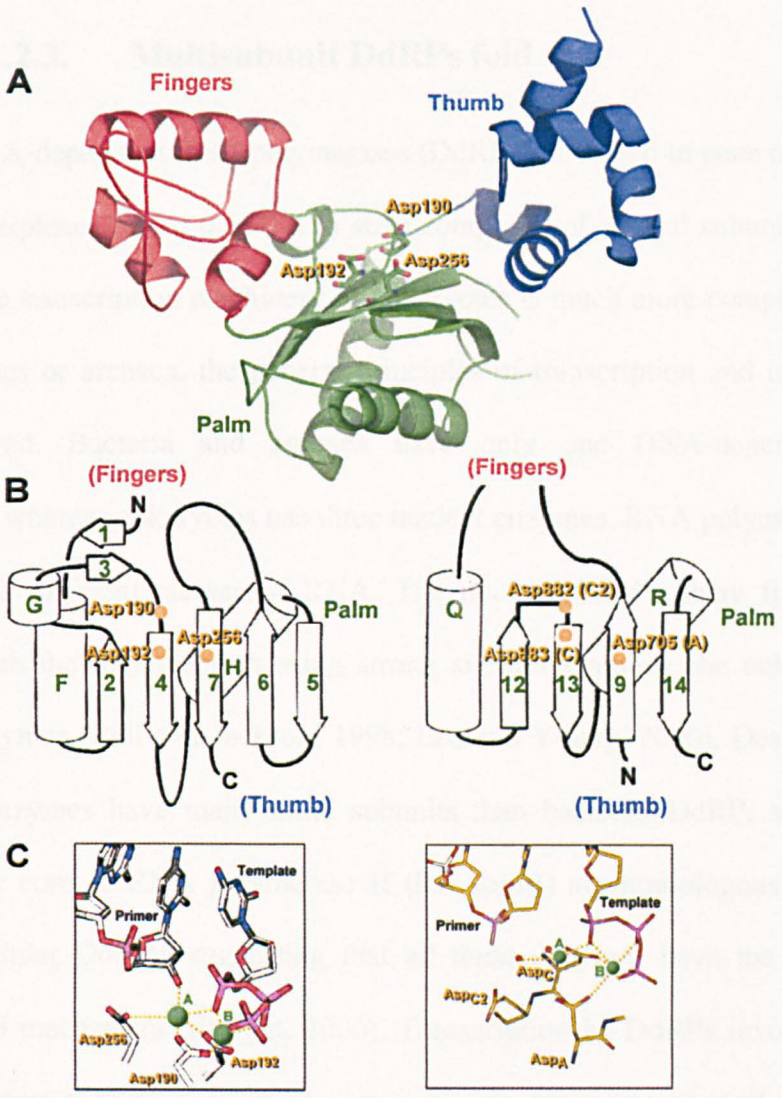


Figure 1.4. Polβ fold

A. Structure of rat DNA polymerase β (Davies *et al.*, 1994). The divergent equivalents of right hand sub-domains are coloured green for the palm, blue for the thumb and red for the fingers. Only the catalytical 31kDa domain of Polβ is shown. Catalytic residues D190, D192 and D256 are shown as ball-and-stick representation (coloured by atom type: C – green, O – red, N – blue).

B. Schematic representation of the "palm" subunit of Polβ (left) and Klenow fragment (right) (as example of right hand fold). β-strands (numbered arrows) in the central β-sheet and helices (lettered cylinders) are shown, with the aspartates at the active site highlighted (orange dots). [Modified from (Joyce and Steitz, 1995)].

C. Ball-and-stick representation of the active site of Polβ (left) and right hand (right) fold polymerases in equivalent orientation. Metal ions A and B (as numbered in Fig. 1.3) are represented as green spheres. The catalytic aspartates are highlighted. Interactions between the ions and the protein, template and primer nucleotides are shown as dotted yellow lines.

1.2.3. Multisubunit DdRPs fold

DNA-dependent RNA polymerases (DdRPs), involved in gene transcription, are large complexes (up to 0.6MDa in size) composed of several subunits (5 to 15). Although the transcription machinery of eukaryotes is much more complex than that of prokaryotes or archaea, the general principles of transcription and its regulation are conserved. Bacteria and archaea have only one DNA-dependent RNA polymerase, whereas eukaryotes use three nuclear enzymes, RNA polymerases I–III, to synthesize different classes of RNA. The nuclear DdRPs share five common subunits, with the remainder showing strong similarity among the eukaryotic and archaeal enzymes (Bell and Jackson, 1998; Lee and Young, 2000). Despite the fact that these enzymes have many more subunits than bacterial DdRP, subunits that make up the core of RNA polymerase II (RNApolII) are homologous to subunits from all cellular DdRPs, suggesting that all these enzymes have the same basic structure and mechanism (Ebright, 2000). Transcription by DdRPs involves a large number of subunits and is regulated by interactions with a wide range of transcription factors and other components. Therefore, the mechanism is highly complex and a detailed description is beyond the scope of this thesis. Instead, focus is given to the organization of the active site cleft and interactions of elements in the core formed by the two large subunits with template and substrate molecules.

The catalytically active core of DdRPs is composed of the two largest subunits of the multimeric complex: β' and β in bacteria, RNA binding protein (Rbp) 1 and Rbp2 in yeast, respectively the largest and second largest subunits. Biochemical and sequence studies identified three invariant aspartates - DxDxD motif – in the largest subunit, crucial for catalysis. However, unlike the right-hand

and Polβ folds, that have a catalytic domain (the palm sub-domain) which shows some similarities, no corresponding features are present in multisubunit DdRPs. Structural studies of RNA polymerase II from *Thermus aquaticus* (Zhang *et al.*, 1999), yeast (Cramer *et al.*, 2001) and thermophilic bacteria *Thermus thermophilus* (Vassylyev *et al.*, 2002) revealed a completely different structural arrangement (Fig. 1.5.A). Interactions of the two largest subunits create a positively charged nucleic acid binding cleft. One side of the cleft is defined by the largest subunit (β'/Rbp1) and is formed by a mobile "clamp" involved in the stabilisation of the DNA-RNA hybrid (reviewed by Cramer, 2002). The other side is defined by two sub-domains of the β/Rbp2 subunit: the "lobe" and "protrusion", also involved in stabilisation of the active complex (Fig. 1.5.A). The active site is defined by two β-barrels with the typical double-psi topology (Castillo *et al.*, 1999), one from subunit β/Rbp2, the other from β'/Rbp1. The double-psi β-barrel from subunit β'/Rbp1 contributes to the active site with the three strictly conserved aspartates in the DxDxD motif (D481, D483, D485 in yeast; D739, D741, D743 in *T. aquaticus* and D728, D730, D732 *T. thermophilus*) that coordinate a Mg^{2+} ion. A range of positively charged residues from the β-barrel in subunit β/Rbp2 completes the cleft (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Vassylyev *et al.*, 2002; Artsimovitch *et al.*, 2004; Westover *et al.*, 2004). Beyond the active site, a sub-domain from the β/Rbp2 subunit - "wall" - blocks the cleft.

In the apo-enzyme structures solved to date (Zhang *et al.*, 1999; Cramer *et al.*, 2001; Vassylyev *et al.*, 2002), one Mg^{2+} ion (metal A) was found to be persistently coordinated by the aspartate residues. However, as for single subunit RdRPs, a second Mg^{2+} - metal B - is found to accompany incoming NTPs

(Kettenberger *et al.*, 2004; Westover *et al.*, 2004). The mechanism of catalysis has been proposed to follow the "two metal ion" mechanism found in single subunit polymerases. One key difference is that one of the metal ions is intrinsically bound to the multisubunit polymerases, with only metal B accompanying the NTPs, whereas both metal ions coordinate the incoming NTPs in the single subunit polymerases. Furthermore, two possible binding modes have been described for the incoming NTP, one when the correct nucleoside is present and can form Watson-Crick base-pairing interactions, and another when a mismatched NTP is present (Westover *et al.*, 2004). Westover and co-workers (2004) proposed a functional significance of this NTP binding position: incoming NTPs initially occupy the mismatch "E" site and then rotate around the second metal ion to sample the base pairing in the matched "A" site. When the correct NTP is present, base-pairing locks the NTP in position and polymerization occurs, with subsequent translocation of the RNA-DNA duplex (Fig. 1.5.C). Contacts with a long helix – "bridge helix" (Rbp2 residues 810-846) – that spans the cleft just before the active site (Fig. 1.5) have been proposed to be relevant for nucleic acid-protein interactions during translocation of the RNA-DNA duplex in yeast (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Westover *et al.*, 2004). In the bacterial DdRP, this helix is broken down into two shorter equivalent α-helices (residues 1067-1081 and 1083-1093).

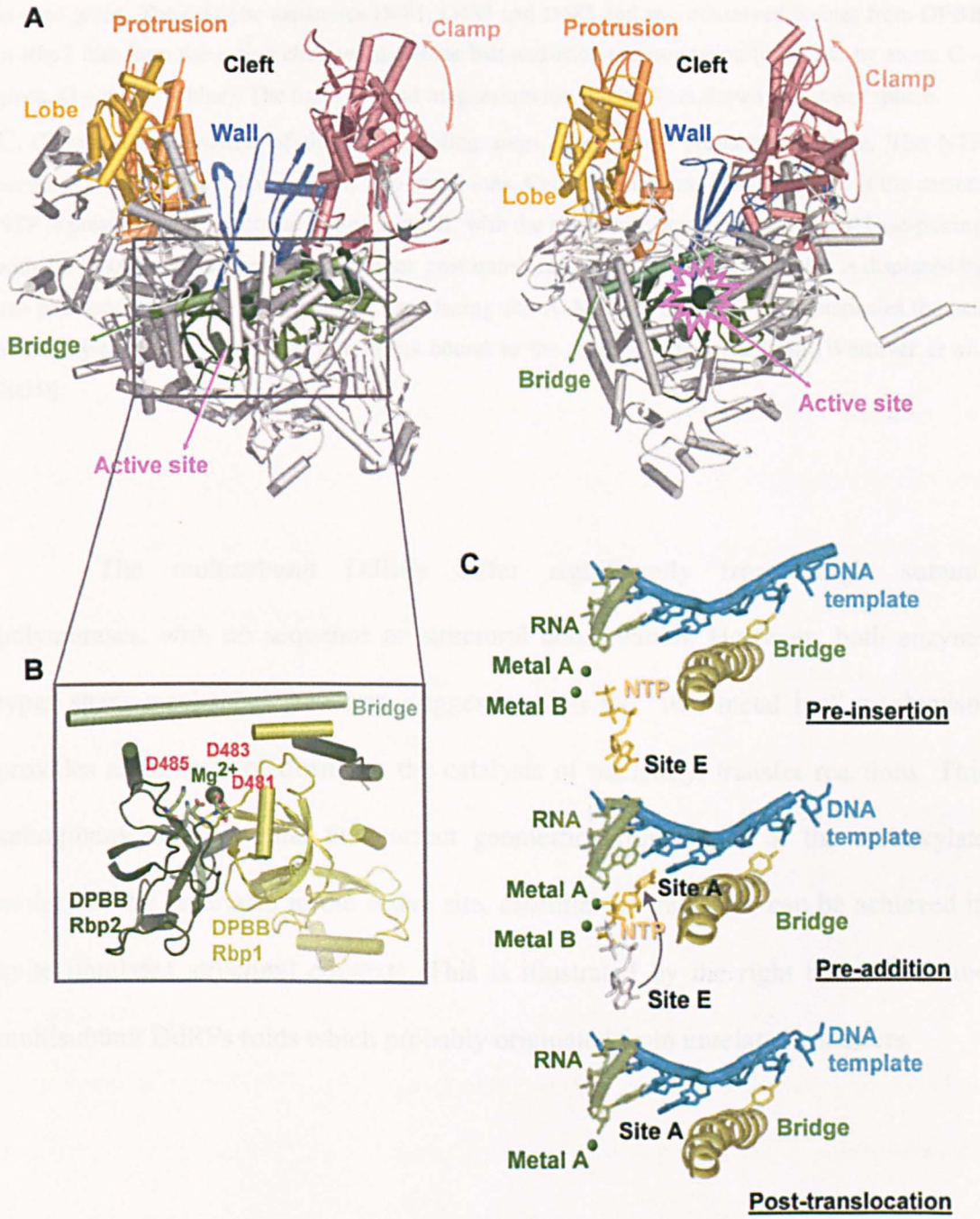


Figure 1.5. Structure of multisubunit DdRP PolII from yeast and bacteria

A. Structure of β and β' and Rbp1 and Rbp2 subunits subunits from *E. coli* (left) and *T. thermophilus* (right) RNA polymerase II. In both panels, important functional/structural sub-domains and the active site (magenta) location are highlighted. β/Rbp2 subunit sub-domains: protrusion (orange), lobe (yellow), wall (blue). B'/Rbp1 sub-domains: DPBBs (green), clamp (red), bridge (dark green), funnel (light green).

B. Zoom view of the active cleft in yeast DdRP, formed by two double-psi β-barrels (DPBB). DPBB from the Rbp1 is coloured light green, DPBB from Rbp2 is coloured dark green and the bridge helix

as lime green. The catalytic aspartates D481, D483 and D485 and two conserved lysines from DPBB in Rbp2 that form the active cleft are shown as ball-and-stick representation (coloured by atom: C – green, O – red; N – blue). The tightly bound magnesium ion (metal A) is shown as a green sphere. **C.** Cartoon representation of the NTP binding steps. Top panel: pre-insertion stage. The NTP occupies site E and is coordinated by two metal ions. Central panel: pre-addition stage. If the correct NTP is present, the base rotates around metal B, with the nucleotide occupying site A and base-pairing with the DNA template base. Bottom panel: post translocation. The RNA-DNA duplex is displaced by one position, leaving the next template base facing site A. Metal B presumably accompanies the exit of PPi by-product whilst metal A remains bound to the protein. [Modified from (Westover *et al.*, 2004)].

The multisubunit DdRPs differ significantly from single subunit polymerases, with no sequence or structural conservation. However, both enzyme types share mechanistic features, suggesting that the "two metal ion" mechanism provides an optimal platform for the catalysis of nucletidyl transfer reactions. This strengthens the view that the correct geometric arrangement of the carboxylate moiety of the aspartates in the active site, essential for catalysis, can be achieved in quite unrelated structural contexts. This is illustrated by the right hand, Polβ and multisubunit DdRPs folds which probably originated from unrelated ancestors.

1.3. Viral RNA-dependent RNA polymerases

Virus encoded RNA-dependent RNA polymerases (vRdRPs) are essential components in the life cycle of RNA viruses, being responsible for RNA synthesis. However, they are only effective *in vivo* when associated with other proteins of both viral and possibly cellular origin in a complex generally termed the "polymerase complex" (PC) (Lai, 1998), despite having the catalytic residues essential for polymerisation. Those interactions vary across virus classes and detailed descriptions are beyond the scope of this thesis. Here, I will focus on the structure and mechanism of the catalytically active polymerase.

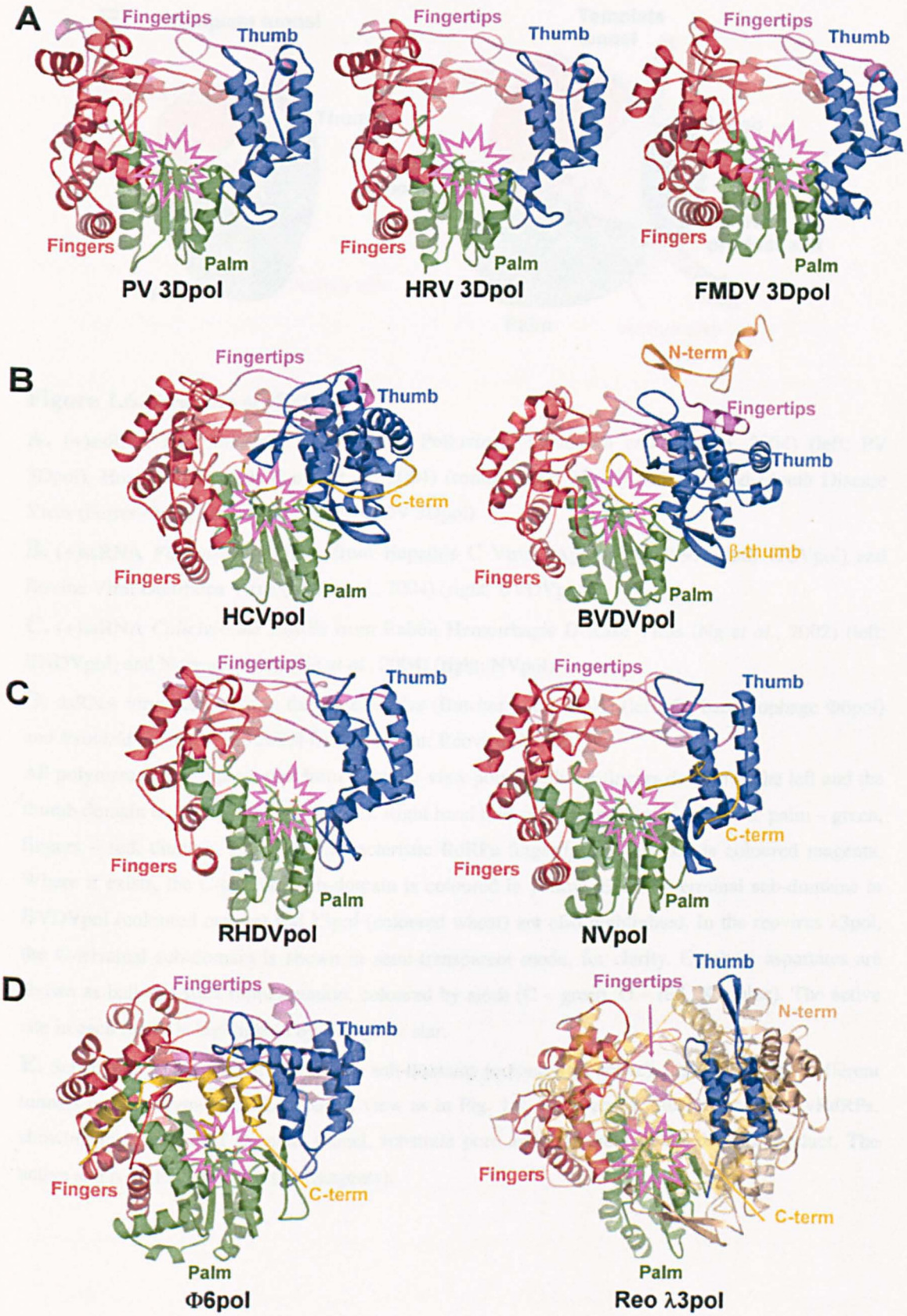
In recent years, a number of vRdRPs have been structurally determined, both in their apo form and, in some cases, in complex with template, NTP and/or metal ions (Table 1.1). Interestingly, so far no X-ray models of RdRPs from negative single-stranded RNA [(-)ssRNA] have been reported. However, several positive single-stranded [(+)ss] and double-stranded (ds) RNA virus polymerases have been determined (Table 1.1.). As mentioned above for the 3Dpol from poliovirus, they exhibit the canonical right hand motif with palm, fingers and thumb sub-domains. Nevertheless, vRdRPs assume a more spherical structure due to an extra structural feature that links the fingers and thumb sub-domains, usually termed the "fingertips". This feature gives RdRPs a "closed hand" appearance, as opposed to the "open hand" conformation of DNA-dependent and RT polymerases (Fig. 1.6). This closed conformation creates two positively charged tunnels in vRdRPs: the template and NTP binding channels [see for example, Butcher *et al.* (2001), Bressanelli *et al.* (2002), Ferrer-Orta *et al.* (2004)].

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

Polymerase	Virus type	Virus family	PDB code	Resl. (Å)	Structural information	Reference
PV 3Dpol	(+)ssRNA	*Picornaviridae*	1RDR	2.4	Disordered residues 1-12;38-66;98-181;270-290	(Hansen *et al.*, 1997)
PV 3Dpol (mutant)	(+)ssRNA	*Picornaviridae*	1RA6	2.0	Mutant L446D/R455D	(Thompson and Peersen, 2004)
PV 3Dpol (mutant) GTP complex	(+)ssRNA	*Picornaviridae*	1RA7	2.0	Mutant L446D/R455D + GTP	(Thompson and Peersen, 2004)
HRV16 pol (serotype 16)	(+)ssRNA	*Picornaviridae*	1XR7	2.3	Serotype 16	(Love *et al.*, 2004)
HRV1B pol (serotype 1B)	(+)ssRNA	*Picornaviridae*	1XR6	2.3	Serotype 1B + K^+	(Love *et al.*, 2004)
HRV14 pol (serotype 14)	(+)ssRNA	*Picornaviridae*	1XR5	2.3	Serotype 14 + Sm^+	(Love *et al.*, 2004)
FMDV 3Dpol	(+)ssRNA	*Picornaviridae*	1U09	1.9	Full length	(Ferrer-Orta *et al.*, 2004)
FMDV 3Dpol RNA complex	(+)ssRNA	*Picornaviridae*	1WNE	3.0	Full length Template:primer RNA	(Ferrer-Orta *et al.*, 2004)
HCVpol 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 (1); Mn^{2+} (1)	(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 1: 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 1: 79-678 79-91; 675-678 disordered	(Choi *et al.*, 2004)
RHDVpol	(+)ssRNA	*Caliciviridae*	1KHV	2.5	1-4; 181-184; 502-516 disor. + Lu^{3+}	(Ng *et al.*, 2002)
RHDVpol	(+)ssRNA	*Caliciviridae*	1KHW	2.7	1-4; 181-184; 502-516 disor. + Mn^{2+}	(Ng *et al.*, 2002)
NVpol (cryst. form I)	(+)ssRNA	*Caliciviridae*	1SH0	2.2	Crystal form I (*P1*) 1-5; 508-510 disordered	(Ng *et al.*, 2004)

Polymerase	Virus type	Virus family	PDB code	Resl. (Å)	Structural information	Reference
NVpol (cryst. form II)	(+)ssRNA	*Caliciviridae*	1SH2	2.3	Crystal form II ($C222_1$) 1-5; 508-510 disordered	(Ng *et al.*, 2004)
NVpol (cryst. form III)	(+)ssRNA	*Caliciviridae*	1SH3	3.0	Crystal form III ($P2_12_12_1$) 1-5; 508-510 disordered + Mg^{2+}	(Ng *et al.*, 2004)
Φ6pol	dsRNA	*Cystoviridae*	1HI8	2.5	Full length (SeMet $P3_2$) + Mg^{2+}	(Butcher *et al.*, 2001)
Φ6pol	dsRNA	*Cystoviridae*	1HHS	2.0	Full length (native $P2_1$) + Mn^{2+}	(Butcher *et al.*, 2001)
Φ6pol ATP complex	dsRNA	*Cystoviridae*	1HI1	3.0	Full length + ATP; + Mn^{2+}	(Butcher *et al.*, 2001)
Φ6pol DNA complex	dsRNA	*Cystoviridae*	1HHT	2.5	Full length 5'-TTTCC-3' DNA + Mn^{2+}	(Butcher *et al.*, 2001)
Φ6pol initiation complex	dsRNA	*Cystoviridae*	1HI0	2.5	Full length, initiation 5'-TTTCC-3' DNA + GTP (2); + Mg^{2+}(2) + Mn^{2+}	(Butcher *et al.*, 2001)
Reovirus (type 3) λ3pol	dsRNA	*Reoviridae*	1MUK	2.5	Full length, apo 1; 957-964; 1266-167 disord.	(Tao *et al.*, 2002)
Reovirus (type 3) λ3pol mRNA cap analogue	dsRNA	*Reoviridae*	1MWH	2.5	Full length 1; 957-964; 1266-167 disord. + mRNA cap analogue GPPG; + Mn^{2+}	(Tao *et al.*, 2002)
Reovirus (type 3) λ3pol Initiation complex	dsRNA	*Reoviridae*	1N1H	2.5	Full length, initiation 1; 957-964; 1266-167 disord. 5'-AUUAGC-3' RNA + 3'dCTP; 3'dGTP; GDP; N7-metGMP; Mn^{2+}(2)	(Tao *et al.*, 2002)
Reovirus (type 3) λ3pol Elongation complex I	dsRNA	*Reoviridae*	1N38	2.5	Full length, elongation I (1 added bond) 1; 957-964; 1266-167 disord. 5'-GC-3' 3'-AUUGC-5' + 3'dUTP; 3'dCTP; Mn^{2+}(2)	(Tao *et al.*, 2002)
Reovirus (type 3) λ3pol Elongation complex II	dsRNA	*Reoviridae*	1N35	2.8	Full length, elongation II (4 added bonds) 1; 957-964; 1266-167 disord. 5'-GGGG-3' 3'-AUUGCCCCC-5' + 3'dCTP; Mn^{2+}(2)	(Tao *et al.*, 2002)

PV – Poliovirus; **HRV** – Human Rhinovirus; **FMDV** – Foot –and-mouth Disease Virus; **HCV** – Hepatitis C virus; **BVDV** – Bovine Viral Diarrhoea Virus; **RHDV** – Rabbit Hemorrhagic Disease Virus; **NV** – Norwalk Virus; **Φ6** – Bacteriophage Φ6


A
Fingertips
Thumb
Fingers
Palm
PV 3Dpol
Fingertips
Thumb
Fingers
Palm
HRV 3Dpol
Fingertips
Thumb
Fingers
Palm
FMDV 3Dpol
B
Fingertips
Thumb
C-term
Fingers
Palm
HCVpol
N-term
Fingertips
Thumb
β-thumb
Fingers
Palm
BVDVpol
C
Fingertips
Thumb
Fingers
Palm
RHDVpol
Fingertips
Thumb
C-term
Fingers
Palm
NVpol
D
Fingertips
Thumb
Fingers
C-term
Palm
Φ6pol
Thumb
Fingertips
N-term
Fingers
C-term
Palm
Reo λ3pol

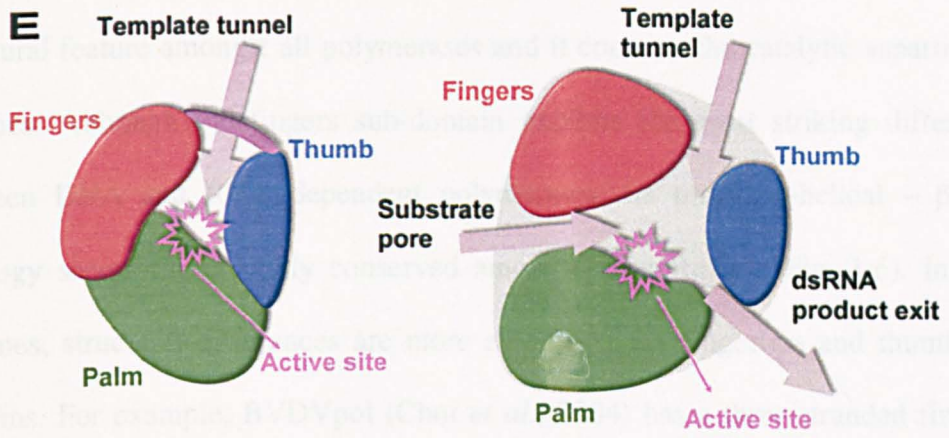


Figure 1.6. Structure of vRdRPs

A. (+)ssRNA *Picornaviridae* RdRPs from Poliovirus (Thompson and Peersen, 2004) (left; PV 3Dpol), Human Rhinovirus (Love *et al.*, 2004) (centre; HRV 3Dpol) and Foot-and-Mouth Disease Virus (Ferrer-Orta *et al.*, 2004) (right; FMDV 3Dpol)

B. (+)ssRNA *Flaviviridae* RdRPs from Hepatitis C Virus (Ago *et al.*, 1999) (left; HCVpol) and Bovine Viral Diarrhoea Virus (Choi *et al.*, 2004) (right; BVDVpol)

C. (+)ssRNA *Caliciviridae* RdRPs from Rabbit Hemorrhagic Disease Virus (Ng *et al.*, 2002) (left; RHDVpol) and Norwalk Virus (Ng *et al.*, 2004) (right; NVpol)

D. dsRNA virus RdRPs from the *Cystoviridae* (Butcher *et al.*, 2001) (left; Φ6 bacteriophage Φ6pol) and *Reoviridae* (Tao *et al.*, 2002) families (right; Reovirus λ3pol).

All polymerases are represented from a similar view point, with the fingers domain to the left and the thumb domain to the right (as in Fig. 1.2). Right hand fold sub-domains are highlighted: palm – green, fingers – red, thumb – blue. The characteristic RdRPs fingertips sub-domain is coloured magenta. Where it exists, the C-terminal sub-domain is coloured in yellow. Extra N-terminal sub-domains in BVDVpol (coloured orange) and λ3pol (coloured wheat) are also highlighted. In the reovirus λ3pol, the C-terminal sub-domain is shown in semi-transparent mode, for clarity. Catalytic aspartates are shown as ball-and-stick representation, coloured by atom (C – green, O – red, N – blue). The active site in each RdRP is highlighted by a magenta star.

E. Schematic representation of vRdRPs sub-domains (coloured as above), highlighting the different tunnels in the polymerases. Left panel: view as in Fig. 1.2. Right panel: Slice view of the vRdRPs, showing the position of template tunnel, substrate pore and exit route of the dsRNA product. The active site is highlighted by a star (magenta).

As described previously, the palm sub-domain is the most conserved structural feature amongst all polymerases and it contains the catalytic aspartic acid residues. Although the fingers sub-domain exhibits the most striking differences between DNA and RNA dependent polymerases, its mixed α-helical – β-sheet topology seems to be highly conserved amongst viral RdRPs (Fig. 1.6). In these enzymes, structural differences are more evident at the fingertips and thumb sub-domains. For example, BVDVpol (Choi *et al.*, 2004) has a three-stranded fingertip structure whilst HCVpol (Bressanelli *et al.*, 1999) and caliciviral polymerases (Ng *et al.*, 2002; Ng *et al.*, 2004) have four-stranded structures and Φ6pol (Butcher *et al.*, 2001) has a more elaborate fingertip sub-domain with six strands. Significantly, the differences in the thumb sub-domain might correlate with the mode of initiation, as described in detail in section 1.4.3. Moreover, some vRdRPs, such Φ6pol (Butcher *et al.*, 2001), HCVpol (Bressanelli *et al.*, 1999), BVDVpol (Choi *et al.*, 2004) and NVpol (Ng *et al.*, 2004) have an extra feature at their C-termini that confers on them a more spherical appearance (Fig. 1.6). This C-terminal structural feature seems to also be relevant to the initiation mode adopted by the polymerase, as discussed in detail below.

The clear structural homologies between vRdRPs extend beyond viral classes, with striking similarities between dsRNA and (+)ssRNA virus polymerases such as Φ6pol, HCVpol and BVDVpol (Fig. 1.6). These (+)ssRNA polymerases seem to share more structural features with dsRNA *Cystoviridae* Φ6pol than with other (+)ssRNA such as picornavirus polymerases. Φ6pol (Butcher *et al.*, 2001), HCVpol (Bressanelli *et al.*, 1999), BVDVpol (Choi *et al.*, 2004) and NVpol (Ng *et al.*, 2004) all have a structural feature at their C-termini that occludes the template

tunnel exit, implying that considerable structural rearrangements are necessary for the dsRNA product to be translocated. Conversely, picornaviral RdRPs structurally determined so far (Table 1.1) exhibit a wider, more open template tunnel (Fig. 1.6). These differences are likely to be related to the preferred initiation mode adopted by each viral class. Picornaviruses use a protein primer-dependent initiation mode and therefore need to accommodate a bulkier template in a wider tunnel. Conversely, dsRNA, flavivirus and calicivirus initiate replication in a primer-independent (*de novo*) mode and the C-terminal structures are thought to play a role in stabilizing the initiation competent complex (described below).

Surprisingly, reovirus polymerase, the other example of a dsRNA vRdRP structurally determined to date, exhibits a much more elaborate structure (Tao *et al.*, 2002) (Fig. 1.7.A). It has a "cage-like" appearance that seems to be unique amongst vRdRPs and is a result of an elaboration on the canonical right hand fold. λ3pol has two extra domains: an N-terminal that bridges the fingers and thumb sub-domains at one side of the catalytic cleft and a C-terminal "bracelet" sub-domain on the opposite side (Fig. 1.7.A). Four channels are defined by these elaborated features: "front", defined by the C-terminal bracelet opening; "left", at the interface of the bracelet and polymerase domains; "rear", at the interface of the polymerase and N-terminal domains and "bottom", at the interface of all three domains (Fig. 1.7.B). The front tunnel is the exit route of the dsRNA product, whilst the left pore defines the template tunnel and the rear channel forms the substrate tunnel. The bottom tunnel has been proposed to play a role in mRNA exit, but its precise function is still unknown (Tao *et al.*, 2002). This more elaborate structure of reovirus λ3pol is thought to be related with the fact that the enzyme also catalyses RNA capping and

the need for a structural organization that allows control of both functions at different stages of the viral life cycle (Tao *et al.*, 2002).

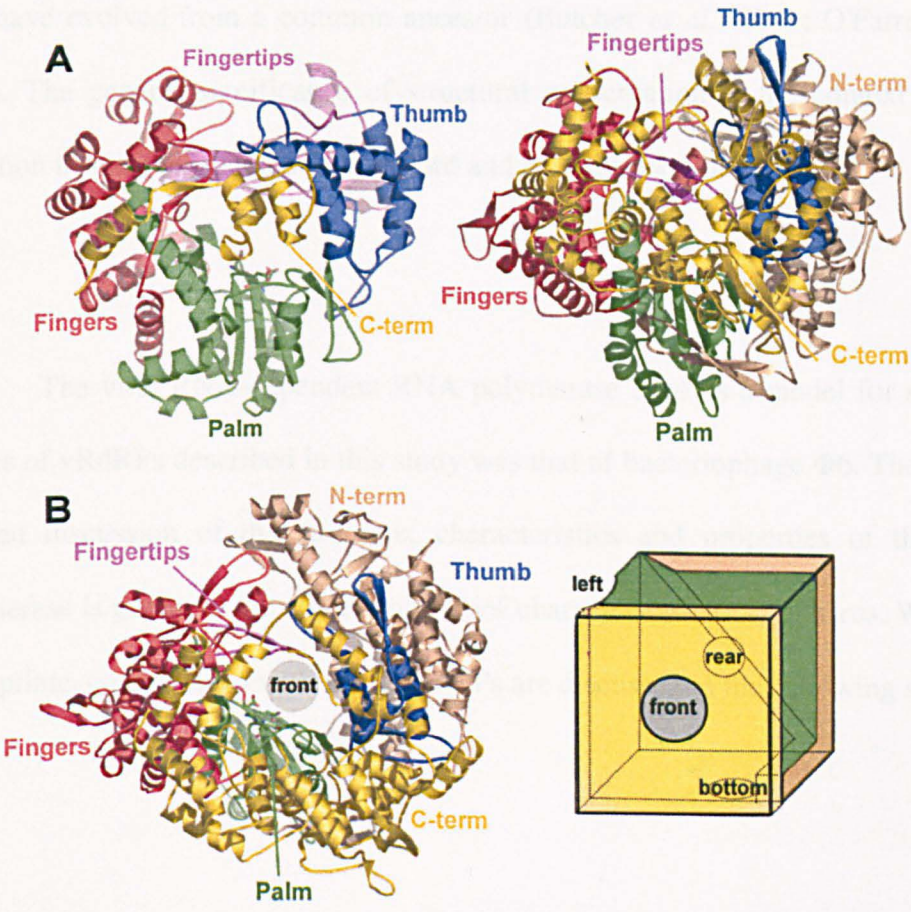


Figure 1.7. Comparison of structures of RdRPs from dsRNA viruses

A. dsRNA virus RdRPs from the *Cystoviridae* (Butcher *et al.*, 2001) (left; Φ6 bacteriophage Φ6pol) and *Reoviridae* (Tao *et al.*, 2002) families (right; Reovirus λ3pol). The C-terminal in λ3pol forms an elaborate "bracelet" structure that encircles the active cleft at the front of the protein, whilst the N-terminal bridges the fingers and thumb domain on the opposite side of the catalytic cleft. This creates a "cage-like" structure that encloses the canonical right hand fold within a largely hollow centre.

B. "Cage-like" conformation of the λ3pol. Left panel: view from the tunnel created by the C-terminal "bracelet". Domains are coloured as above. The "front" tunnel is shown as a grey circle. Right panel: schematic representation of λ3pol structure, with the positions of the four channels highlighted, viewed as in the left panel. C-terminal bracelet is shown in yellow, the N-terminal at the "back" in wheat colour and the catalytic right hand domain in green.

The overall structural similarity and the conservation of secondary and tertiary structure elements in the palm and thumb domains of polymerases of the families *Picorna-*, *Flavi-*, *Cysto-* and *Retroviridae* has led to speculation that they may have evolved from a common ancestor (Butcher *et al.*, 2001; O'Farrell *et al.*, 2003). The general significance of structural conservation in the context of virus evolution has been discussed by Bamford and co-workers (2002; 2003).

The viral RNA-dependent RNA polymerase used as a model for structural studies of vRdRPs described in this study was that of bacteriophage Φ6. Therefore, a detailed discussion of the structure, characteristics and properties of the phage polymerase is given, in the context of a brief characterisation of the virus. Whenever appropriate, comparisons with other vRdRPs are discussed in the following sections.

1.4. dsRNA viruses - *Cystoviridae*

Viruses with double-stranded RNA (dsRNA) genomes infect a wide variety of hosts: bacteria, fungi, plants and animals, including humans. There are six dsRNA viruses families: *Cystoviridae*, *Reoviridae*, *Birnaviridae*, *Totiviridae*, *Partitivridae* and *Hypoviridae*. Their genomes are either monopartite (*Totiviridae*, *Partitivridae* and *Hypoviridae*) or segmented – *Birnaviridae* (2 segments), *Cystoviridae* (3 segments) and *Reoviridae* (10-12 segments). As they infect a cell, dsRNA viruses face two major problems: dsRNA is not recognised by the cellular replicative machinery and it induces a strong apoptotic response. Therefore, dsRNA viruses have developed a protective strategy that involves retention of their genome in a closed icosahedral particle. One of the constituents of this particle is the RNA-dependent RNA polymerase (RdRP).

One of the best studied dsRNA virus has been bacteriophage Φ6 from the *Cystoviridae* family. It shares many of the key mechanistic characteristics of more complex dsRNA viruses, such as rotavirus and bluetongue virus, causal agents of major human and animal disease, but it can be handled without risk to humans or other vertebrates. Bacteriophages Φ7-Φ13 from the *Cystoviridae* family have been studied recently, with identified polymerase complexes similarly to that of Φ6pol (Mindich *et al.*, 1999; Hoogstraten *et al.*, 2000; Qiao *et al.*, 2000; Sun *et al.*, 2003). The biochemical properties of these polymerases are briefly discussed, within the context of the Φ6pol description since this was the vRdRP studied in this work and is therefore the major focus of this thesis. A brief characterisation of Φ6 bacteriophage is also given.

1.4.1. Bacteriophage Φ6

Bacteriophage Φ6 is the prototype virus of the family *Cystoviridae*. Viruses belonging to this family infect bacteria and have a spherical morphology with a lipid envelope surrounding the icosahedral nucleocapsid (Fig. 1.8.).

The bacteriophage Φ6 virion is made up of 3 concentric layers:

(i) Core or polymerase complex (PC) – dsRNA surrounded by the protein components of the PC

(ii) Nucleocapsid (NC) – core + P8 coat (Bamford *et al.*, 1976; Butcher *et al.*, 1997)

(iii) Lipid envelope (Mindich and Bamford, 1988)

Core particles are actively transcribing units, comprising the genome and the polymerase complex. Procapsids are the newly produced empty polymerase complex particles (Butcher *et al.*, 1997; de Haas *et al.*, 1999). The capsid shell is comprised principally of the major capsid P1 protein, with hexamers of the RNA translocase protein P4 sitting at the 5 fold vertices (Fig. 1.8). The positions of P2, the polymerase, and P7 are less certain, although it has been proposed that P2 might be internalised beneath the 5-fold vertices (de Haas *et al.*, 1999; Poranen *et al.*, 2001a). A similar architecture is observed for members of the *Reoviridae*, in particular bluetongue virus (BTV) (Grimes *et al.*, 1998; Diprose *et al.*, 2001; Nason *et al.*, 2004) and rotavirus (Prasad *et al.*, 1996). Apart from its structural role, Φ6 P1 has been implicated in the specific recognition of Φ6 RNA packaging signals (Onodera *et al.*, 1998). P4 has an NTP-binding consensus motif characteristic of helicases and NTPases and the protein has been shown to hydrolyse NTP and to unwind dsRNA when part of the polymerase complex. It is, therefore, thought to package ssRNA

genome precursors into the PC at the expense of NTP hydrolysis. Mancini *et al.* (2004) determined the structure of the related bacteriophage Φ12 P4 protein, together with several complexes of intermediate states of RNA translocation driven by ATP hydrolysis. They propose a mechanism that explains how ATP hydrolysis at external sites is coordinated to the translocation of RNA, located at the centre of the ring of the hexameric P4. This mechanism provides a highly efficient unidirectional motor driven by well-defined and localized conformational changes triggered by ATP hydrolysis (Mancini *et al.*, 2004).

P7 has been shown to modulate the RNA metabolism of Φ6, although its exact role is still uncertain. It has been proposed that it is involved in regulation of the RNA packaging reaction and in dsRNA transcription. Φ6 protein P2 has been shown to have both transcriptase and replicase activities *in vitro* and was hence identified as the phage RdRP (Makeyev and Bamford, 2000a, 2000b).

The genome of Φ6 is internalised within the core particle, organised as three dsRNA segments: small (S), medium (M) and large (L), with each virus particle containing one copy of each segment (Day and Mindich, 1980). Their lengths are 2948bp, 4063bp and 6374bp, respectively (McGraw *et al.*, 1986; Gottlieb *et al.*, 1988; Mindich *et al.*, 1988). All segments contain a conserved regulatory sequence at their 5' and 3' termini. Segment L codes for all the components of the core particle, whereas the other genes are shared between S and M. This clustering of sequences coding for the core proteins allows Φ6 to modulate their synthesis during its life cycle.

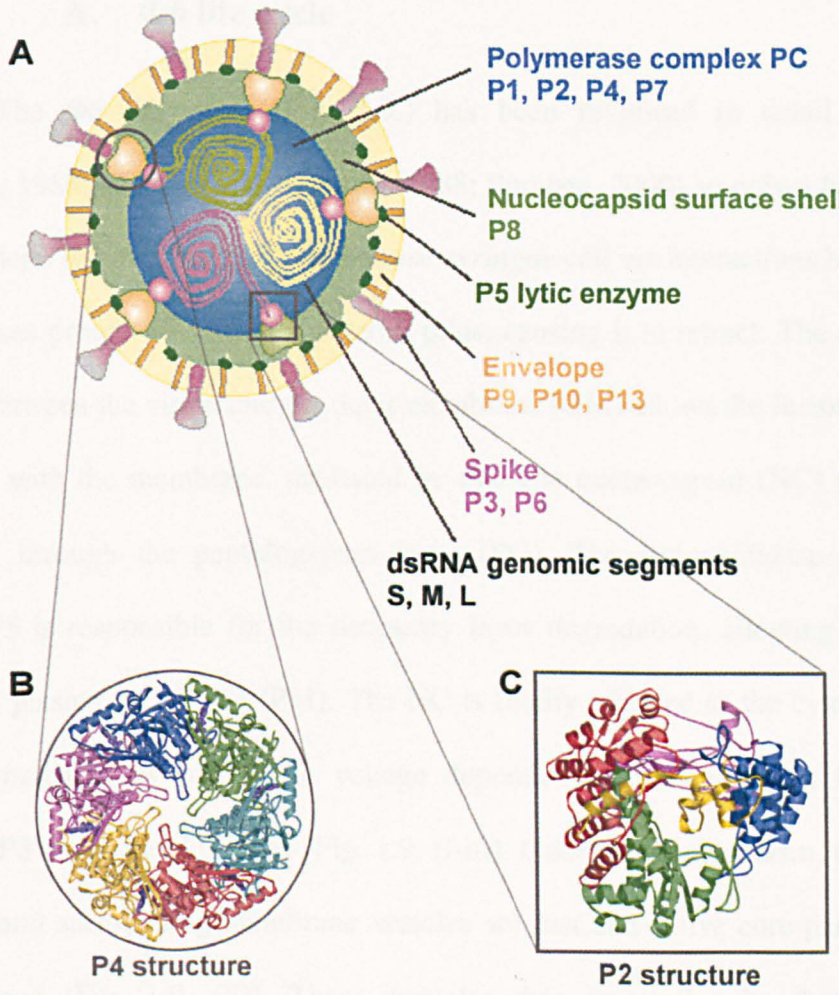


Figure 1.8. **The Φ6 virion**

A. Schematic representation of the virion. **PC** - Polymerase complex: P1 – shell constituent; P2 – RdRP; P4 – packaging NTPase; P7 – modulator. **Nucleocapsid**: core + surrounding P8 coat, with P5 (endopeptidase) in the T=13 sites of the P8 layer. **Lipid envelope**: P6, P9, P10, P13 integral membrane proteins and phospholipids, with P3 cell attachment protein anchored to P6, forming the outer spikes.

B. Cartoon representation of the hexameric structure of the P4 helicase (Mancini *et al.*, 2004). View is from the "external" side of the hexamer, as seen by the RNA being translocated. Monomeric units are coloured yellow, red, cyan, green, blue, magenta (clockwise). [Figure kindly provided by E. Mancini.]

C. Cartoon representation of Φ6pol (Butcher *et al.*, 2001), as seen in Fig. 1.6.D. Canonical palm (green), fingers (red) and thumb (blue) and characteristic fingertips (magenta) and C-terminal sub-domains are highlighted.

A. Φ6 life cycle

The Φ6 life cycle (Fig. 1.9.) has been reviewed in detail elsewhere (Mindich, 1988; Mindich and Bamford, 1988; Poranen, 2000) so only a brief outline is given here. Φ6 attaches to *Pseudomonas syringae* cell via interactions between the outer spikes protein P3 and the bacterial pilus, causing it to retract. The established contact between the virion and the outer membrane (OM) allows the fusion of the Φ6 envelope with the membrane, mediated by P6. The nucleocapsid (NC) then has to penetrate through the peptidoglycan layer (PG). The endopeptidase activity of protein P5 is responsible for the necessary layer degradation, allowing the NC to reach the plasma membrane (PM). The NC is finally released to the cytoplasm in a process mediated by membrane voltage dependent and independent interactions between P8 protein and the PM [Fig. 1.9. (i-ii)]. Once in the cytoplasm, the external P8 shell and surrounding membrane vesicles are lost and active core particles (CP) are released [Fig. 1.9. (ii)]. These particles then transcribe the three genomic segments S, M and L, extruding the (+) sense copies (s^+, m^+ and l^+) into the cytoplasm [Fig. 1.9. (iii)]. The cellular translating machinery uses them as mRNAs [Fig. 1.9. (iv)] to synthesise Φ6 viral proteins. The same s^+, m^+ and l^+ transcripts can also act as precursors of genomic dsRNA segments. In order for replication to occur, the single-stranded transcripts are packaged into newly synthesised and preformed PCs [Fig. 1.9. (v)] where Φ6pol synthesises the daughter strand [Fig. 1.9. (vi-vii)]. These PCs, now packed with dsRNA, can then direct additional rounds of transcription [Fig. 1.9. (viii)]. Alternatively, they can produce new NC and subsequently mature virions via a complex morphogenesis pathway. At that stage, phage proteins P10, P5, and probably some other uncharacterized factor(s) disrupt

the cell membranes and lyse the cell, releasing new infective particles [Fig. 1.9. (ix-x)].

Analysing the life cycle of Φ6, it becomes clear that the virus must have a mechanism to switch from transcription to replication. *In vitro* studies revealed the differences between replication and transcription that could account for the regulation of the viral RNA metabolism (Makeyev and Bamford, 2000b). However, the molecular basis of both processes is the same: reading of a ssRNA template and synthesis of the complementary strand. In replication, (+)ssRNA is directly read to produce the dsRNA product. However, for transcription to occur, unwinding of dsRNA strand is necessary so that the polymerase can synthesise the (+)ssRNA product.

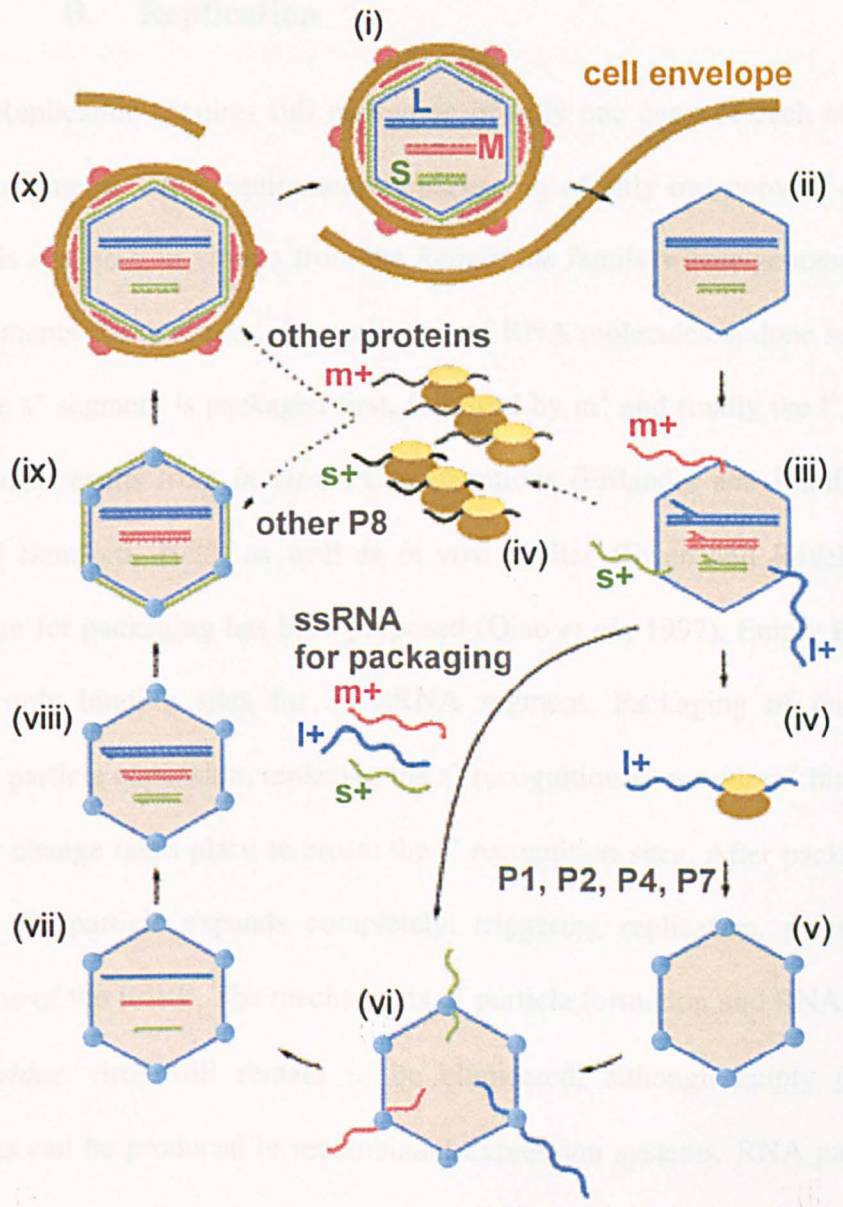


Figure 1.9. **Φ6 life cycle**

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

B. Replication

Replication requires full packaging of only one copy of each of the three genomic segments. This requirement of packaging of only one copy of each RNA segment is also seen in viruses from the *Reoviridae* family whose genomes have 10 to 12 segments. Furthermore, the packaging of RNA molecules is done sequentially in Φ6: the s^+ segment is packaged first, followed by m^+ and finally the l^+. Based on experimental results from *in vitro* PC preparations (Frilander and Bamford, 1995; Juuti and Bamford, 1995) as well as *in vivo* studies (Ewen and Revel, 1990), a mechanism for packaging has been proposed (Qiao *et al.*, 1997). Empty PC initially contains only binding sites for s^+ ssRNA segment. Packaging of this segment promotes particle expansion, replacing the s^+ recognition sites with m^+ binding sites. A similar change takes place to create the l^+ recognition sites. After packaging of l^+ segment, the particle expands completely, triggering replication, presumably by stimulation of the RdRP. The mechanisms of particle formation and RNA packaging in *Reoviridae* virus still remain to be elucidated, although empty polymerase complexes can be produced in recombinant expression systems. RNA packaging in *Reoviridae* occurs within cytoplasmic inclusions (viroplasms), and is facilitated by virus-encoded non-structural proteins (see Patton and Spencer, 2000 and references therein). Non-structural proteins NSP2 in rotavirus, NS2 in BTV and σNS in orthoreovirus form large multimeric complexes co-localized with the PCs in infected cells and bind ssRNA with high affinity. It has been postulated that these complexes could be functional analogues of P4 in Φ6, implicated in the RNA packaging. Therefore, a similar mechanism to that proposed to occur in Φ6 could be relevant in

reoviruses, although it is still poorly understood how the segmented dsRNA viruses can specifically select for one copy of each segment to be packaged sequentially in order to reconstitute a functional genome (reviewed in Patton and Spencer, 2000).

Φ6 PCs support end-to-end polymerization, where ssRNA and dsRNA templates are copied without the loss of genetic information. Importantly, isolated Φ6pol retains this property initiating RNA synthesis at the very 3′-terminal nucleotide of the template (Makeyev and Bamford, 2000a). Replicase activity of Φ6pol is stimulated by Mn^{2+}concentrations up to 2mM, above which it is inhibited. A similar effect is observed in related enzymes of Φ8, Φ12 and Φ13 (Yang *et al.*, 2001; Yang *et al.*, 2003a; Yang *et al.*, 2003b). Furthermore, a stimulatory effect of manganese has been reported for other polymerases: Qβ replicase (Blumenthal and Carmichael, 1979; Blumenthal, 1980), HCVpol (Alaoui-Lsmaili *et al.*, 2000; Zhong *et al.*, 2000), brome mosaic virus polymerase (BMVpol) (Sun *et al.*, 1996), and polioviral 3Dpol (Arnold *et al.*, 1999). In addition, Φ6pol requires a relatively high NTP concentration for optimal polymerisation, as observed for other vRdRPs that use *de novo* initiation, namely Qβ replicase (Blumenthal, 1980), BMVpol (Kao and Sun, 1996), BVDVpol (Kao *et al.*, 1999), and cellular and viral DdRPs (Losick and Chamberlin, 1976). Significantly, these stimulatory effects are also observed in the replication within Φ6 PC (Emori *et al.*, 1983; Bamford *et al.*, 1995; van Dijk *et al.*, 1995). Importantly, Φ6pol replication activity is inhibited by Ca^{2+} at concentrations higher than 0.5mM (Makeyev and Bamford, 2000b).

All genomic segments have a similar efficiency rate of initiation of replication, which implies that it does not depend on any signal sequence at the (+)ssRNA 3 end. Indeed, isolated vRdRPs from Φ6 and other cystoviruses accept

many heterologous ssRNA templates *in vitro*. The yield of dsRNA product depends on the template 3′-terminal sequence and template secondary structure. Studies with different chimeric ssRNAs (Makeyev and Bamford, 2001; Yang *et al.*, 2001) revealed that, despite the differences in the template preferences between Φ6, Φ8 and Φ13 polymerases, a general preference for 3′-terminal cytosines is observed. These results strongly suggest that replication efficiency is controlled at the initiation step, and that vRdRPs prefer pyrimidine-rich 3′-terminal initiation sites (C-3′ > U-3′). Indeed, similar template preferences are observed in Qβ replicase (Blumenthal and Carmichael, 1979), HCVpol (Sun *et al.*, 2000; Zhong *et al.*, 2000) and BVDVpol (Kao *et al.*, 1999). Studies of isolated Φ6pol initiation using RNA templates with insertions at the 3' end that can assume secondary structures such as hairpin-tetraloops (Laurila *et al.*, 2002) and highly structured ssRNAs (Makeyev and Bamford, 2000a) revealed a considerable lowered polymerisation activity. Therefore, Φ6pol seems to require a single-stranded 3' terminus to efficiently initiate reaction. Interestingly, Φ6pol can also accept ssDNA as a template *in vitro* (Makeyev and Bamford, 2001).

C. Transcription

Transcription in Φ6 is semi-conservative: the old (+) strand is displaced as the new is synthesised using the (-)ssRNA as a template. Significantly, this activity is less efficient *in vitro* than replication, at least by one order of magnitude, apparently due to inefficient initiation from a dsRNA template (Makeyev and Bamford, 2000b). Access to the 3'-terminus of the template strand in a single-stranded form in order to initiate transcription requires "unzipping" of the dsRNA terminus, which is apparently inefficient in the Φ6pol–dsRNA *in vitro* systems and must be somehow stimulated within the polymerase complex. Indeed, when assembled into the PC particle, Φ6pol can catalyze multiple rounds of transcription (Poranen *et al.*, 2001b). Past the initiation step, purified Φ6pol is capable of normal elongation, albeit at a lower rate (Makeyev and Bamford, 2000b). Therefore, Φ6pol does not seem to require the assistance of other proteins to unwind RNA duplex during elongation. This might be a common feature amongst RdRPs, since poliovirus and alfalfa mosaic virus polymerases can also displace non-template strand during elongation (Cho *et al.*, 1993; de Graaff *et al.*, 1995).

Importantly, genomic Φ6 S and M segments are transcribed more efficiently than the L segment, especially at later stages of infection when the virus needs to assemble complete particles instead of producing more NCs. These different efficiency rates for initiation of transcription are probably related to the (-)ssRNA 3' terminal sequence: 5'-...CC- 3' for the S and M segments and 5'-...AC-3' for the L segment, likely to be related to the preference for 3' terminal cytidines described above. The facts that the L segment encodes for all the proteins of the polymerase

complex and has different (-)ssRNA 3' end sequence are thought to be related to mechanisms of control during the viral life cycle.

Transcription is stimulated by Mg^{2+} and inhibited by concentrations of Ca^{2+} as low as 0.2mM. This effect has also been observed in *in vitro* transcription of reovirus (Sargent and Borsa, 1984) and DNA-dependent RNA polymerase II (Okai, 1982).

The subtle differences between replication and transcription are closely related to the virus life cycle. The polymerase recognizes the 3'-end of the (+) sense RNA segment after RNA encapsidation and synthesizes the daughter (–) strand during genome replication, to produce the dsRNA genome. It then switches mode, preferentially recognising the 3'-end of the (–) sense RNA strand, to synthesise a daughter (+) strand. The parental (+) strand is displaced from the particle during this process of semi-conservative transcription. This is likely to be a general model for the role of RdRPs of dsRNA bacteriophages that transcribe RNA in a semi-conservative way.

1.4.2. Φ6pol structure

The structure of Φ6pol was determined by Butcher and co-workers (2001) in its apo form, in complex with a DNA template, with a bound ATP and in an initiation competent state. As mentioned previously, the polymerase of bacteriophage Φ6 is a compact spherical molecule, due to two elaborations of the basic right hand architecture (Fig. 1.10.A). The first is an N-terminal extension constituted by six strands that strap together the tips of the fingers and thumb (fingertips). The C-terminal 64 residues form the second elaboration. Two positively charged tunnels allow the access of NTP substrates and RNA template to the active site. In the following descriptions, the template oligonucleotides are numbered sequentially from 3' to 5', with the 3' end of the template (which is used to initiate *de novo* RNA synthesis) named as T1. Nucleotides of the daughter strand are denoted D1, D2, etc. from the 5' to the 3', such that D1 base pairs with T1, D2 with T2, and so on.

The template tunnel is wide enough to accommodate ssRNA but not dsRNA, and the distance from the surface to the active site can be spanned by a ssRNA oligonucleotide of 5nt. In fact, Butcher et al. (2001) determined the structure of Φ6pol with a 5nt DNA oligonucleotide template that completely fills the tunnel (Fig. 1.10.B). The DNA template sequence 5'-TTTCC-3' was chosen so it mimics the 3' end sequence of (-)ssRNA of the genomic S and M segments (5'-...CC-3'), the preferred templates for transcription (McGraw *et al.*, 1986; Gottlieb *et al.*, 1988; Yang *et al.*, 2001). The oligonucleotide binds inside a tunnel, lined with the side-chains of predominantly basic amino acids, that leads to the active site, in a position and conformation similar to that seen for template bound to HIV-1 RT (Huang *et al.*, 1998). Surprisingly, mapping the double-stranded DNA/DNA hybrid seen for the RT

ternary complex onto the Φ6pol structure places it exactly through the body of the C-terminal sub-domain. A similar physical barrier imposed by the C-terminus is present in HCVpol (Bressanelli *et al.*, 2002). Unexpectedly, the DNA template binds the polymerase in a position that places T1 well past the expected catalytic residues, with the base threaded into a specific binding pocket (denoted S) within the C-terminal sub-domain (Fig. 1.10.B). The described specificity of Φ6pol towards a cytidine at the 3' end is explained by interactions in the S pocket (a hydrogen bond between an O^{ε} of E634 and N4 of cytidine would discriminate against the O4 of uracil, and the observed pocket is too small to accommodate either adenine or guanidine). T2 also inserts beyond the catalytic residues, and is base stacked with T3, which seems to be aligned to engage the incoming substrate NTP (the catalytically relevant binding site for incoming NTPs is hereafter denoted as site C). Discrimination at T1 and T2 by structural elements present in the template tunnel explains how the virus regulates the level of transcription of each genomic strand and promotes the production of (+)ssRNA rather than (-)ssRNA from the dsRNA template.

The edge of the template channel is shaped like a plough, adjacent to a positively charged path over the polymerase surface (Fig. 1.10.C). The molecular surface around the entrance to the template tunnel is highly charged, rich in basic residues. Therefore, it seems suitable to establish non-specific interactions with the phosphate backbone of RNA. Butcher *et al.* (2001) proposed that the plough-like protuberance from this molecular surface might separate the strands of RNA, feeding one strand directly into the similarly basic template channel and the other out of the capsid, perhaps through a positively charged surface channel (Fig. 1.10.C). Selective

attachment of the minus strand in the template tunnel ensures that the correct (+) strand leaves the polymerase complex.

The NTP (substrate) tunnel, formed by the fingertips, is another common structural motif observed for RdRPs. This region is positively charged in a number of RdRPs and interacts with the incoming negatively charged NTPs. The structure of Φ6pol in complex with an ATP molecule (Butcher *et al.*, 2001) reveals a binding site in the substrate pore (hereafter denoted as site I) that overlaps with, but is displaced from by ~5 Å, the inferred NTP-binding site in the active site (site C). The interactions of triphosphate moieties of the incoming NTP with basic residues K223, R225, R268 and R270 might facilitate the interrogation of the correct NTP (Fig. 1.10.D).

These complexes, together with a crystal structure of an initiation competent state allowed Butcher and co-workers (2001) to propose the series of events that lead to RNA synthesis in Φ6pol and related vRdRPs, as described in detail in the following section.

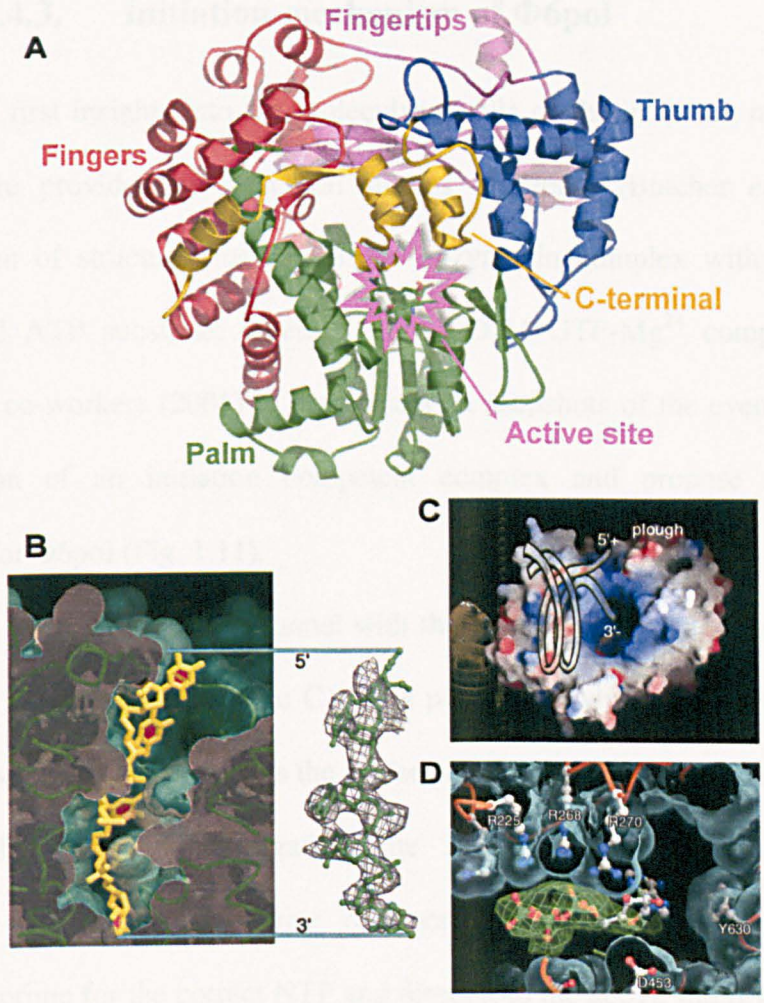


Figure 1.10. Φ6pol structure

A. Cartoon representation of Φ6pol structure. Right hand sub-domains are coloured: palm (green), fingers (red) and thumb (blue). The fingertips sub-domain characteristic of vRdRPs is coloured magenta. The C-terminal sub-domain is coloured yellow.

B. A section through the template channel with the bound 5nt DNA oligonucleotide drawn in yellow. The surface of the polymerase and the embedded polypeptide chain are coloured green. The electron density map covering the template is highlighted on the right, with the template drawn in green as ball—and-sticks representations. The 3' and 5' ends are marked.

C. Surface charge representation of Φ6pol viewed from above, showing the entrance to the template tunnel. Putative positions for the strands before initiation are shown.

D. Difference electron density map for the NTP bound to site I. Conserved residues R225, R268 and R270 line the tunnel and interact with the phosphate backbone from the ATP molecule. Catalytic aspartate D453 and Y630 are drawn. All the residues drawn and the ATP molecule are shown in ball-and-stick representation (C – white, N – blue, O – red, P – green).

[Modified from (Butcher *et al.*, 2001)]

1.4.3. Initiation mechanism of Φ6pol

The first insights into the molecular details of the initiation mechanism in vRdRPs were provided by structural studies of Φ6pol (Butcher *et al.*, 2001). Determination of structural models of the enzyme in complex with a 5nt DNA template and ATP substrate, together with a DNA-GTP-Mg^{2+} complex allowed Butcher and co-workers (2001) to obtain several snapshots of the events leading to the formation of an initiation competent complex and propose an initiation mechanism for Φ6pol (Fig. 1.11).

The template enters the tunnel with the 3' cytidine binding in a pocket (site S) well past the catalytic site (site C). This pocket lies within the C-terminal sub-domain and specifically recognizes the preferred cytidine T1. At the substrate tunnel, different NTPs occupy interrogation site I, presumably in rapid exchange. Interactions with a set of arginine residues at the fingertips are proposed to sequentially prime for the correct NTP and direct it to the active site (Fig. 1.11; steps II-IV). Once the correct nucleoside (GTP) is directed to the active site, it interacts with the initiation platform (P), constituted by the 629QYWK632 loop in the C-terminal sub-domain. Watson-Crick base-pairing interactions with the T2 cytidine of the template and stacking interactions with Y630 stabilize the GTP molecule (denoted D2) (Fig. 1.11; step V). Due to interactions with D2 at the initiation platform and electrostatic attraction to R268 and R270, the template ratchets back, causing T1 to be displaced from the specificity pocket (Fig. 1.11; step VI). A second GTP, D1, enters the P site to lock the initiation complex into its active form: D1 forms Watson-Crick base pair interactions with T1, D2 base-pairs with T2, further stabilized by base-stacking interactions with Y630 (Fig. 1.11; step VII). Two

magnesium ions are coordinated by the phosphate backbones of D1 and D2 and the catalytic aspartates (as described for the general "two metal ion mechanism"; see Fig. 1.3. for details). Once catalysis occurs, pyrophosphate is released and the polymerised product ratchets down, displacing the C-terminal sub-domain of the protein (Fig. 1.11; step VIII). This may be facilitated by attraction of the phosphates of the GTP in site P to the Mn^{2+} ion present in Φ6pol in a somewhat distant position (~6Å from the active site). The next NTP slips into the catalytic site C, from site I, which sets the ratchet for the chain elongation to start.

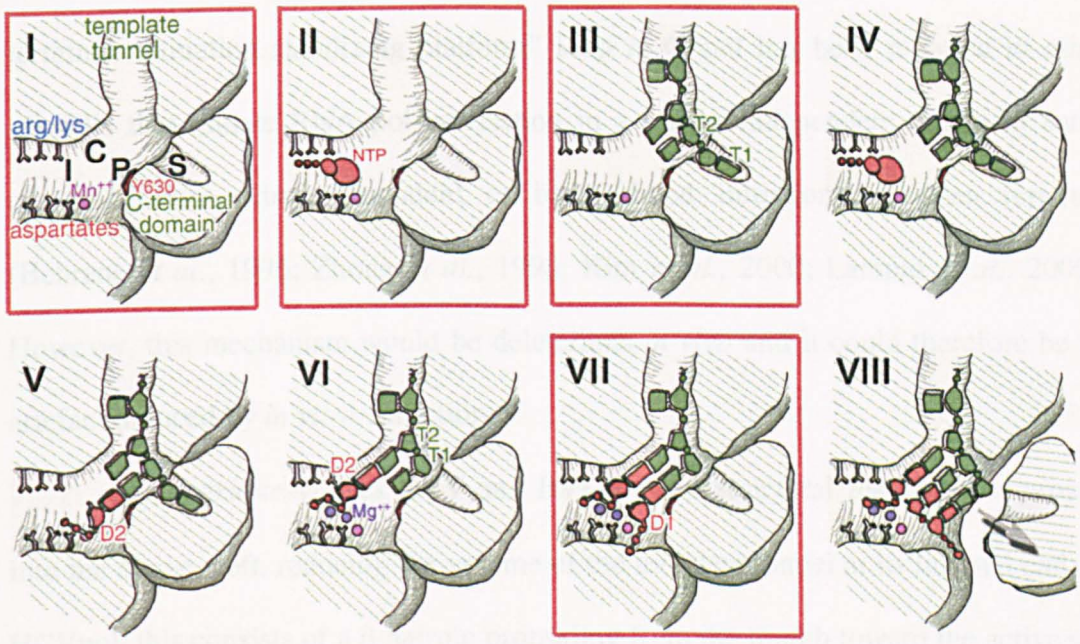


Figure 1.11. Φ6pol initiation model

I-IV - The template enters the tunnel, fitting snugly in site S (specificity pocket S). Site I (interrogation) is occupied by NTPs, presumably in rapid exchange. **V** - GTP binds to site P (initiation stabilizing platform). **VI** - The template ratchets backwards, moving T1 away from pocket S. **VII** - A second GTP binds into site P, locking the initiation complex into its final form. **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 sub-domain of the protein is then necessary for elongation to occur. Red boxes indicate stages captured by X-ray crystallography.

[From (Butcher *et al.*, 2001)]

The presence of structural elaborations that might be equivalent to the C-terminal "initiation stabilizing platform" loop in Φ6pol has been reported in other vRdRPs that initiate RNA polymerisation in a primer-independent mode. In some cases, *in vitro* primer-dependent or back-primed initiation has been reported (Behrens *et al.*, 1996; Zhong *et al.*, 1998; Kim *et al.*, 2000; Lampio *et al.*, 2000). However, this mechanism would be deleterious *in vivo* and it could therefore be an artefact induced by *in vitro* conditions.

Flaviviridae viruses HCV and BVDV have structural features that project into the active cleft, reducing the volume of the template tunnel at its internal end. In HCVpol, this consists of a β-hairpin protruding from the thumb toward the active site at the base of the palm domain (Bressanelli *et al.*, 1999; Lesburg *et al.*, 1999) (see Fig. 1.6). It allows only the single-stranded 3' terminus of an RNA template to bind productively to the active site and may function as a gate preventing the 3' terminus of the template RNA from slipping through the active site, thus ensuring terminal initiation of replication (Hong *et al.*, 2001). Recently, a regulatory motif in the C-terminal non-catalytic region of the HCV RdRP has been identified upstream of the membrane anchor domain. It comprises a unique conserved hydrophobic pocket, which protrudes into the RNA-binding cavity. Together with the β-hairpin it forms a rigid bulky loop at the active site that likely serves as an initiation platform similar to that of Φ6pol (Ranjith-Kumar *et al.*, 2003). It could also play a role in preventing primer-dependent and back-primed initiation and recognition of the correct secondary structure at the 3'-terminal end of the HCV genome (Leveque *et al.*, 2003). The thumb domain of BVDV exhibits a different, more elaborate topology. It has a structural element – the "β-thumb" region – constituted by two β-strands and

connecting loops that occlude the active site cavity in a similar way to the β-hairpin in HCV (Choi *et al.*, 2004) (see Fig. 1.6). It interacts with the fingers and palm sub-domains through a long C-terminal loop that fits snugly between the thumb and palm sub-domains. Thus, the β-thumb might have a similar function as the β-hairpin feature in HCV, allowing only ssRNA to access the active site during initiation. The crystal structures of the HCV and BVDV RdRPs indicate that there is not enough space underneath the β-hairpin or β-thumb to allow the passage of a dsRNA product. Therefore, theses features should be flexible enough to give way to the nascent dsRNA (Hong *et al.*, 2001). It was hypothesized that a change in conformation of the thumb domain takes place upon template binding to allow efficient *de novo* initiation of RNA synthesis (Bressanelli *et al.*, 2002; Choi *et al.*, 2004).

Viruses of the family *Reoviridae* also initiate RNA synthesis in a primer-independent mode. The reovirus RdRP also has a special priming loop proposed to have an effect in stabilizing the initiation competent complex (Tao *et al.*, 2002). This loop is a unique insertion within a strand present in the palm sub-domain that supports the stacking of the priming NTP, D1. In the apo structure and in the initiation complex, the loop protrudes into the active site. Conversely, in the fully active polymerase complex, the loop retracts towards the palm to fit into the minor groove of the product duplex (Tao *et al.*, 2002).

The observation of "initiation stabilizing platforms" in these polymerases indicates that the model proposed by Butcher et al. (2001) might be adopted by RdRP from other viral families, providing a possible evolutionary link between viral classes (Bamford *et al.*, 2002; Bamford, 2003).

Calicivirus and picornavirus initiate polymerization in a primer-dependent manner, using a protein-linked primer and template. The most dramatic difference between the Rabbit Hemorrhagic Disease Virus (RHDV) and poliovirus RdRPs and that of other polymerases occurs at the thumb sub-domain: primer-dependent polymerases have a smaller, less elaborate thumb and a wider template tunnel. The absence of "initiation stabilizing platform" loops in the poliovirus and RHDV enzymes is consistent with their ability to utilize dsRNA as templates *in vitro* (Arnold *et al.*, 1999; Hong *et al.*, 2001; Lopez Vazquez *et al.*, 2001) and the fact that they need to accommodate a bulkier template since they utilise a protein-primed initiation for genome replication (Paul *et al.*, 1998).

Supporting the view that the protruding loops observed in HCVpol, BVDVpol, λ3pol and Φ6pol are related to primer-independent initiation, it has been reported that initiation can revert to a back-priming mode when the size of the initiation platform is reduced. Thus, a Φ6pol mutant where the 630YKW632 loop was changed to smaller, less bulky residues GSG (Laurila *et al.*, 2002) was found to preferentially initiate by a back-priming mechanism. Hong and co-workers (2001) reported a similar result for HCVpol when the β-hairpin 443LDCQIYGACYSI454 was mutated to LGGI. This indicates that the structural features protruding into the active site are important for stabilizing the initiation complex thus insuring correct *de novo* initiation and preventing back-priming from occuring. Furthermore, it provides a structural explanation for its absence in vRdRPs that utilise primer-dependent mechanisms.

1.5. Cellular RNA-dependent RNA polymerases

RNA-dependent RNA polymerisation activity is usually associated with viral RNA replication and transcription. However, this activity has also been reported in uninfected cells. Two distinct activity types were identified: firstly, cellular DNA-dependent RNA polymerases can utilise RNA in particular circumstances and, secondly, RNA polymerisation can be catalysed by cellular encoded RNA-dependent RNA polymerases (cRdRP). Plant viroid replication and RNA genome replication of hepatitis Δ virus (Lai, 2005) by DNA-dependent RNA polymerase II (RNApolII) are examples of situations were cellular DdRPs are recruited to use RNA as a template. The first cellular encoded RdRP identified was that of the tomato plant (Schiebel *et al.*, 1993a, b). Since then, cRdRPs have been identified in several organisms, including plants, fungi, protozoa and some metazoans (Cogoni and Macino, 1999a; Sijen *et al.*, 2001; Makeyev and Bamford, 2002; Schramke *et al.*, 2005). In those studies, the key essential role of cRdRPs in RNA silencing was also identified.

1.6. RNA silencing mechanisms

RNA silencing, also known as RNA interference (RNAi), refers to a group of RNA-induced gene silencing mechanisms conserved in most eukaryotic organisms and playing essential roles in cellular immunity against viruses and transposons, aspects of development, modulation of chromatin structure, and some other processes such as genome rearrangement in ciliates (Baulcombe, 2004; Lippman and Martienssen, 2004; Mello and Conte, 2004). RNAi can generally function on two different levels, inducing either transcriptional gene silencing (TGS) via establishing repressed chromatin state or posttranscriptional gene silencing (PTGS) by degradation of target RNAs. A key step in known silencing pathways is the processing of dsRNAs into short RNA duplexes of characteristic size and structure. These short dsRNAs guide RNA silencing by specific and distinct mechanisms. Many components of the RNA silencing machinery still need to be identified and characterized, but a reasonably complete understanding of the process is emerging, as schematically represented in Fig. 1.12.

All studied RNA silencing pathways rely on the use of dsRNA triggers that are processed by a dsRNA-specific RNase-III-type endonuclease termed Dicer (Bernstein *et al.*, 2001b; Hutvagner *et al.*, 2001; Lee *et al.*, 2004). Long dsRNA and microRNAs (miRNAs) precursors are processed down to small interfering RNAs (siRNAs) that appear as 21-25nt long fragments with 2nt 3' protruding termini (Bernstein *et al.*, 2001a; Zhang *et al.*, 2004). One of the two siRNA strands is recruited by an effector complex comprising an Argonaute (Ago) subunit and used as a guide for sequence-specific degradation of target mRNAs (in PTGS) or directed

silencing of cognate chromatin domains (in TGS) (Hammond *et al.*, 2001; Song *et al.*, 2004a; Verdel *et al.*, 2004). In PTGS, the effector complex comprising the Argonaute protein together with the siRNA and other uncharacterised components is named RNA-induced silencing complex (RISC), whilst the similar complex directing heterochromatin silencing in TGS is termed RNA-induced initiation of transcriptional silencing (RITS). The single-stranded siRNA in RISC guides sequence-specific degradation of complementary or near-complementary target mRNAs (Martinez *et al.*, 2002; Martinez and Tuschl, 2004). RISC cleaves the target mRNA in the middle of the complementary region, ten nucleotides upstream of the nucleotide paired with the 5' end of the guide siRNA (Elbashir *et al.*, 2001b).

Dicer and Argonaute proteins contain a common RNA-binding Piwi/Argonaute/Zwille (PAZ) domain which is suggestive of evolutionary relatedness of the two enzymes. Recently, the crystal structure of an archae-bacterial Ago protein revealed striking similarity of the PIWI domain with members of the RNase H family (Song *et al.*, 2004b). As RNase H cleaves the RNA strand of RNA/DNA duplexes, it was proposed that Ago proteins act by cleaving target RNA in target RNA/siRNA hybrids. Recent biochemical and structural studies, however, converged on the view that PAZ is an RNA binding domain (RBD) that specifically recognizes the terminus of the base-paired helix of siRNA duplexes, including the characteristic 2-nucleotide 3' overhangs (Lingel *et al.*, 2004; Ma *et al.*, 2005). This siRNA-duplex specific interaction with PAZ ensures the safe transitioning of small RNAs into RISC by minimizing the possibility of unrelated RNA-processing or RNA-turnover products entering the RNA silencing pathway. The preferred recognition of the termini of dsRNA precursors by PAZ-domain-containing Dicer

(Zhang *et al.*, 2002) suggests that the processing reaction is guided by the PAZ domain docking at the terminus of long dsRNAs.

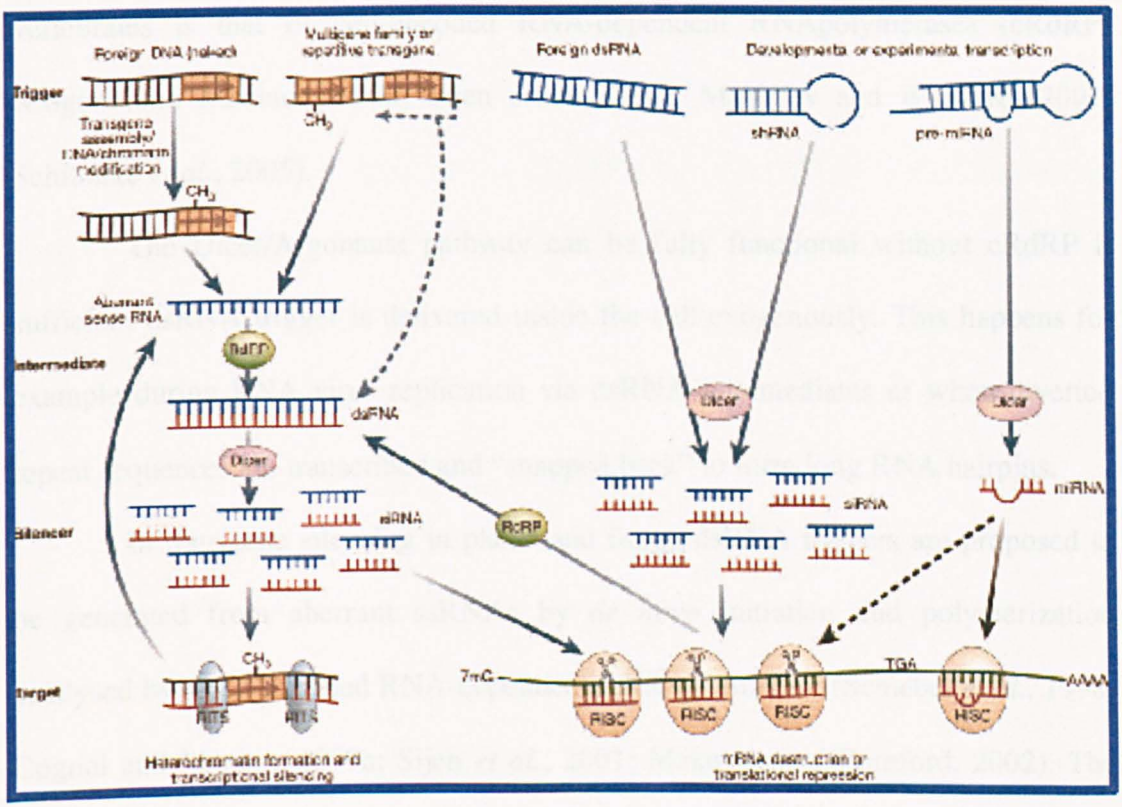


Figure 1.12. Model for RNA silencing mechanism

dsRNA can act as the initial trigger, for example when foreign dsRNA is introduced experimentally. In other cases, dsRNA acts as an intermediate, for example when 'aberrant' RNAs are copied by cellular RdRPs. Transcription can produce dsRNA by read-through from adjacent transcripts, as may occur for repetitive gene families or high-copy arrays (blue dashed arrows). Alternatively, transcription may be triggered experimentally or developmentally, for example in the expression of short hairpin (shRNA) genes and endogenous hairpin (miRNA) genes. The small RNA products of the Dicer-mediated dsRNA processing reaction guide distinct protein complexes to their targets. These silencing complexes include the RNA-induced silencing complex (RISC), which is implicated in mRNA destruction and translational repression, and the RNA-induced transcriptional silencing complex (RITS), which is implicated in chromatin silencing. Sequence mismatches between a miRNA and its target mRNA lead to translational repression (black solid arrow), whereas near perfect complementarity results in mRNA destruction (black dashed arrow). Feedback cycles permit an amplification and long-term maintenance of silencing. CH3, modified DNA or chromatin; 7mG, 7-methylguanine; AAAA, poly-adenosine tail; TGA, translation termination codon.

[From Mello and Conte, 2004]

Another protein family critical for RNA silencing in plants, fungi, protozoa, and some animals species like nematodes but apparently not in insects and vertebrates is that of cell-encoded RNA-dependent RNApolymerases (cRdRP) (Cogoni and Macino, 1999a; Sijen *et al.*, 2001; Makeyev and Bamford, 2002; Schramke *et al.*, 2005).

The Dicer/Argonaute pathway can be fully functional without cRdRP if sufficient dsRNA trigger is delivered inside the cell exogenously. This happens for example during RNA virus replication via dsRNA intermediates or when inverted repeat sequences are transcribed and "snapped back" to form long RNA hairpins.

In transgene silencing in plants and fungi, dsRNA triggers are proposed to be generated from aberrant ssRNAs by *de novo* initiation and polymerization catalysed by a cell-encoded RNA-dependent RNA polymerase (Schiebel *et al.*, 1998; Cogoni and Macino, 1999a; Sijen *et al.*, 2001; Makeyev and Bamford, 2002). The cRdRP is also thought to play a role in the amplification of dsRNA triggers, thus making the RNAi response more robust. However, the exact mechanism of this process still remains unknown. In addition to their role in cytoplasmic PTGS, cRdRPs have also been implicated in RNA-dependent DNA methylation. In *Arabidopsis*, SDE1/SGS2/RDR6 proteins are required for this process to occur (Mourrain *et al.*, 2000). Furthermore, recent studies in fission yeast *S. pombe* show that Rdp1, the only cRdRP gene present, is essential for TGS at centromeric repeats and the silent mating-type region, together with two other proteins (Motamedi *et al.*, 2004).

Several species seem to encode more than one cRdRP: three in *N. crassa* and *Dictyostelium*, four in *C. elegans* and as many as six genes in *Arabidopsis*.

Importantly, in *C. elegans*, one of the cRdRPs, - EGO-1 - catalyzes RNAi in the germ line, whist RRF-1 is reserved for somatic RNAi (Sijen *et al.*, 2001). Conversely, two different cRdRPs of *Arabidopsis* - RDR1 and SDE1/SGS2/RDR6 – are thought to play non-redundant roles during virus-induced gene silencing caused by different viral species (Mourrain *et al.*, 2000; Xie *et al.*, 2004). These observation support the hypothesis that different cRdRPs might support silencing in different tissues or/and be used in different silencing mechanisms.

1.6.1. Quelling in Neurospora Crassa

Quelling, as the RNA silencing mechanism has been named in *N. crassa*, is one of the best studied models for RNA silencing. It was originally described as reversible inactivation of gene expression by transformation with repeated homologous sequences. Quelling occurs during the vegetative phase of growth and, as for co-suppression in plants, it affects both transgenes and endogenous genes (Romano, 1992).

Three genes have been identified in the initial genetic screen for quelling defective phenotype: *qde-1*, *qde-2* and *qde-3* (Pickford *et al.*, 2002). The *qde-1* mutant was defective in an RNA-dependent RNA polymerase (RdRP) (Cogoni and Macino, 1999a). The protein product encoded by *qde-2* was shown to correspond to the Ago protein in *Neurospora* (Catalanotto *et al.*, 2000). The *qde-3* gene encodes a putative RecQ-type DNA helicase (Cogoni and Macino, 1999b). Recently, QDE-3 RecQ helicase and its homologue, RecQ-2, have been shown to play a role in recombination repair (Pickford *et al.*, 2003; Kato *et al.*, 2004), suggesting that QDE-3 RecQ helicases may have a dual role in *N. crassa*. Two Dicer-like proteins - DCL-1 and DCL-2 – that play overlapping roles in the generation of siRNAs have been reported (Catalanotto *et al.*, 2004). Although *Neurospora* supports chromatin-based silencing via locus-specific methylation of lysine 9 on histone H3 and DNA methylation, this level of regulation does not seem to interact with quelling (Chicas *et al.*, 2005).

1.6.2. QDE-1 cellular RNA-dependent RNA polymerase

Makeyev and Bamford (2002) have recently isolated recombinant QDE-1, providing direct evidence of RNA-dependent RNA polymerisation activity involvement in quelling. The polymerase activity of QDE-1 resides in the sub-domain since deletion of the N-terminal residues 1-376 has no detectable effect on the activity of the protein (Makeyev and Bamford, 2002). The catalytic active truncated version of the protein – QDE-1 ΔN (residues 377-1402) – accepts a number of ssRNA templates *in vitro*, producing two distinct types of RNA products:

(i) full-length copies

(ii) short RNA oligonucleotides complementary to the template, spread throughout the template

For the synthesis of full length products, the polymerase can initiate RNA polymerisation in a *de novo* or back-priming manner, depending on the secondary structure of the 3' end of the template. These long dsRNAs can then be degraded by a Dicer-like ribonuclease into 21–25nt long siRNAs (Makeyev and Bamford, 2002). Surprisingly, QDE-1 can also initiate *de novo* RNA synthesis internally, producing short copies of input ssRNA of variable length – 9 to 21nt. Furthermore, *in vitro*, this reaction is considerably more efficient than the synthesis of the full-length dsRNA (Makeyev and Bamford, 2002). If QDE-1 uses this reaction mode *in vivo*, at least a subset of the QDE-1 reaction products, close to 19–21bp, would provide an ideal target for a RISC-like nuclease complex, thus inducing a localized mRNA cleavage (Elbashir *et al.*, 2001a; Elbashir *et al.*, 2001b; Sharp, 2001). However, relevance of these products and this reaction mode *in vivo* has yet to be proven. Surprisingly, dsRNA substrates are not recognized by recombinant QDE-1 *in vitro*. Moreover,

primer extension *in vitro* is inefficient as compared with other RdRPs from viral origin. These results may indicate that the primer extension is not a major function of QDE-1 *in vivo*. Taken together, these observations indicate that QDE-1 may be needed for the synthesis of long dsRNA triggers and/or production of small guide RNAs.

Recent studies have provided further insight into the function of QDE-1 in *N. crassa*. Interestingly, QDE-1 seems to no longer be required upon the direct expression of dsRNA (Catalanotto *et al.*, 2004). This suggests that the main role of cRdRP in transgene-induced gene silencing in *Neurospora* is the conversion of transgenic RNA into dsRNA. Furthermore, over-expression of QDE-1 resulted in an increase in the production of siRNAs. In fact, high levels of QDE-1 allow the maintenance of silencing even when the number of transgenic copies are reduced by increasing the production of dsRNA and in turn siRNA molecules. Therefore, it was proposed that, in *Neurospora*, silencing activation and maintenance appear to rely on both the cellular amount of QDE-1 and the amount of transgenic copies producing RNA molecules that act as a substrate for the RdRP, implicating QDE-1 as a rate-limiting factor in PTGS (Forrest *et al.*, 2004).

Chapter 2

Φ6 RdRP – Experimental Procedures

The structure of apo Φ6 polymerase (Φ6pol) was reported by Butcher *et al.* (2001), who also determined the structures of several complexes with DNA oligonucleotides and NTPs, allowing a mechanism for the initiation of polymerisation to be proposed (section 1.4.2). However, several questions remained unanswered:

(i) what is the molecular basis of RNA specificity for Φ6pol?

(ii) what structural rearrangements occur once polymerisation occurs?

(iii) what is the molecular basis effects of different divalent cations?

(iv) what structural features are crucial to form the initiation complex?

(v) what is the importance of the initiation platform loop?

Strategies to try to fully understand the Φ6pol mechanism involved both structural and biochemical approaches, in a collaborative project with Prof. Dennis Bamford's group at the University of Helsinki. A number of atomic structures of Φ6pol complexed with RNA oligonucleotide template, several NTP substrate conditions and in the presence of different cations were determined. In addition, structures of mutant versions of the protein, where key residues were altered in order to understand their relevance, were also determined.

Structural studies of Φ6pol described in this thesis are grouped into different subcategories (Table 2.1), according to which of the above questions the experiment was designed to address:

(i) To understand the preference of Φ6 polymerase for RNA as a template, co-crystals of the protein with RNA oligos with 5, 6, and 7 nucleotides were obtained (5nt Φ6pol-RNA, 6nt Φ6pol-RNA and 7nt Φ6pol-RNA, respectively).

(ii) A Φ6pol-RNA co-crystal soaked with GTP and Mg^{2+} provided detailed structural information on the rearrangements observed when one round of polymerization occurs (Φ6pol-RNA-GTP-Mg^{2+}).

(iii) To determine the molecular basis of calcium inhibition of Φ6 polymerase reaction, a Φ6pol-RNA co-crystal was soaked with Ca^{2+} and GTP (Φ6pol-RNA-GTP-Ca^{2+}). To address the influence of manganese in the reaction, "manganese-free" Φ6pol models were obtained – in the presence of EDTA and with Mg^{2+} substituting Mn^{2+} (Φ6pol-EDTA and Φ6pol-Mg^{2+}, respectively).

(iv) To further understand the role of manganese, a mutant of Φ6pol was made, altering E491, which coordinates the Mn^{2+} ion, to glutamine. The apo structure of this mutant (E491Q) was determined, together with a set of complexes with templates and ligands: E491Q-DNA-GTP-Mg^{2+}, E491Q-DNA-GTP, E491Q-DNA-GTP-Mg^{2+}-Mn^{2+}, E491Q-RNA-Mn^{2+} and E491Q-RNA-Mn^{2+}-GTP-Mg^{2+}. The results shed light on the different steps leading to initiation.

(v) A mutant where the initiation platform loop 629QYWK632 is replaced with non-bulky residues SG was also produced and crystallised. The structure of this mutant (SG) gives insight into the importance of this loop.

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 polymerase mechanism is described in Chapter 3.

Table 2.1. Φ6 polymerase structures described in this thesis

Structure	Φ6pol	Nucleotide	Ion	NTP	PDB ac. code	
7nt Φ6pol-RNA	WT	5'-UUUUUCC-3'	Mn^{2+} (cryst.)	-	1UVJ	RNA specificity
6nt Φ6pol-RNA	WT	5'-UUUUCC-3'	Mn^{2+} (cryst.)	-	1UVI	RNA specificity
5nt Φ6pol-RNA	WT	5'-UUUCC-3'	Mn^{2+} (cryst.)	-	1UVM/ 1UVL	RNA specificity
Φ6pol-RNA-GTP-Mg^{2+}	WT	5'-UUUCC-3'	Mn^{2+} (cryst.) Mg^{2+} (soak)	GTP	1UVK	Reaction mechanism
Φ6pol-RNA-GTP-Ca^{2+}	WT	5'-UUUUCC-3'	Mn^{2+} (cryst.) Ca^{2+} (soak)	GTP	1UVN	Ca^{2+} effect
Φ6pol-EDTA	WT	-	-	-	n/a	Mn^{2+} effect
Φ6pol-Mg^{2+}	WT	-	Mg^{2+} (soak)	-	n/a	Mn^{2+} effect
E491Q	E491Q mutant	-	-	-	n/a	Mn^{2+} effect
E491Q-RNA-Mn^{2+}-GTP-Mg^{2+}	E491Q mutant	5'-UUUUCC-3'	Mn^{2+} (cryst.) Mg^{2+} (soak)	GTP	n/a	Mn^{2+} effect Reaction mechanism
E491Q-DNA-GTP-Mg^{2+}	E491Q mutant	5'-TTTTCC-3'	Mg^{2+} (soak)	GTP	n/a	Mn^{2+} effect Reaction mechanism
E491Q-DNA-GTP	E491Q mutant	5'-TTTTCC-3'	-	GTP	n/a	Mn^{2+} effect Reaction mechanism
E491Q-DNA-GTP-Mg^{2+}-Mn^{2+}	E491Q mutant	5'-TTTTCC-3'	Mg^{2+} (soak) Mn^{2+} (soak)	GTP	n/a	Mn^{2+} effect Reaction mechanism
E491Q-RNA-Mn^{2+}	E491Q mutant	5'-UUUUCC-3'	Mn^{2+} (cryst.)	-	n/a	Mn^{2+} effect Reaction mechanism
SG	SG mutant	-	-	-	1WAC	Initiation platform

2.1. Protein expression and purification

Φ6 polymerase expression and purification were usually carried out by Minni R. L. Koivunen, as part of the collaboration with Prof. Dennis Bamford at the University of Helsinki, whilst structural studies were performed in Oxford as my contribution to this project. To strengthen my understanding of the procedures involved in the expression and purification of Φ6pol, I was directly involved in producing one batch of the SG mutant pure protein when visiting the University of Helsinki. Procedures for wild type (WT) Φ6 polymerase protein expression and purification were described by Makeyev and Bamford (2000a). Slightly modified protocols were carried out for the SG mutant and are described in detail here. For completeness, a short description of the plasmids used for the expression of the WT, E491Q and SG mutant Φ6 polymerase is also included.

For expression of the wild type Φ6 polymerase, plasmid pEMG2 was used (Makeyev, 2001). This plasmid was used to construct the SG mutant plasmid by PCR amplification of the small fragment of Φ6 polymerase gene using oligonucleotides 5'-CCAGTTCAGCCCTGAGTACGGTGT-3' and 5'-GCCATGCATCAGTACCTCGTGGATATTCGCCGAGACATCGGCCTCGGTAC CGGAGAGTTTGTT-3' as upstream and downstream primers, respectively. The PCR product was digested with *Nru*I-*Nsi*I and ligated with similarly cut pEMG2. The insert in the resultant plasmid pRT2 [629QYKW632 - SG mutant] was verified by sequencing. For expression of E491Q Φ6 polymerase mutant, a point mutation was introduced into plasmid pEM33 encoding for WT Φ6pol with a C-terminal hexahistidine tag tethered to the protein with an AALE linker (Makeyev and

Bamford, 2000a). The QuickChange mutagenesis kit (Stratagene) and 5'-CACCAAGGAAGGCGCCACCGTGCTGGTAGGAGATCTTCATGTAAGG-3' and 5'-CCTTACATGAAGATCTCCTACCAGCACGGTGGCGCCTTCCTTGGTG-3' oligonucleotides as upstream and downstream primers, respectively, were used to introduce the mutation, creating plasmid pNL18. To construct a plasmid for the expression of E491Q polymerase without histidine tag (His-tag), pNL18 was cut with *Nde*I and *Nsi*I and ligated with similarly cut pEMG2 (expression vector for WT Φ6pol gene). The resultant plasmid, pSVe4, was partially sequenced to verify the mutation.

Escherichia coli DH5α (Gibco-BRL) cells were used for plating and production of plasmids. Competent cells were transformed by adding of 100μl of plasmid DNA to 200μl of competent cells DH5α and incubated on ice for 30-40min, followed by a heat shock treatment for 30s at 42°C. The transformed cells were incubated for 30min at 37°C with shaking, after addition of 1ml of LB medium. 100μl of the cell culture was then plated on to LB medium plates containing ampicillin and left to grow overnight at 37°C. Pure plasmid DNA was obtained using the QIAprep Miniprep (Qiagen), according to manufacturer's instructions. Briefly, overnight pelleted bacterial cells are ressuspended in a set of buffers (alkaline lysis) before centrifugation for 10min at 14,000g. The supernatant is then applied into a QIAprep column and centrifuged for 30-60s at 14,000g. The flow-through is discarded and the column washed with a buffer containing 70% ethanol (PE buffer, Qiagen) before eluting the DNA with 10mM TrisHCl (pH8.5) by centrifugation for 1min (14,000g). To verify that the cells had been transformed with the correct plasmid, restriction enzyme reactions were carried out and analysed in 0.8% agarose

gels. As an example, a description of procedures for restriction enzyme analysis of an His-tag containing plasmid derived from pEM33 is given. pNL9 plasmid has a point mutation in the Φ6pol gene (K541L) that eliminates a *Nru*I digestion sequence, preserving the *Nde*I digestion site (Fig. 2.1.A). Therefore, when the plasmid is digested with both enzymes, linear plasmids 7372bp long should result, whilst false positives would include two fragments (1612bp and 5760bp). 2µl DNA plasmid were incubated for 1h at 37°C with 2µl *Nru*I (10 units/ µl), 2µl *Nde*I (10 units/ µl) and 2µl *Nru*I (10 units/ µl)+ 2µl *Nde*I (10 units/ µl) in 200mM Tris-HCl pH 7.5, 70mM $MgCl_2$, 100mM KCl, 20mM 2-mercaptoethanol buffer and loaded to the agarose gel, together with undigested pNL9, DNA size markers and λ DNA cut with *Pst*I as digestion marker. Undigested DNA and plasmid incubated with N*ru*I resulted in two bands of ~1600 and 5700bp, whilst double digestion and N*de*I result in a single band with ~7400bp (Fig. 2.1.B). This indicates that the plasmid was uncut by *Nru*I but digested by *Nde*I, as expected.

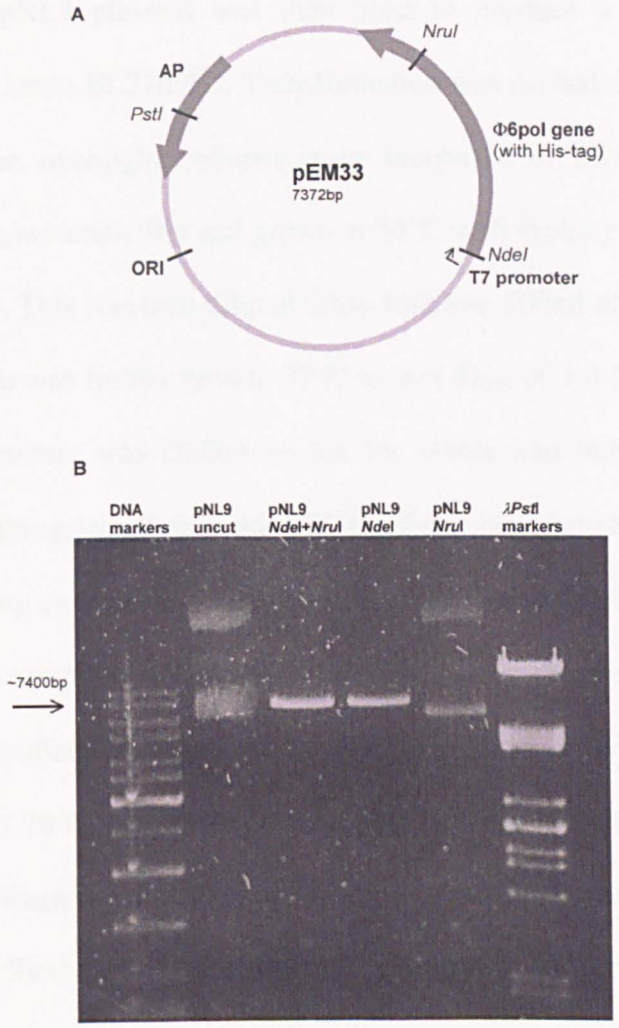


Figure 2.1. Example of an expression plasmid production of Φ6pol with a His-tag and a restriction enzyme analysis of a similar plasmid

A. A schematic representation of Φ6pol expression plasmid (pEM33) presenting the location of a Φ6pol gene and the recognition sites for the restriction enzymes used here (Makeyev and Bamford, 2000a)

B. Agarose gel analysis of the digestion of plasmid pNL9 (based on pEM33) either by *NdeI*, *NruI* or a mixture of these restriction enzymes.

Purified pNL9 plasmid was then used to produce a starter culture of expression *E. coli* strain BL21/DE3. Transformation was carried out as described for DH5α cells. Three overnight colonies were incubated in 20ml of LB medium containing 100 mg/ml ampicillin and grown at 37°C with shaking (250rpm) until the OD_{540} reached 0.5. This was then diluted thirty-fold into 500ml of the same medium. The diluted culture was further grown, 37°C to an OD_{600} of 1.0 (~2h), with shaking at 250rpm. The culture was chilled on ice for 10min and induced with 100mM isopropyl-beta-D-thiogalactopyranoside (IPTG). Each step in protein production was analysed by running each protein sample on a 12.5% SDS-PAGE gel. Prior to IPTG induction a 200µl sample was collected, centrifuged at 13,000rpm and ressuspended in 20µl of 1xSB (sodium boric acid electrophoresis buffer). IPTG-induced cells were then transferred to 20°C where they were shaken for 18h at 240rpm. Bacteria were collected by centrifugation at 7,000rpm for 10min at 4°C and ressuspended in 6.3ml of 100mM NaCl, 50mM Tris-HCl (pH 8.0), 1mM EDTA. All purification steps were carried out at 4°C. The suspension was passed three times at ~105MPa through a pre-cooled French pressure cell. Phenylmethylsulfonyl fluoride was added to a final concentration of 1mM after the first passage. The lysate was centrifuged at 120,000 g for 2.5h. The supernatant fraction was loaded onto a dye affinity column (Cibacron Blue 3GA; Sigma). Proteins bound to the column were eluted with 500mM NaCl, 50mM Tris-HCl (pH 8.0), 1mM EDTA (Fig. 2.2.).

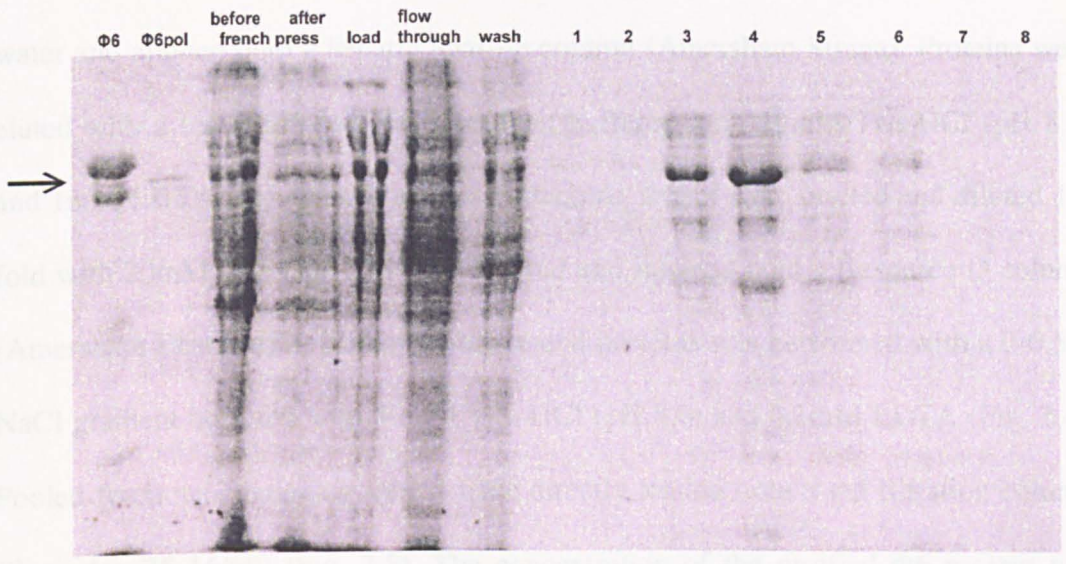


Figure 2.2. Analysis of Φ6pol purification products after blue agarose affinity column

SDS-PAGE gel analysis of samples after the blue agarose gel (Sigma) column. Purified Φ6 genome and Φ6pol were used as molecular size markers. Samples at each step of purification were analysed: before and after French press cell disruption, after centrifugation at 10,000rpm for 1h (load), flow through after loading (flow through) and washing (wash) steps, and eluted fractions 1 to 8. Migration position of Φ6pol is marked on the left (black arrow).

Pooled fractions containing Φ6pol were diluted 5-fold with ice-cold distilled water and applied onto a heparin agarose column (Amersham Sigma). Proteins were eluted with a linear 0.1-1M NaCl gradient buffered with 50mM Tris-HCl (pH 8.0) and 1mM EDTA (Fig. 2.3). Fractions containing Φ6pol were pooled and diluted 10-fold with 20mM Tris-HCl (pH 8.0), filtered and injected onto a Resource Q column (Amersham Pharmacia). Elution of the bound proteins was performed with a 0-0.5M NaCl gradient buffered with 50mM Tris-HCl (pH 8.0) and 0.1mM EDTA (Fig. 2.4). Pooled fractions containing Φ6pol were directly loaded onto a gel filtration column (Superdex 75 16/60) (Fig. 2.5). The concentration of the purified Φ6 protein was determined by the absorbance at 280nm (ε_{280}=1.395 $M^{-1}cm^{-1}$).

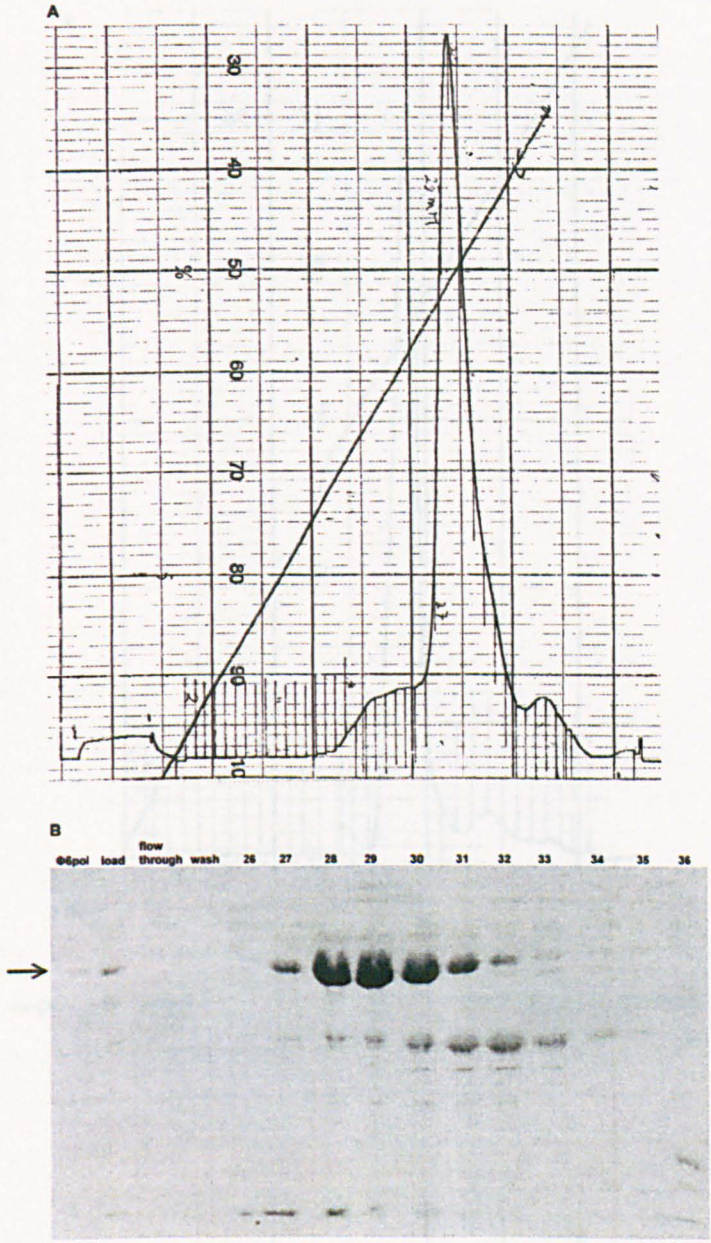


Figure 2.3. Analysis of Φ6pol purification products after heparin column

A. The elution pattern (A_{280}) from the heparin column, with a peak corresponding to the Φ6pol containing fractions. Elution gradient is indicated by diagonal line.

B. SDS-PAGE gel analysis of samples eluted form the heparin column (Hitrap, Amersham Biosciences). Purified Φ6pol was used as molecular size marker. Loaded sample from the blue agarose column (load), flow through of the loading (flow through) and washing (wash) steps and eluted fractions 26-36 were analysed. The molecular size of Φ6pol is marked by a black arrow on the left.

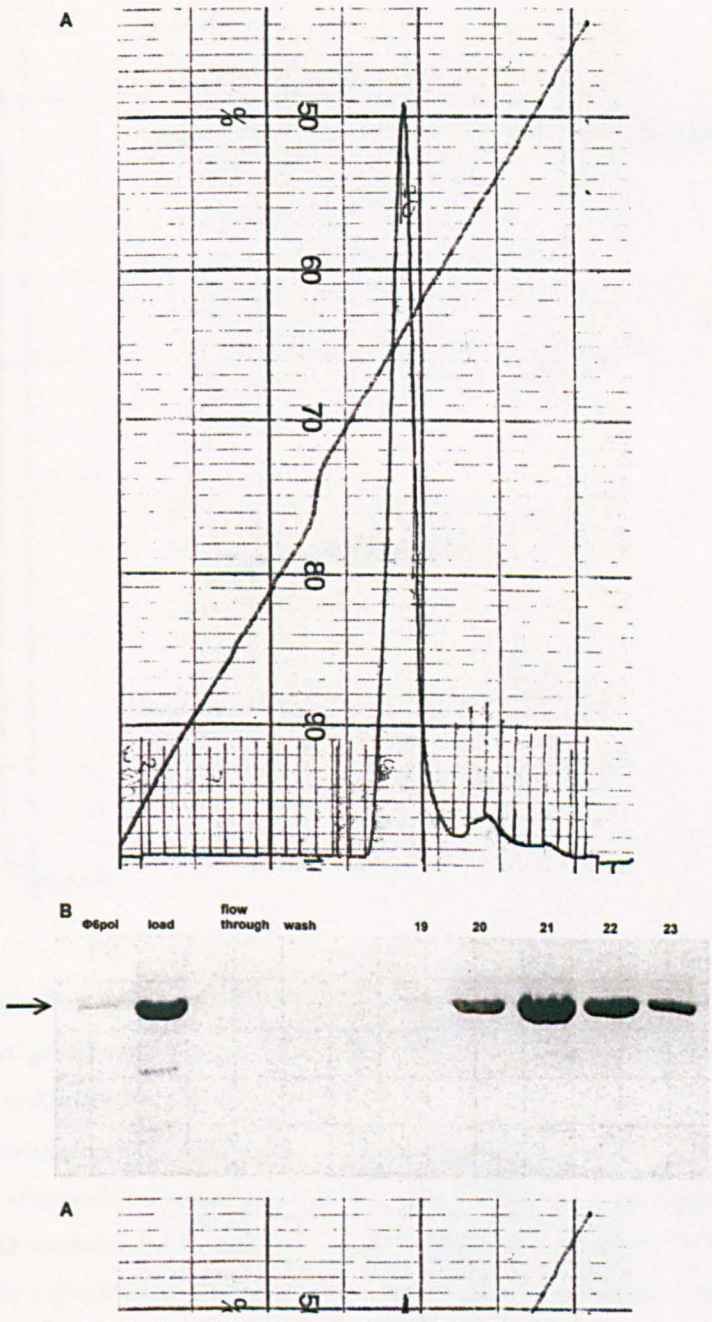


Figure 2.4. Analysis of Φ6pol purification products after Q column

A. Q column elution pattern (A_{280}), with a peak corresponding to fractions containing Φ6pol. Elution gradient is indicated by diagonal line.

B. SDS-PAGE (containing 12% acrylamide) of samples after the Q column. Purified Φ6pol was used as molecular size marker. Loaded sample from the heparin column (load), flow through of the loading (flow through) and washing (wash) steps and eluted fractions 19-23 were analysed. The molecular size of Φ6pol is marked by a black arrow on the left.

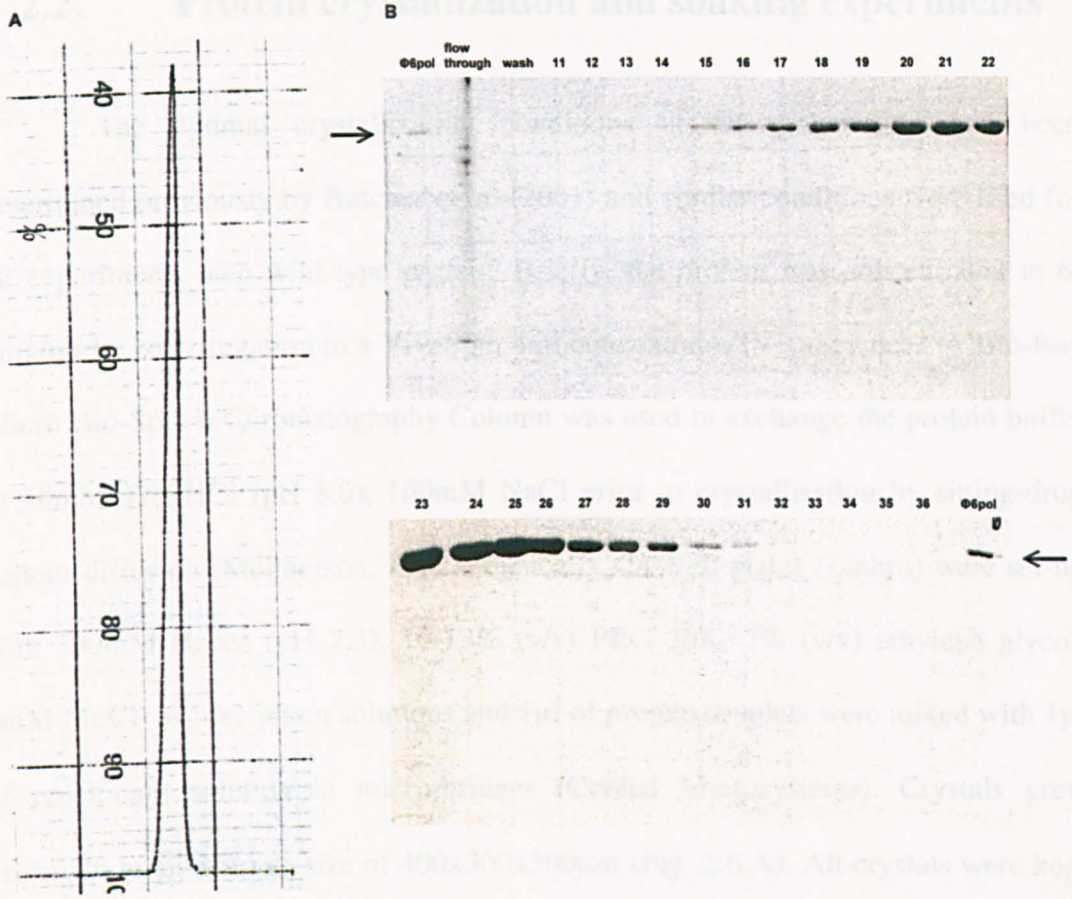


Figure 2.5. Analysis of Φ6pol purification products after gel filtration

A. A_{280} curve from gel filtration column (Superdex75 16/60), showing a sharp peak corresponding to the Φ6pol pooled containing fractions.

B. SDS-PAGE (containing 12% acrylamide) of samples after the gel filtration. Purified Φ6pol was used as molecular size marker. Loaded sample from the Q column (load), flow through of the loading (flow through) and washing (wash) steps and fractions 11-36 were analysed. The molecular size of Φ6pol is marked by a black arrow. Top right bottom panel: load, flow through, wash and fractions 11 to 22; bottom right panel: fractions 23-36 (plus purified Φ6pol as a marker)

2.2. Protein crystallization and soaking experiments

The optimal crystallization conditions of Φ6 polymerase had been determined previously by Butcher *et al.* (2001) and similar conditions were used for all experiments with wild type protein. Briefly, the protein was concentrated to 6-8mg/ml by centrifugation in a VivaSpin 4ml concentrator (VivaScience). A Bio-Rad Micro Bio-Spin 6 Chromatography Column was used to exchange the protein buffer to 10mM Tris-HCl (pH 8.0), 100mM NaCl prior to crystallisation by sitting-drop vapour diffusion (McPherson, 1982). Typically, 24-well plates (Linbro) were set-up with 100mM Hepes (pH 7.3), 10-13% (v/v) PEG 20K, 2% (v/v) ethylene glycol, 2mM $MnCl_2$ as precipitant solutions and 1μl of protein droplets were mixed with 1μl of precipitant solution in micro-bridges (Crystal Microsystems). Crystals grew overnight to an average size of 400x300x300μm (Fig. 2.6.A). All crystals were kept at room temperature until data collection.

2.2.1. Φ6pol-RNA co-crystallisation

Co-crystallisation experiments were performed on Φ6pol by incubation with RNA molecules consisting of 5, 6 or 7 nucleotides (5'-UUUCC-3', 5'-UUUUCC-3' and 5'-UUUUUCC-3', respectively). The sequence of these molecules mimics the conserved 3' end (5'-...CC-3') of the negative sense strand of the S and M segments of the Φ6 genome, the preferred templates for transcription by Φ6pol (McGraw *et al.*, 1986; Gottlieb *et al.*, 1988; Yang *et al.*, 2001). Oligonucleotides with the chosen sequence were supplied by Eurogentec (EGT Group) and stored at a controlled

temperature of -20°C until co-crystallization experiments were set up. Typically, 6µl of a Φ6pol solution at a concentration of 6-8mg/ml (80-107µM) was incubated with 0.2, 0.3, 0.4 and 0.6 µl (320, 475, 625 and 910 µM, respectively) of a solution of oligonucleotides at a concentration of 10mM, for one hour on ice. An excess of RNA (from 1:4 up to 1:11 RNA/protein) was used to ensure binding of the oligos. Sitting drop crystallization trays (Linbro) in the described conditions were then prepared.

2.2.2. Φ6pol-RNA-GTP-Mg^{2+} complex

To study further the reaction mechanism of the polymerase, co-crystals of WT protein and a 5nt RNA oligonucleotide (5'-UUUUCC-3') were used for soaking experiments with GTP and Mg^{2+}. A concentration of 125mM $MgCl_2$ was added to the crystallization drop, followed shortly after by 100mM GTP (lithium salt). The total soaking time was around 2min (at which point cracking of the crystal become apparent).

2.2.3. Φ6pol-RNA-GTP-Ca^{2+} complex

To study the effect of calcium on the polymerase mechanism, co-crystals of WT protein and a 6nt RNA oligonucleotide were used for soaking experiments with GTP and Ca^{2+}. $CaCl_2$ at up to 125mM and, shortly after, 100mM GTP (lithium salt) were added to the crystallization drop and left to diffuse through the crystal for up to 2min or until the crystal started to show cracks on its surface when observed under the microscope.

2.2.4. Φ6pol-EDTA and Φ6pol-Mg^{2+}

In order to obtain crystals of "manganese-free" polymerase, the protein was incubated with 50mM of the chelating agent EDTA for one hour, prior to crystallization in the same conditions as the native crystallization experiments where 2mM $MnCl_2$ are added to the mother liquor. With the aim of obtaining crystals of Φ6pol with unambiguously placed magnesium ions, crystals grown in the presence of 50mM EDTA were soaked with $MgCl_2$ at a concentration of 100mM in order to completely wash out the chelating agent and allow the Mg^{2+} to penetrate the crystals for around 2-3min, until cracking of the crystal became apparent when observed under a microscope.

2.2.5. E491Q mutant crystallization

To determine the optimal crystallization conditions for the E491Q mutant protein, a screen of 480 conditions was carried out using the sitting drop vapour diffusion method with a 100:100nl drop size (1:1 protein/precipitant ratio) by a Cartesian Robot available at the Oxford Protein Production Facility (Brown *et al.*, 2003; Walter *et al.*, 2003). The protein was typically concentrated to 6-8mg/ml in 10mM Tris-HCl (pH 8.0) and 100mM NaCl in a procedure similar to that described for wild type protein. Optimised conditions were found using the automated procedure established by Walter *et al.* (2005). Once optimal conditions had been established, crystallisation volumes were scaled up to 2μl - 3μl drops (1:1, 1:2 or 2:1 protein/precipitant ratio) in 24-well plates, resulting in crystals of 200x110x80μm in

15-20% PEG 4K, 8.5% isopropanol, 15% glycerol and 100mM Hepes (pH 7.5) (Fig. 2.6.B).

2.2.6. E491Q co-crystallizations with RNA and DNA oligos

Φ6pol E491Q mutant was co-crystallized with a 6nt RNA oligonucleotide (5'-UUUUCC-3') and equivalent DNA oligonucleotide (5'-TTTTCC-3'), chosen with the same rational as the oligos used for wild type co-crystallisation and supplied by Eurogentec (EGT Group). As for wild type experiments, 6µl of a protein solution at a concentration of 6-8mg/ml was incubated with 2 to 8 µM of the oligonucleotides for one hour on ice, before setting up sitting drop crystallization trays in the conditions pre-established for the mutant protein crystals. For both oligonucleotides, crystallisation were carried out in the presence and absence of 2mM $MnCl_2$. For the RNA co-crystallization experiments, 2mM $MnCl_2$ was necessary, whilst in successful DNA co-crystallization experiments, Mn^{2+} ions were absent from the protein buffer and the crystallization solution.

2.2.7. E491Q RNA/DNA co-crystals soaking experiments

E491Q-oligonucleotide co-crystals were used for soaking experiments with GTP and Mg^{2+} and/or Mn^{2+}. A concentration of 5mM $MgCl_2$ and, shortly after, 25mM GTP (lithium salt) was added to the crystallization drop containing E491Q-RNA co-crystals grown in the presence of 2mM $MnCl_2$, and left to diffuse through

the crystal. In a similar experiment, 25mM GTP was added to the crystallization droplet containing E491Q-DNA co-crystals. In a separate experiment, 5mM $MgCl_2$ was added to the crystallization droplet containing E491Q-DNA co-crystals, shortly followed by the addition of 25mM GTP. Finally, since these E491Q-DNA co-crystals were grown in the absence of Mn^{2+}, 5mM $MnCl_2$ was added to the crystallization drop, at a concentration of 5mM, prior to the addition of 5mM $MgCl_2$ followed by 25mM GTP. All soaks were carried out for between 30 and 120s, until cracking of the crystal became apparent.

2.2.8. SG mutant

For structural studies, the SG mutant protein was buffered using a solution of 10mM Tris-HCl (pH 8.0), 100mM NaCl and concentrated to 4mg/ml, using the same procedure as described for WT. Initially, a screen of 480 conditions was carried out using the sitting drop vapour diffusion method with a 200nl drop size (100nl protein plus 100nl precipitant ratio) (Brown *et al.*, 2003; Walter *et al.*, 2003). Optimal crystal growth conditions - 100mM sodium citrate pH 5.6, 19% (v/v) isopropanol, 19% (v/v) PEG 4000, 5% (v/v) glycerol - were scaled up to 24-well plate set-up, with 2μl droplet size (1:1 protein/precipitant solution). These conditions gave very small crystals within a few hours, that grew to a size of 300x200x100μm after four months (Fig. 2.6.C). Co-crystallisation using 6nt RNA and DNA oligos described were unsuccessful. Furthermore, all attempts to soak SG mutant crystals in solutions containing GTP, non-hydrolysable analogues of GTP or ATP, with similar procedures used for WT experiments, were also unsuccessful.

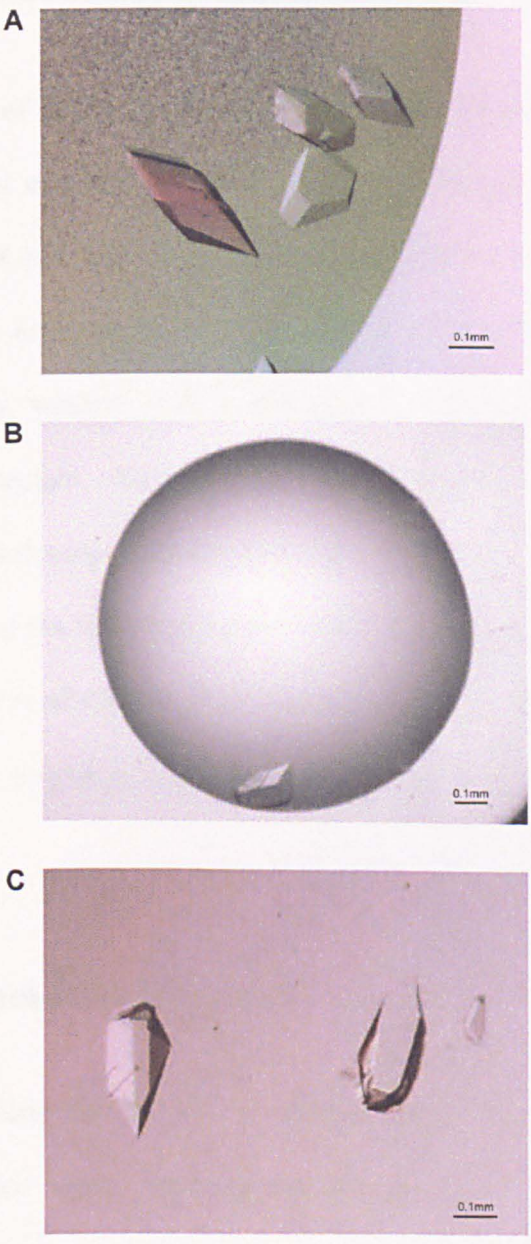


Figure 2.6. Φ6pol crystals

A. Φ6pol wild type crystals, grown over night in 100mM Hepes (pH 7.3), 10-13% (v/v) PEG 20K, 2% (v/v) ethylene glycol and 2mM $MnCl_2$.

B. E491Q Φ6pol crystal grown to full size within 35h in 100mM Hepes (pH 7.5), 17% PEG 4K, 8.5% isopropanol and 15% glycerol.

C. SG Φ6pol crystals grown after four months in 100mM sodium citrate pH 5.6, 19% (v/v) PEG 4000, 19% (v/v) isopropanol, and 5% (v/v) glycerol

2.3. Data collection and processing

A summary of data collection and processing statistics for data sets used to determine each Φ6pol structure described in this thesis is given in Table 2.2. Unless otherwise stated, all X-ray diffraction data sets were collected from crystals cooled in a stream of gaseous nitrogen at a temperature of 100K. Prior to data collection, crystals were briefly washed with a precipitant solution containing 25% (v/v) glycerol as cryoprotectant. All data sets were integrated and scaled using the HKL2000 suite of software (Otwinowski and Minor, 1997). All crystals of the different complexes of the wild type protein were isomorphous with those previously described by Butcher *et al.* (2000, 2001), belonging to space group $P2_1$ with unit cell dimensions: a≈105Å, b≈93Å, c≈140Å, α=γ=90°, β≈101°.

2.3.1. Φ6pol-RNA complexes

Diffraction data for the WT polymerase co-crystallised with 6nt and 7nt RNA oligonucleotides were collected on station PX 14.2 at the Synchrotron Radiation Source (SRS), Daresbury, UK, using an ADSC Q4R CCD detector, to resolutions of 2.2 and 1.9 Å, respectively. X-ray diffraction data from the 5nt Φ6pol-RNA co-crystal were collected to a resolution of 2.0Å, at station ID29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, using an ADSC Q210 CCD detector. For all three data sets, the wavelength of the X-ray radiation was 0.98Å, with 1° oscillation between frames over a continuous rotation range. Exposure times of 10 seconds per frame were used for the 5nt Φ6pol-RNA co-

crystal. Exposure times of 15 and 30s were used for the 7nt and 6nt Φ6pol-RNA co-crystals, respectively.

2.3.2. Φ6pol-RNA-GTP-Mg^{2+} complex

Data from WT Φ6pol-RNA co-crystals soaked with GTP and $MgCl_2$ to a resolution of 2.5Å were collected in-house, using a MAResearch 345 image plate detector, with a wavelength of 1.54Å, as oscillations of 1° and with exposure times between 10 and 15 minutes.

2.3.3. Φ6pol-RNA-GTP-Ca^{2+} complex

X-ray diffraction data from Φ6pol-RNA-Ca^{2+}-GTP crystals were collected on station BM14 at the ESRF, using a MAResearch 225 CCD detectors, with the wavelength of the X-ray radiation set to the *K* edge of manganese, 6.547 KeV (1.89 Å). Each of the 360 images collected covered an oscillation of 1°, exposure times were 10 seconds and the resolution at the edge of the detector was 3.0Å.

2.3.4. Φ6pol-EDTA and Φ6pol-Mg^{2+} complex

Data from Φ6pol crystals grown in the presence of EDTA were collected on station BM14, at the ESRF, using a MAResearch 165 CCD detector, to a resolution of 2.5Å. The wavelength of the X-ray beam was tuned to the *K* edge of manganese,

6.547 KeV (1.89Å). Diffraction data for the Φ6pol-Mg^{2+} complex were collected to 2.5Å resolution on station ID14EH2 at the ESRF, using an ADSC Q4 CCD detector. For both data sets, each image covered an oscillation of 1° and exposure times were typically 5-10 seconds.

2.3.5. E491Q mutant

X-ray data from E491Q Φ6pol mutant apo crystals and all co-crystallisation and soaking experiments were collected on station PX 9.6 at the SRS, using an ADSC Q4R CCD detector. Each image covered an oscillation of 1° and exposure times were typically 45-60 seconds per frame.

For Φ6pol E491Q mutant crystals, diffracting to 3.2Å resolution, 180 frames were collected. These crystals were approximately isomorphous to the WT crystals previously described (Butcher *et al.*, 2000; Butcher *et al.*, 2001) belonging to space group $P2_1$, with unit cell dimensions a≈106Å, b≈92Å, c≈142Å, β≈102°.

2.3.6. E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} complex

E491Q co-crystals with RNA and Mn^{2+}, soaked with Mg^{2+} and GTP, diffracted very poorly to 3.8Å and were approximately isomorphous to crystals of the SG initiation platform mutant described below and belonging to space group $P2_1$ with unit cell dimensions a=78Å, b=107Å, c=158Å, β=99°.

2.3.7. E491Q-DNA-GTP-Mg^{2+}, E491Q-DNA-GTP, E491Q-DNA-GTP-Mg^{2+}-Mn^{2+} and E491Q-RNA-Mn^{2+} complexes

The E491Q-DNA-GTP-Mg^{2+}, E491Q-DNA-GTP, E491-DNA-Mg^{2+}-Mn^{2+} and E491Q-RNA-Mn^{2+} datasets, collected on station PX 9.6 at the SRS, using an ADSC Q4R CCD detector as oscillations of 1° and with exposure times typically 45-60 seconds per frame, belonged to space group $P3_2$, with similar unit cell dimensions: a≈b≈109Å, c≈159Å. These crystals are approximately isomorphous to the $P3_2$ SeMet crystal structure determined by Butcher *et al.* (2001). For this space group there are four possible ways to index the data. To facilitate later comparisons, all datasets were reindexed with SCALEPACK (Otwinowski and Minor, 1997) to be consistent with the SeMet WT model (Butcher *et al.*, 2001). The best data set for this space group was collected from a crystal of the E491Q-DNA-GTP-Mg^{2+} complex to a resolution of 2.4Å. The E491Q-DNA-GTP data set was collected to a resolution of 2.6Å, whilst the E491Q-DNA-GTP-Mg^{2+}-Mn^{2+} complex diffracted to 2.8Å resolution. Mutant co-crystals with RNA in the presence of Mn^{2+} diffracted to 3.2Å.

2.3.8. SG mutant

Φ6pol SG mutant X-ray diffraction data to 3.0Å were collected at ID29, at the ERSF, using an ADSC Q210 detector, with 1° oscillation images over a continuous rotation range. Data processing revealed that the SG mutant crystals are not isomorphous with WT crystals described, belonging to space group $P2_1$ with unit cell dimensions of a=76.6Å, b=105.9Å, c=157.7Å, β=98.8°.

Table 2.2. Data collection and processing statistics for Φ6pol data sets

Data set	Wavl. (Å)	N. img.	Sp. Gr.	Unit cell a,b,c (Å); α=γ=90° β (°)	Resol. Range (Å)	Observ.	Unique reflect.	Cmplt. (%)	I/δ(I)	R_{merge} [a] (%)
7nt Φ6pol-RNA	0.98	359	$P2_1$	105.1, 93.5, 104.8; 101.2	30.0–1.9 (2.0–1.9)	1733793	209896	99.7 (99.8)	13.5 (1.2)	8.5
6nt Φ6pol-RNA	0.98	239	$P2_1$	105.1, 93.7, 140.7; 101.2	20.0–2.2 (2.3–2.2)	1522219	142686	97.8 (83.7)	13.3 (1.6)	12.6
5nt Φ6pol-RNA	0.98	297	$P2_1$	105.9, 91.9, 140.8; 101.6	50.0–2.0 (2.1–2.0)	2110477	176955	99.2 (93.5)	20.9 (1.8)	6.6
Φ6pol-RNA-GTP-Mg^{2+}	1.54	172	$P2_1$	106.3, 93.8, 140.8; 101.4	20.0-2.5 (2.6–2.5)	4701397	132353	97.7 (96.0)	25.3 (3.4)	13.6
Φ6pol-RNA-GTP-Ca^{2+}	1.89	360	$P2_1$	105.5, 92.1, 140.8; 101.0	50.0–3.0 (3.1-3.0)	880290	98376	96.9 (78.4)	11.9 (1.6)	15.1
Φ6pol-EDTA	1.89	280	$P2_1$	105.2, 94.0, 141.0; 101.2	50.0-2.5 (2.6-2.5)	1276051	97851	94.7 (72.8)	7.0 (1.1)	16.5
Φ6pol-Mg^{2+}	0.93	280	$P2_1$	105.7, 91.0, 140.9; 101.2	50.0-2.5 (2.6-2.5)	3442868	88823	99.2 (100)	17.7 (3.5)	9.1
E491Q	0.98	180	$P2_1$	106.5, 91.6, 142.7; 101.6	20.0-3.2 (3.3-3.2)	458511	40323	90.5 (80.8)	7.2 (1.6)	17.9
E491Q-RNA-Mn^{2+}-GTP-Mg^{2+}	0.98	180	$P2_1$	78.1, 107.0, 158.4; 98.9	20.0-3.8 (3.9-3.8)	359558	25268	99.5 (99.4)	6.6 (3.3)	21.8
E491Q-DNA-GTP-Mg^{2+}	0.98	180	$P3_2$	108.1, 108,1, 158.4; 120.0	30.0-2.4 (2.5-2.4)	1187116	105101	100.0 (100.0)	6.8 (1.7)	17.6
E491Q-DNA-GTP	0.98	180	$P3_2$	109.5, 109.5, 159.0; 120.0	30.0-2.6 (2.7-2.6)	951362	107922	100.0 (99.5)	6.3 (1.1)	18.8
E491Q-DNA-GTP-Mg^{2+}-Mn^{2+}	0.98	130	$P3_2$	109.0, 109.0, 158.8; 120.0	20.0-2.8 (2.9-2.8)	610471	72626	99.8 (99.9)	6.2 (1.4)	22.3
E491Q-RNA-Mn^{2+}	0.98	180	$P3_2$	110.0, 110.0, 159.1; 120.0	18.0-3.2 (3.3-3.2)	668591	52552	100.0 (100.0)	6.7 (1.6)	28.9
SG	0.98	185	$P2_1$	76.6, 105.9, 157.7; 98.8	30.0–3.0 (3.1–3.0)	620947	57385	99.2 (98.6)	7.4 (1.9)	24.2

Values in parenthesis are for the outermost resolution shell

[a] $R_{merge} = \Sigma_i\Sigma_h(|I_{j,h} - \langle I_h \rangle|) / \Sigma_j\Sigma_h(\langle I_h \rangle)$, where h are unique reflections indices, $I_{j,h}$ are intensities of symmetry-related reflections and $\langle I_h \rangle$ is the mean intensity

2.4. Structure determination and refinement

Φ6pol and derivatives, including different complexes and mutated variant forms, crystallise in three space groups:

A - $P2_1$ with unit cell dimensions a≈105Å, b≈93Å, c≈140Å, β≈101°

B - $P2_1$ with unit cell dimensions a≈77Å, b≈107Å, c≈158Å, β≈99°

C - $P3_2$ with unit cell dimensions a≈b≈109Å, c≈159Å.

Crystals of WT Φ6pol and apo E491Q mutant described in this chapter belong to A crystal form and were isomorphous with crystals of structures determined previously by Butcher *et al.* (2001). Rigid body refinement of monomers in the asymmetric unit was sufficient to obtain initial phases for these structures.

The structures of crystals belonging to crystal forms B and C, different from the $P2_1$ A crystal form were solved by molecular replacement, as described bellow.

Final refinement statistics for all structural models described in this chapter are summarised in Table 2.3. Unless otherwise stated, manual model building and maps inspection were carried out using O (Jones *et al.*, 1991).

2.4.1. 6nt and 7nt Φ6pol-RNA complexes

For the Φ6pol-RNA complexes, initial models were obtained by rigid body refinement of the apo WT model with the package CNS (Brunger *et al.*, 1998).

Visual inspection of the resulting SIGMAA |Fo|-|Fc| maps allowed identification of the presence of RNA oligonucleotides in the template tunnel. Since there are three molecules in the asymmetric unit related by non-crystallographic symmetry, the electron density maps were three-fold averaged using the program GAP (DIS, JMG, JMD, unpublished) in order to improve the signal-to-noise ratio. RNA oligonucleotides models were manually fitted in the electron density maps in O for all three complexes. The structures were then refined using CNS (Brunger *et al.*, 1998) with positional and individual B-factor refinement and bulk solvent correction, imposing strict three-fold non-crystallographic symmetry constraints. For the 6nt and 7nt RNA oligos, the data and calculated electron density maps were of high quality, so final refinement to a resolution of 2.2Å and 1.9Å (respectively) was carried out using non-crystallographic symmetry restraints for the different protein sub-domains [palm, fingers, thumb and C-terminal, as defined by Butcher et al. (2001)] in each molecule, resulting in models with R_{factor} of 23.3% and 24.3%, respectively (R_{free}=25.9% and 27.1%) and good stereochemistry (rmsd bond length of 0.015 and 0.012 Å; rmsd bond angle of 1.6 and 1.4°, respectively). The protein root-mean square deviation between equivalent atoms of the three protein molecules in the asymmetric unit was 0.12 $Å^2$ and 0.14 $Å^2$, respectively for the 6nt and 7nt models. Differences in the strength of the electron density of SIGMAA 2|Fo|-|Fc| maps in the template tunnel between the three molecules in the asymmetric unit for these complexes suggested that the RNA oligos in each molecule had different occupancies. To check this possibility, different occupancies were assigned to the 6nt and 7nt RNA models in each molecule (0.2 for molecule "A", 0.4 for molecule "B" and 0.7 for molecule "C") and refined in a final cycle of positional and individual B-

factor refinement in CNS (Brunger *et al.*, 1998). The resulting B-factors of the RNA templates were reasonably similar indicating that the apparent differences in occupancies were real. This can be explained by the modulation of the RNA entry by the varied contacts with surrounding molecules, restrained by crystallographic contacts. The entrance of the RNA tunnel of molecule "C" is less obscured by neighbouring molecules than the template tunnels of molecules "A" and "B" (Fig. 2.7), allowing easier access of RNA oligonucleotides. Therefore, the average occupancy of the RNA molecule in the template tunnel in molecule "C" is higher than for molecule "A" or "B". Due to local disorder of the more external nucleotides in the 6nt and 7nt oligonucleotide structures, only 4 nucleotides were fitted into the difference-Fourier maps and refined.

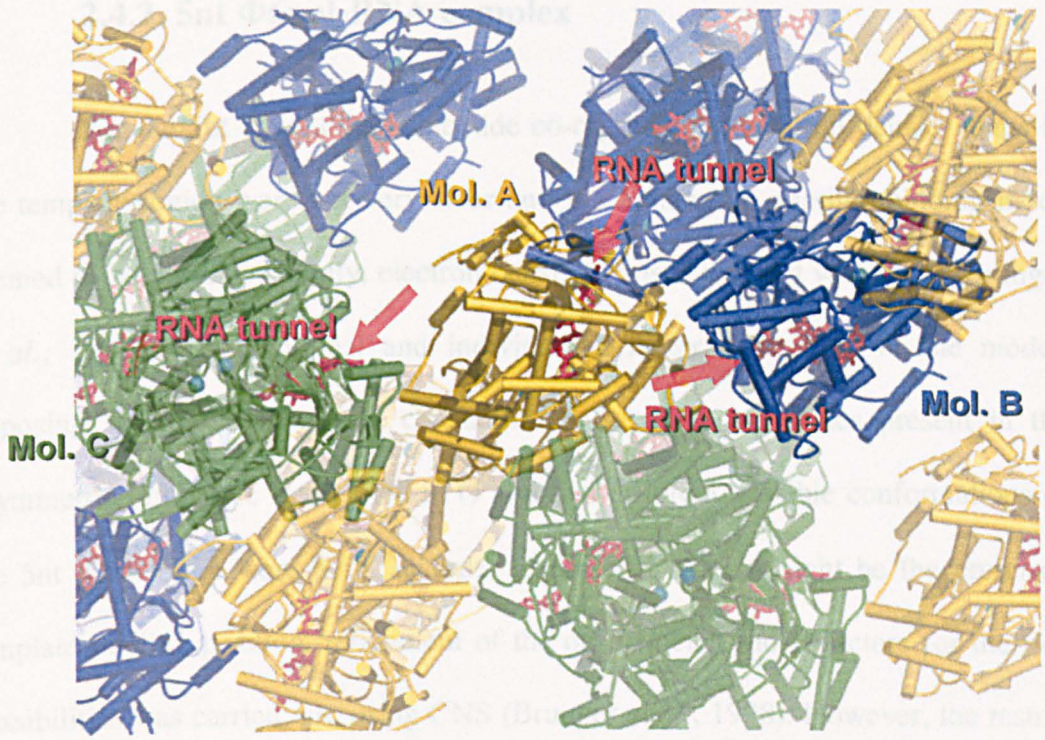


Figure 2.7. Φ6pol-RNA $P2_1$ crystallographic packing

Cartoon representation of the crystallographic packing of the Φ6pol co-crystallised with a 6nt RNA oligonucleotide (ball and stick representation, red). The three molecules present in the asymmetric unit are represented in different colours: molecule "A" – yellow; molecule "B" – blue; molecule "C" – green. Crystallographically related molecules are shown in transparent mode, with an equivalent colour code. Manganese ions are represented by cyan spheres. The entrance through the template tunnel for each molecule in the asymmetric unit is highlighted by a red arrow. Note the more accessible entrance in molecule C due to less tight crystallographic contacts with surrounding molecules.

[Figure drawn and rendered with PyMOL (DeLano, 2004)]

2.4.2. 5nt Φ6pol-RNA complex

For the 5nt RNA oligonucleotide co-crystals, the electron density maps in the template region were not very clear and more than one possible binding mode seemed to be present. Initially, electron density maps calculated with CNS (Brunger *et al.*, 1998) after positional and individual B-factor refinement of the model, imposing strict-NCS three-fold constrains on the three molecules present in the asymmetric unit, were interpreted in O and two different possible conformations of the 5nt RNA oligo were built. In order to establish which might be the preferred template binding position, refinement of the occupancies and B-factors for the two possibilities was carried out, using CNS (Brunger *et al.*, 1998). However, the results were inconclusive and visual inspection of the resulting electron density maps to try to establish which conformation had the highest occupancy still proved unsuccessful. These observations lead us to conclude that there is a certain degree of flexibility in the binding of the 5nt RNA oligo along the template tunnel and refinement was done separately for the two distinct but overlapping positions for the oligonucleotides. For the buried conformation, refinement was done to a resolution of 2.0Å, imposing strict three-fold NCS constrains in CNS (Brunger *et al.*, 1998), resulting in a model with R_{factor} of 27.9% (R_{free}=29.7%) and average stereochemistry (rmsd bond length = 0.023Å and rmsd bon angle = 2.2°). Conversely, for the intermediate template conformation, refinement was carried out imposing NCS restraints as defined for the longer RNA oligos complexes, resulting in a final model with R_{factor}=24.1% and R_{free}=27.6% (resolution of 2.0Å) and good stereochemistry (rmsd bond length = 0.015Å and rmsd bond angle = 1.7°). Maps show a similar density distribution of

density in all three copies present in the asymmetric unit. However, initial averaging of the three molecules to improve the signal to noise ratio might have masked differences in the binding position in each molecule. To confirm this, original unaveraged maps were visually inspected and a similar electron density disorder was observed. This indicates that flexibility is an intrinsic property of the binding of 5nt RNA oligonucleotide and not an artefact produced when refining the model.

2.4.3. Φ6pol-RNA-GTP-Mg^{2+} complex

For the WT Φ6pol-RNA co-crystals soaked with GTP and $MgCl_2$, an initial model was obtained after rigid body calculations in CNS (Brunger *et al.*, 1998). Visual inspection of the resulting SIGMAA |Fo|-|Fc| maps revealed density in the active site region and substrate tunnel but none in the template tunnel. The presence of Mn^{2+} in its high affinity binding site was confirmed by analysis of the peak height. Similarly, the presence of two magnesium ions was detected. Close examination of the density at the active site allowed modelling of a reaction product to be identified: guanylyl(3'-5')-guanosine-5'-triphosphate (PPP-G-P-G). This molecule was manually built into the electron density together with an inorganic phosphate (PPi) molecule. As for other WT structural models, internal three-fold averaging of electron density maps with GAP improved the signal-to-noise ratio. Consecutive iterative cycles of manual model building and refinement in CNS (Brunger *et al.*, 1998) to a resolution of 2.5Å, imposing strict non-crystallographic constrains, resulted in a model with R_{factor}=24.5% and R_{free}=26.6% and good stereochemistry (rmsd bond length = 0.016Å and rmsd bond angle = 1.9°).

2.4.4. Φ6pol -RNA-Ca^{2+}-GTP complex

The data collected at the manganese *K* edge for the Φ6pol-RNA-Ca^{2+}-GTP crystals provided a good contrast in anomalous scattering signal between Mn^{2+} (f" = 4.0 e^-) and Ca^{2+} (f" = 1.0 e^-). The three-fold averaged anomalous difference maps showed three peaks above 4σ in each polymerase molecule. The highest peak was 6.5σ, whereas the next two peaks each had a height of some 4.5σ, reflecting the greater f" value for Mn^{2+} and demonstrating that we have a single Mn^{2+} and two Ca^{2+} ions bound per molecule. Initial rigid body refinement with CNS (Brunger *et al.*, 1998) and resulting SIGMAA 2|Fo|-|Fc| maps showed electron density for the oligonucleotides in the template binding tunnel and two GTP molecules in the substrate binding region. An RNA oligonucleotide molecule and two GTP molecules were then built in O (Jones *et al.*, 1991). The three ions identified by observation of anomalous difference Fourier maps also explain all the strong peaks in standard difference Fourier maps, confirming that no Mg^{2+} ions are present in the structure. All electron density maps were then improved by internal three-fold averaging using GAP. Disorder at the external side of the template tunnel is observed in the Φ6pol - RNA-Ca^{2+}-GTP complex, where a 6nt oligonucleotide was used for co-crystallization. Therefore, as for the Φ6pol-6ntRNA complex, only four nucleotides from the 3' end were built and included in refinement. Final positional and individual B-factor refinement of the Φ6pol-RNA-Ca^{2+}-GTP against the data to a resolution of 3.0Å, using strict three-fold NCS constraints, resulted in a model with R_{factor}=23.3% (R_{free}=25.9%) and good stereochemistry (rmsd bond length = 0.015Å and rmsd bond angle = 1.9°).

2.4.5. Φ6pol-EDTA

Data collected at the manganese *K* edge for the Φ6pol-EDTA crystals optimised detection of the anomalous signal from Mn^{2+} via anomalous difference Fourier maps. These maps revealed no significant anomalous signal in the usual Mn^{2+} binding site or elsewhere in the protein, confirming that the experiment had been successful. As with other WT models described, all electron density maps were improved by internal three-fold NCS averaging using GAP. Final refinement of the Φ6pol-EDTA model against its data to 2.5Å resolution, using strict three-fold NCS constraints, resulted in a model with R_{factor}=23.8% (R_{free}=26.7%) and good stereochemistry (rmsd bond length = 0.015Å and rmsd bond angle = 1.7°).

2.4.6. Φ6pol-Mg^{2+} complex

Although data for the Φ6pol-Mg^{2+} complex was not collected at the manganese *K* edge, interpretation of the differences in peak height of SIGMAA |Fo|-|Fc| maps calculated after rigid body refinement allowed identification of the ion. A single significant peak was observed in the electron density map and was interpreted as a Mg^{2+} ion based on the reduced peak height (6σ) compared to that seen for Mn^{2+} in the apo-structure (8σ). The final model after positional and B-factor individual refinement in CNS (Brunger *et al.*, 1998) to a resolution of 2.5Å, under strict three-fold NCS constraints, has an R_{factor} of 26.4% (R_{free} of 29.6%) and good stereochemistry (rmsd bond length = 0.018Å and rmsd bond angle = 2.1°).

2.4.7. E491Q Φ6pol mutant

The apo structure of the E491Q Φ6pol mutant was determined by rigid body refinement of the three molecules present in the asymmetric unit with REFMAC5 (Murshudov *et al.*, 1997), using the WT apo structure as a starting model. Visual inspection of the initial model and manual model rebuilding were carried out in COOT (Emsley and Cowtan, 2004). Final refinement in REFMAC5 (Murshudov *et al.*, 1997) carried out to a resolution of 3.2Å, imposing non-crystallographic symmetry restraints as defined for the Φ6pol-RNA models, resulted in a model with R_{factor}=21.4% (R_{free}=26.8%) and good stereochemistry (rmsd bond length = 0.009Å and rmsd bond angle = 1.2°).

2.4.8. E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} complex

For the E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} model, rigid body refinement of the three molecules of Φ6pol in the asymmetric unit was carried out with CNS (Brunger *et al.*, 1998) using the SG mutant structural model described below (pag. 112), resulting in a model with R_{factor} of 29.3% (R_{free}=29.9%) and good stereochemistry (rmsd bond length = 0.009Å and rmsd bond angle = 1.2°). Due to the poor resolution (3.8Å) of this dataset, no further refinement or rebuilding was carried out. Visualization of electron density maps carried out in COOT (Emsley and Cowtan, 2004) showed some very weak density in the template tunnel and some clear density peaks in the substrate tunnel. Due to the poor quality of the data, no attempts were made to build RNA or GTP molecules into the observed density.

2.4.9. E491Q-DNA-GTP-Mg^{2+} complex

Initial work on the E491Q structures which crystallised in space group $P3_2$, with unit cell dimensions similar to the SeMet WT structure previously determined (Butcher *et al.*, 2001) was done using the best data set collected – E491Q-DNA-GTP-Mg^{2+}. Since the SeMet crystals described by Butcher *et al.* (2001) possessed two molecules in the asymmetric unit (a.u.), initial rigid body refinement calculations were done with these two molecules present, using the SeMet WT structure as a search model. However, these refinements proved unsuccessful. Upon inspection of the resulting electron density maps, it became clear why the refinement was not producing improved models: strikingly, despite similar unit cell dimensions, a third molecule is present in the asymmetric unit. Initial analysis of the cell content did not clearly indicate the correct number of molecules present in the asymmetric unit (a.u.) since Matthews coefficients were 3.6Å^3/Da for two molecules in the a.u. with 65% solvent content and 2.4Å^3/Da for three molecules and a solvent content of 48%. Close examination of the crystallographic contacts of the SeMet $P3_2$ crystal form revealed a loose packing of the molecules, with large gaps that could accommodate a third molecule (Fig. 2.8.A). Therefore, molecular replacement calculations were carried out to determine the orientation of all three molecules present in the asymmetric unit of the E491 complexes. EPMR (Kissinger *et al.*, 1999) was used, using the WT as a starting model and gave clear results for all three molecules, as indicated by the R_{factor} (52.7%, 46.0%, 47.7%, respectively for molecule A, B and C) and correlation coefficients (27%, 44%, 39%, respectively) for each molecule. The crystallographic contacts in the E491Q crystals are much tighter than those in the previous $P3_2$ packing (Fig.2.8.B). Only small rearrangements at the protein surface

seem to be needed for the extra crystallographic contacts to be established and a third molecule accommodated, without major shifts in the orientation and packing of the other two molecules (Fig. 2.8.C). Rigid body refinement was then carried out using REFMAC5 (Murshudov *et al.*, 1997) and visual inspection of the resulting SIGMAA |Fo|-|Fc| maps revealed electron density in the template and substrate tunnels in all three molecules. However, clear differences in the strength of the density were detectable so no internal averaging was carried out and 4nt DNA oligonucleotide (the two nucleotides at the 5' end were disordered and could not be traced) and a GTP molecule were modelled for each Φ6pol monomer. No divalent cations were detectable in the electron density maps. Due to the weak electron density levels for the DNA and GTP molecules these were assigned an occupancy of 0.5. Manual model building followed by refinement in REFMAC5 (Murshudov *et al.*, 1997) to a resolution of 2.4Å after addition of 381 solvent molecules with ARP/wARP (Lamzin and Wilson, 1993; Perrakis *et al.*, 2001) resulted in a final model with R_{factor}=18.4% and R_{free}=25.4% and good stereochemistry (rmsd bond length = 0.004Å and rmsd bond angle = 1.0°).

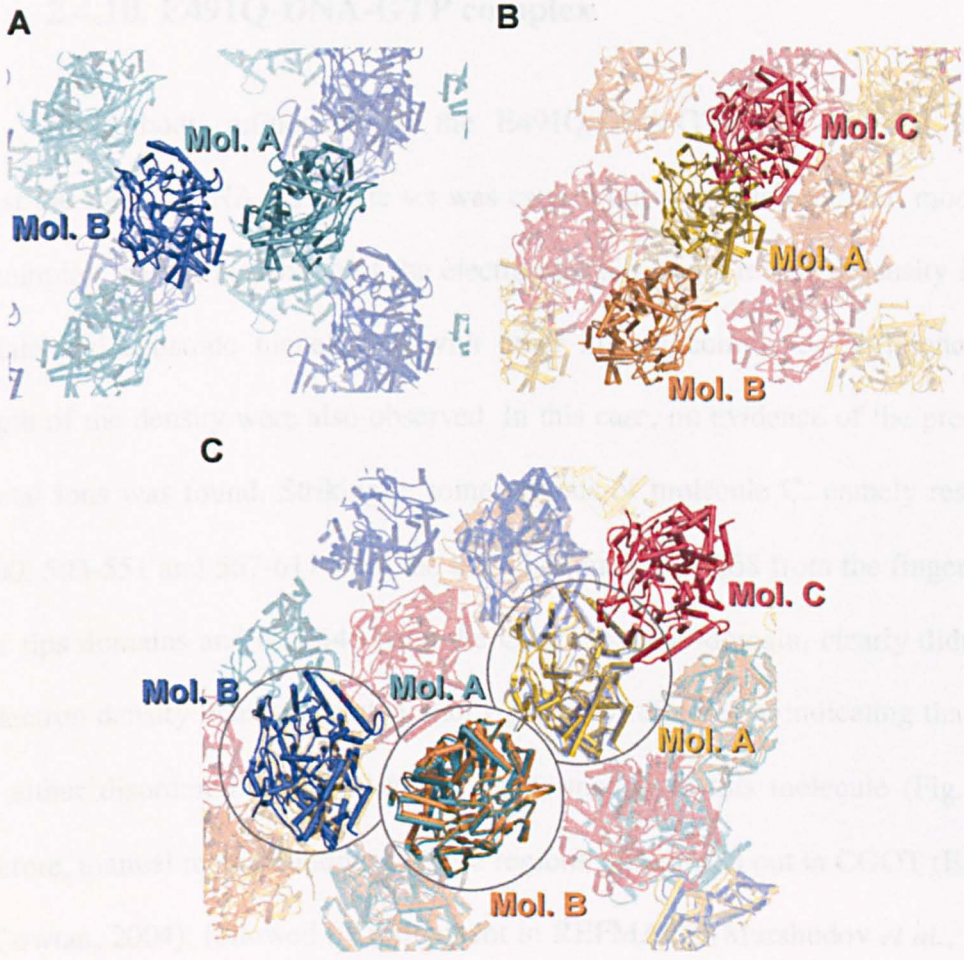


Figure 2.8. **Φ6pol $P3_2$ crystallographic packing**

A. Cartoon representation of the $P3_2$ crystallographic packing of SeMet Φ6pol, with two molecules (A – cyan; B – blue) in the asymmetric unit. Crystallographically related molecules are represented in the same way, in semi-transparent mode. A loose packing, with high solvent content is clearly observable.

B. Cartoon representation of the $P3_2$ crystallographic packing of E491Q Φ6pol complexes, with three molecules (A – yellow; B – orange; C - red) in the asymmetric unit. Crystallographically related molecules are represented in the same way, in semi-transparent mode.

C. Cartoon representation of the $P3_2$ crystallographic packing of SeMet Φ6pol superimposed on the $P3_2$ packing of E491Q complexes in the asymmetric unit. Molecules are represented as above. Differences in the packing are clear, with "molecule C" in the mutated protein occupying the large gaps observed in the SeMet crystallographic packing. Superimposition of the molecules "A" and "B" of each crystal (highlighted by grey circles) shows that the orientation and positioning is equivalent, with only slight rearrangements at the protein surface.

[Figure drawn and rendered with PyMOL (DeLano, 2004)]

2.4.10. E491Q-DNA-GTP complex

Rigid body refinement of the E491Q-DNA-GTP-Mg^{2+} refined model against the E491Q-DNA-GTP data set was carried out to obtain an initial model of this complex. Visual inspection of the electron density maps revealed density in the template and substrate tunnels. As with other mutant complexes, differences in strength of the density were also observed. In this case, no evidence of the presence of metal ions was found. Strikingly, some regions of molecule C, namely residues 37-100, 503-551 and 567-617 from the thumb domain, 204-268 from the fingers and finger tips domains and 618-646 from the C-terminal sub-domain, clearly didn't fit the electron density maps calculated after rigid body refinement, indicating that they were either disordered or in a different conformation in this molecule (Fig. 2.9). Therefore, manual model building of these regions was carried out in COOT (Emsley and Cowtan, 2004), followed by refinement in REFMAC5 (Murshudov *et al.*, 1997) imposing NCS restraints only for molecules A and B. At this stage, a 4nt long DNA oligonucleotide and one GTP molecule were manually fitted into the observed density in each molecule. Final refinement with REFMAC5 (Murshudov *et al.*, 1997) to a resolution of 2.6Å, keeping tight NCS restraints for molecules A and B, resulted in a model with R_{factor}=24.2% and R_{free}=30.2% and good stereochemistry (rmsd bond length = 0.009Å and rmsd bond angle = 1.3°). As with other mutant complex models, the occupancy of the DNA and GTP molecules was kept at 0.5 to reflect the weak density levels, particularly in molecule A.

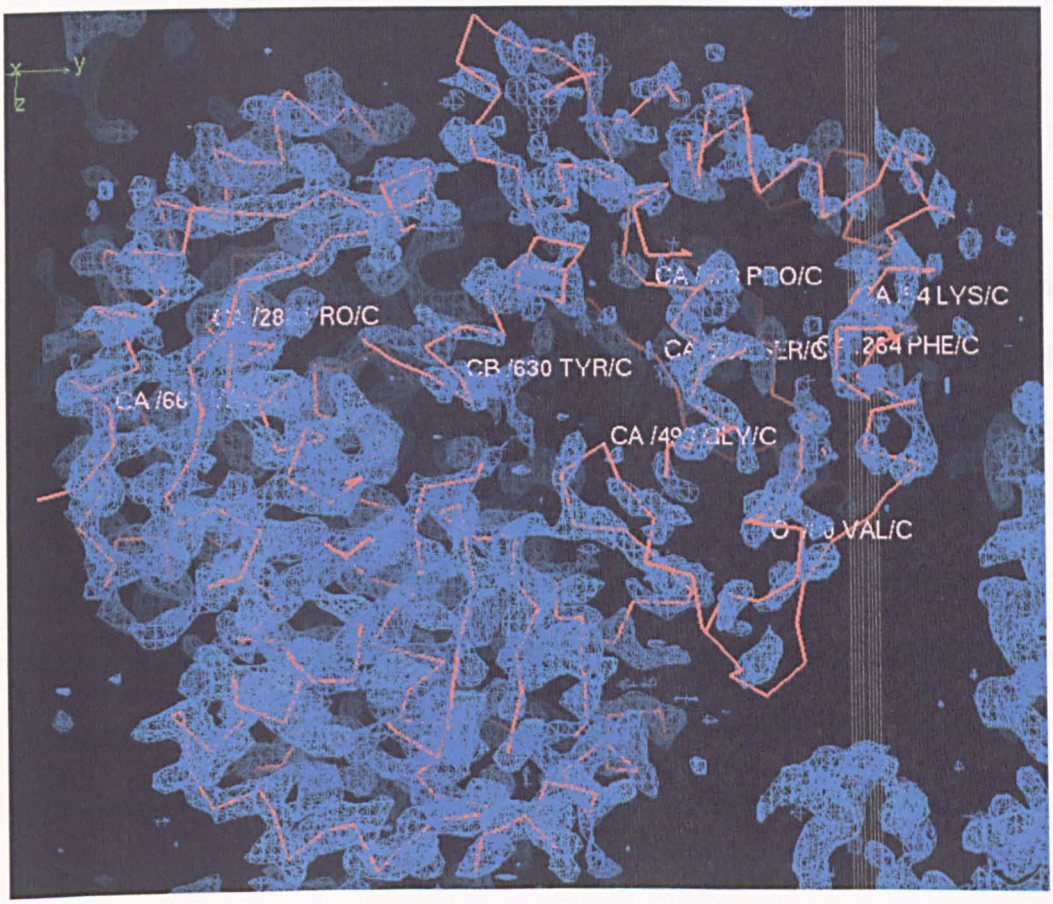


Figure 2.9. Φ6pol E491Q-DNA-GTP (molecule C)

Molecule "C" (ribbon representation, red) in the asymmetric unit of the E491Q-DNA-GTP crystals. SIGMAA |2Fo|-|Fc| maps calculated after rigid body refinement with REFMAC5 (Murshudov *et al.*, 1997) at a contour level of 1.5σ are represented by blue chicken-wire.

[Figure drawn from COOT (Emsley and Cowtan, 2004)]

2.4.11. E491Q-DNA-GTP-Mg^{2+}-Mn^{2+} complex

The E491-DNA-GTP-Mg^{2+}-Mn^{2+} complex initial model (R_{factor}=32.1%, R_{free}=32.4%) was obtained by rigid body refinement with REFMAC5 (Murshudov *et al.*, 1997). For molecules A and B, density observed in the template and substrate tunnels was similar to that seen in the other E491Q-DNA complexes and a 4nt DNA oligo and a GTP molecule were manually modelled. As in other E491Q-DNA complexes, no clear indication of the presence of the divalent cations was detected. Similarly, the occupancy of the ligands was kept at 0.5, due to the weak density levels. Surprisingly, electron density observed in the template and substrate tunnels in molecule C clearly showed a different state had been captured since two molecules of GTP could be identified and DNA seemed to be in a different conformation, away from the S pocket. In fact, when the WT-DNA initiation complex previously determined (Butcher *et al.*, 2001) was superimposed onto molecule C using SHP (Stuart *et al.*, 1979), clear similarities in the DNA and GTP molecules positions were observed. Therefore, those models for DNA and GTP were initially used to model the ligands in molecule C. Furthermore, analysis of the intensity of peaks in |Fo|-|Fc| maps, seemed to indicate that one magnesium ion was present in this molecule, in the substrate binding tunnel. Since the electron density for DNA and GTP was strong in molecule C, occupancies were kept at 1.0 throughout refinement. As for other E491Q complexes, positional and B-factor refinement was carried out in REFMAC5 (Murshudov *et al.*, 1997) to 2.8Å resolution, in this case imposing tight restraints for molecules A and B, resulting in a final model with good stereochemistry and R_{factor}

of 22.3% (R_{free}=29.3%) and good stereochemistry (rmsd bond length = 0.008Å and rmsd bond angle = 1.3°).

2.4.12. E491Q-RNA-Mn^{2+} complex

For the E491Q-RNA-Mn^{2+} data set, rigid body using the *P3₂* mutant model was carried out using REFMAC5 (Murshudov *et al.*, 1997), resulting in a model with R_{factor}=32.6% (R_{free}=33.5%). Visual inspection of the |Fo|-|Fc| maps revealed very poor density in the template region. Strikingly, it was clear that manganese was present in two of the molecules, but not in the third (Fig. 2.10). Therefore, manganese ions were modelled into those two molecules. Due to the low resolution of this data set (3.2Å), no attempts to build the RNA oligos into density were carried out. Final positional and B-factor refinement, imposing tight NCS restraints for each sub-domain in each molecule, resulted in a model an R_{factor} of 21.8% (R_{free}=29.5%) and good stereochemistry (rmsd bond length = 0.007Å and rmsd bond angle = 1.0°).

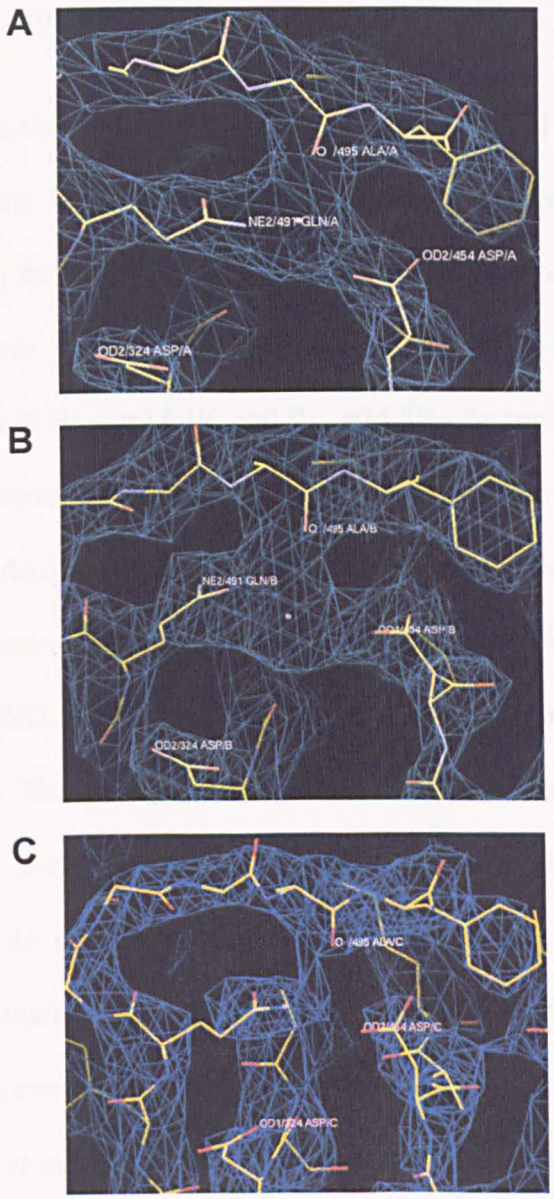


Figure 2.10. E491Q-RNA-Mn^{2+} structure

A. "Mn^{2+} binding site" in molecule "A" (C – yellow, O – red, N – blue). Coordinating residues and neighbouring catalytic D324 are highlighted. SIGMAA |2Fo|-|Fc| maps calculated after rigid body refinement with RFMAC5, contoured at 1.5σ are represented by blue chicken-wire.

B. View in a similar orientation of the "manganese binding site" in molecule "B". Representation as in A.

C. "Manganese binding site" in molecule "C" in the E491Q-RNA co-crystals grown in the presence of Mn^{2+}, viewed similarly as in A and B.

[Figure drawn from COOT (Emsley and Cowtan, 2004)]

2.4.13. SG mutant

To determine the correct positioning and orientation of each molecule in the asymmetric unit of the SG mutant crystals, molecular replacement with AMORE (Navaza, 1994) using the WT protein as a search model was necessary. Rigid body refinement of the three molecules in the asymmetric unit to a resolution of 3.0Å resulted in a model with R_{factor}=34.4% and R_{free}=34.3%. As for WT complexes, GAP was then used to improve the electron density maps by internal three-fold averaging. Absence of electron density in the 2|Fo|-|Fc| map in the mutated loop region as well as strong negative electron density features (-10σ) in the difference Fourier map for residues 629QYKW632, show that the structure of the mutated loop has been disrupted (Fig. 2.11). The mutated residues were modelled using CALPHA (Esnouf, 1997). Finally, the electron density map does not show strong density at the manganese binding site, indicating the absence of the ion. Due to the poor density in the region of the mutation and the overall low resolution of the data, no further manual building was carried out. Final positional and B-factor individual refinement with CNS (Brunger *et al.*, 1998) to a resolution of 3.0Å, imposing three-fold non-crystallographic constraints resulted in a model with R_{factor}=24.1% and R_{free}=28.0% and good stereochemistry (rmsd bond length = 0.014Å and rmsd bond angle = 1.8°).

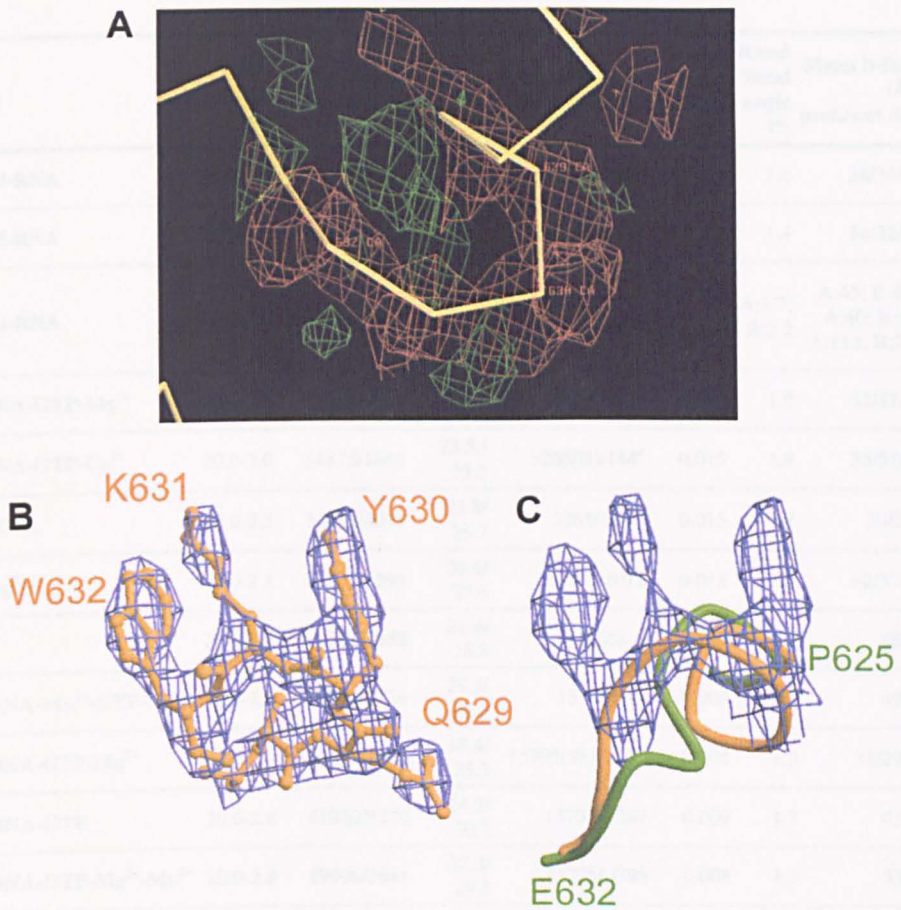


Figure 2.11. SG mutant loop disordered conformation

A. SIGMAA |Fo|-|Fc| electron density maps, contoured at -1.5 (red chicken-wire) and 1.5σ (green chicken-wire) levels, clearly showing the altered path of the loop.

B. The 3-fold averaged difference electron density map reveals negative electron density for residues in loop 629QYWK632, drawn in as ball and stick representation in orange.

C. Cartoon representation showing the change in path of the polypeptide main chain on substitution of QYKW (orange) by SG (lime green). The main chain is drawn from residues 625 to 633 (WT numbering), the approximate positions of residues S and G are marked and the view and electron density contour level are identical to that drawn in B.

Table 2.3. Refinement statistics for Φ6pol structures

Structure	Resol. range (Å)	No. reflct. (work/test)	R_{factor}[a]/ R_{free}[b] (%)	No. of atoms: Prot./wat./lig.	Rmsd bond length (Å)	Rmsd bond angle (°)	Mean B-fact. (Å²) prot./wat./lig.	Rmsd backb. B-fact. (Å²)
7nt Φ6pol-RNA	20.0-1.9	199005/10430	24.3/ 27.1	15795/648/234[+]	0.015	1.6	34/34/78	3.4
6nt Φ6pol-RNA	20.0-2.2	135268/7208	23.3/ 25.9	15795/439/234	0.012	1.4	34/32/65	3.5
5nt Φ6pol-RNA	20.0-2.0	167968/8877	A:24.1; B:27.9 / A:27.6; B:29.7	15795/443/294[§a]; 5265/98[§b]	A:0.015 ; B:0.023	A:1.7; B:2.2	A:45; B:42 / A:40; B:44/ A:113; B:114	A:3.5; B:4.3
Φ6pol-RNA-GTP-Mg^{2+}	20.0-2.5	92226/4874	24.5/ 26.6	5265/96/67[¶]	0.016	1.9	42/47/73	4.3
Φ6pol-RNA-GTP-Ca^{2+}	20.0-3.0	31875/1661	23.5./ 25.3	5265/80/144[‡]	0.015	1.9	55/51/95	3.7
Φ6pol-EDTA	20.0-2.5	84233/4071	23.8/ 26.7	5265/102/-	0.015	1.7	39/51/-	4.1
Φ6pol-Mg^{2+}	20.0-2.5	82175/4392	26.4/ 29.6	5265/101/1	0.018	2.1	62/57/25	4.5
E491Q	20.0-3.2	38111/2038	21.4/ 26.8	15795/-/-	0.009	1.2	48/-/-	0.4
E491Q-RNA-Mn^{2+}-GTP-Mg^{2+}	20.0-3.8	24083/1224	29.3/ 29.9	15795/ -/ -	0.009	1.2	49/-/-	3.2
E491Q-DNA-GTP-Mg^{2+}	20.0-2.4	76778/4055	18.4/ 25.2	15795/381/1151	0.004	1.0	31/29/54	3.1
E491Q-DNA-GTP	20.0-2.6	61930/3270	24.2/ 30.2	15795/-/381	0.009	1.3	42/49	0.5
E491Q-DNA-GTP-Mg^{2+}-Mn^{2+}	20.0-2.8	49096/2641	22.3/ 29.3	15795/-/395	0.008	1.3	33/46	0.6
E491Q-RNA-Mn^{2+}	20.0-3.2	33578/1772	21.8/ 29.5	15795/-/2	0.007	1.0	43/75	0.4
SG	20.0-3.0	46972/2510	24.1/ 28.0	5231/-/-	0.014	1.8	37/-/-	4.0

[a] values for A (intermediate) and B (buried) positions, when appropriate ; [b] values in parenthesis refer to the highest resolution shell ; [c] $R_{merge} = \Sigma_i\Sigma_h(|I_{i,h} - \langle I_h \rangle|) / \Sigma_i\Sigma_h(\langle I_h \rangle)$, where h are unique reflections indices, $I_{i,h}$ are intensities of symmetry-related reflections and $\langle I_h \rangle$ is the mean intensity; [d] R_{factor} 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_{factor}; 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 ; [+] 5'-UUCC-3' (3), Mn^{2+}(3); [§] 5'-UUCC-3'(3), Mn^{2+}(3) ; [§a] 5'-UUUCC-3' (3), Mn^{2+}(3); [§b] 5'-UUUCC-3' (3), Mn^{2+}(3); [¶] G-P-G-PPP (1), PPi (1), Mn2+ (1) ; [‡] 5'-UUCC-3' (1), GTP (2), Ca^{2+} (2), Mn^{2+} (1)

Chapter 3

Φ6 RdRP – Analysis of results

The structural studies on Φ6 RNA-dependent RNA polymerase, described in this thesis, were carried out in an attempt to answer questions about the details of the mechanism of RNA polymerisation. In this chapter, a description of each structure determined is given, together with an analysis of its contribution to the overall understanding of Φ6 polymerase mechanism. As in Chapter 2, the descriptions and analyses of the different structures are grouped into subchapters according to the question that they address:

(i) RNA specificity – RNA complexes

(ii) Structural rearrangements upon catalysis – "Dead-end" complex

(iii) Divalent ion effect: Ca^{2+} effect – Φ6pol-RNA-GTP-Ca^{2+} complex; Mn^{2+} role – "manganese-free" models; E491Q structures

(iv) Molecular details of reaction mechanism and role of Mn^{2+} – E491Q complexes

(v) Initiation platform relevance – SG mutant

To allow comparisons of the structures described with others previously solved, superposition of the structural models was carried out with SHP (Stuart *et al.*, 1979). Unless otherwise stated, figures in this chapter were drawn using BOBSCRIPT (Esnouf, 1999) and rendered with RASTER3D (Merritt, 1997).

3.1. Φ6pol RNA Specificity

Although Φ6pol can utilize both DNA and RNA templates, RNA templates are preferred (Makeyev and Bamford, 2000a). Analysis of the electron density maps for each of the three different lengths of RNA (5nt, 6nt and 7nt) revealed the oligonucleotide bound in the template tunnel (Fig. 3.1). Throughout this chapter, the 3' nucleotide is denoted T1 and oligonucleotides are numbered sequentially from 3' to 5', according to the nomenclature established by Butcher *et al.* (2001).

3.1.1. 6nt and 7nt Φ6pol-RNA complexes

For these longer RNA oligonucleotides, the binding mode is similar to that described by Butcher *et al.* (2001) for a 5nt DNA oligonucleotide. Nevertheless, the presence of the extra hydroxyl group of RNA (at the O2' of the sugar ring) causes changes in the sugar conformation of the nucleotides and slight rearrangements in the neighbouring polymerase residues. This OH group is able to form extra hydrogen bonds, further stabilizing the bound RNA oligonucleotides within the tunnel, when compared to DNA oligonucleotides. The structures of the 6 and 7 nucleotide RNA oligos are very similar, with only slight variations in the orientation of the bases, particularly in the 3' end (Fig. 3.1). It is therefore possible to define an overall mode of RNA template binding and compare it to the DNA binding mode (Butcher *et al.*, 2001).

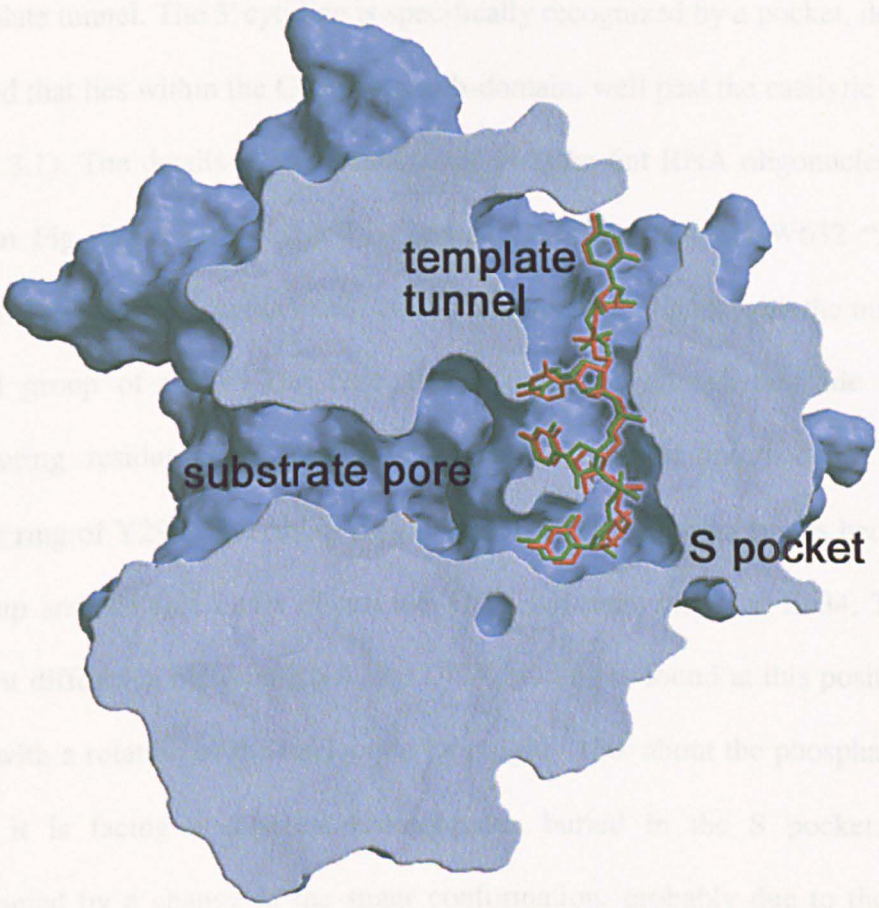


Figure 3.1. Binding of RNA oligonucleotides

A slice through the surface representation of Φ6pol. 6nt (red) and 7nt (green) RNA oligonucleotides are represented in the template tunnel. The 3' cytidine is deeply buried in site S.

[From (Salgado *et al.*, 2004)]

As in the Φ6pol-DNA complex (Butcher *et al.*, 2001), the RNA fits snugly in the template tunnel. The 3' cytidine is specifically recognized by a pocket, defined as site S and that lies within the C-terminal sub-domain, well past the catalytic site (site C) (Fig. 3.1). The details of the interactions with the 6nt RNA oligonucleotide are shown in Fig. 3.2.A (left). The T1 cytidine faces the 629QYKW632 "initiation platform" loop region (Laurila *et al.*, 2002), and hydrogen bonds with the main chain carbonyl group of Q629. The base also hydrogen bonds to the side chain of neighbouring residue K451 and establishes hydrophobic interactions with the aromatic ring of Y295. The ribose ring is stabilized by hydrogen bonds between the O2' group and the side chain of residue T633 and main chain of E634. The most important difference between RNA and DNA binding is found at this position (Fig. 3.2.B), with a rotation of the nucleotide by roughly 180° about the phosphate group, so that it is facing a different environment, buried in the S pocket. This is accompanied by a change in the sugar conformation, probably due to the need to accommodate the extra OH group (O2'), which forms hydrogen bonds with residues T633 and E634. When DNA is bound in the tunnel, the sugar is accommodated within the loop and the base interacts with residues T633 and E634 (Fig. 3.2.A, right).

The T2 cytidine establishes hydrophobic interactions with residues R291 and A272 and base stacks with the uracil in position T3, which forms hydrogen bonds with G275, M273, R204 and K543 (Fig. 3.2.A, left). The last visible uracil nucleotide, T4, has fewer interactions with the polymerase and is less well ordered. In comparison, equivalent hydrogen bonds are not present in the DNA template complex (Fig. 3.2.A, right). The slight variations found in the interactions of the

remaining nucleotides between RNA and DNA are probably due to the chemical differences between DNA and RNA nucleotides.

It is clear that, in complex with Φ6pol, the ribose sugar rings of RNA nucleotides assume a different conformation compared to that seen in the Φ6pol-DNA complex, which induces subtle rearrangements of the surrounding residues. These subtle changes lead to additional favourable interactions between the O2' hydroxyl group and surrounding residues that stabilize the template within the tunnel (Fig. 3.2.A), presumably explaining why ssDNA (from bacteriophage M13 linearized with restriction endonuclease *HinfI*) is only a poor template for this enzyme (Makeyev, unpublished data).

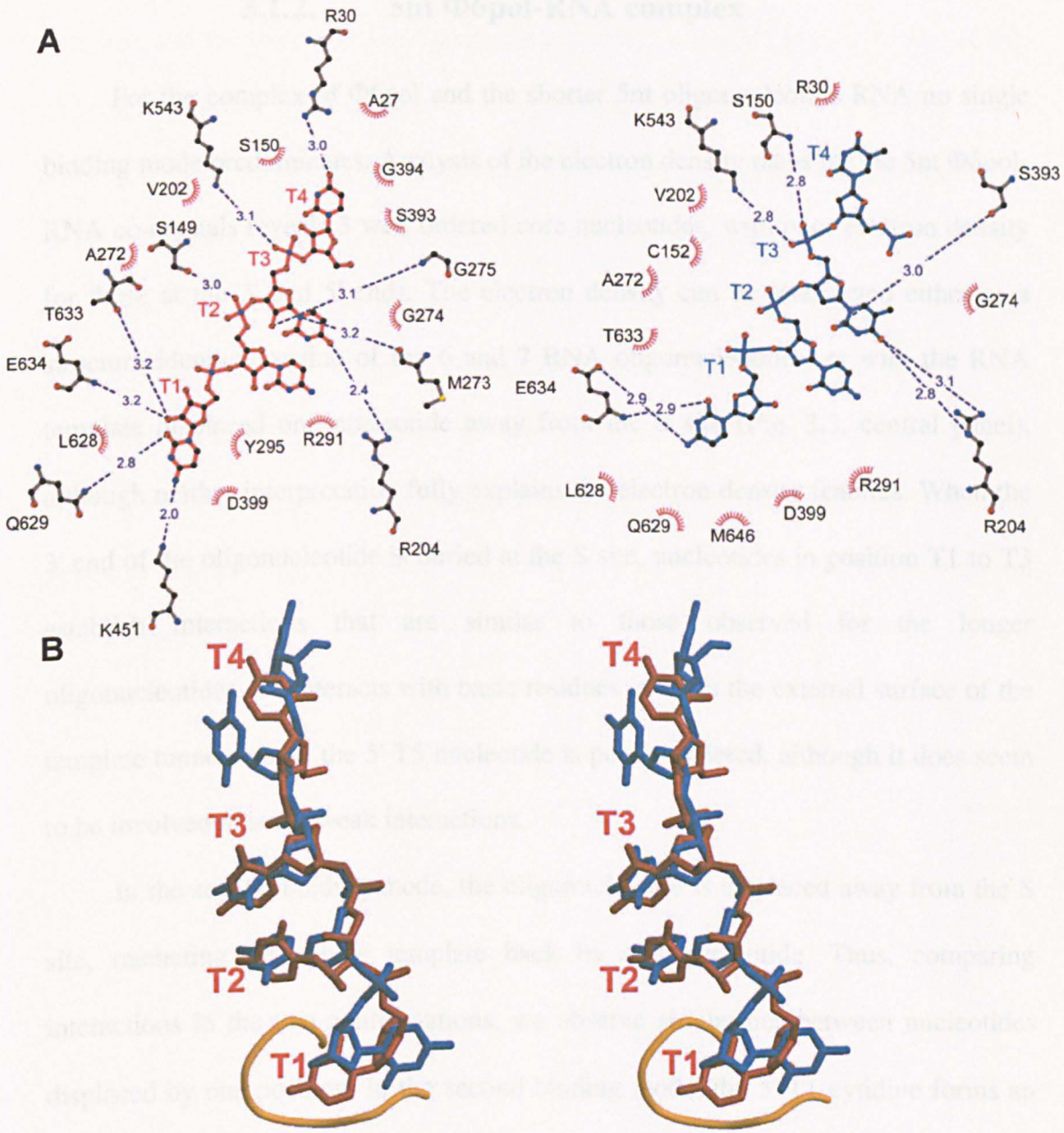


Figure 3.2. RNA template (6nt) vs. DNA template binding to Φ6pol

A. LIGPLOT (Wallace *et al.*, 1995) representations of the template/protein interactions for the RNA (red) oligonucleotide (6nt) (left) and DNA (blue) oligonucleotide (5nt) (right).

B. Stereo view of RNA (red) and DNA (blue) (from Butcher *et al.*, 2001) binding to the template tunnel. The presence of the OH group in the sugar causes rearrangements of the nucleotides with the most significant difference in position T1.

[From (Salgado *et al.*, 2004)]

3.1.2. 5nt Φ6pol-RNA complex

For the complex of Φ6pol and the shorter 5nt oligonucleotide RNA no single binding mode predominates. Analysis of the electron density maps for the 5nt Φ6pol-RNA co-crystals reveals 3 well ordered core nucleotides, with poor electron density for those at the 3' and 5' ends. The electron density can be interpreted either as a structure identical to that of the 6 and 7 RNA oligonucleotides, or with the RNA template displaced one nucleotide away from the S site (Fig. 3.3, central panel), although neither interpretation fully explains the electron density features. When the 3' end of the oligonucleotide is buried at the S site, nucleotides in position T1 to T3 establish interactions that are similar to those observed for the longer oligonucleotides. T4 interacts with basic residues towards the external surface of the template tunnel, whilst the 5' T5 nucleotide is poorly ordered, although it does seem to be involved in some weak interactions.

In the second binding mode, the oligonucleotide is displaced away from the S site, ratcheting the whole template back by one nucleotide. Thus, comparing interactions in the two conformations, we observe similarities between nucleotides displaced by one position. In the second binding mode, the 3' T1 cytidine forms an hydrogen bond with S149 and base stacks with cytidine T2. T2 forms fewer favourable interactions than the equivalent uracil T3 in the other conformation (Fig. 3.3, central panel). Nucleotides T3 and T4 are in similar orientations to T4 and T5 in the first binding mode, with T4 being stabilized mainly by interaction with the previous nucleotide.

The observed variations demonstrate how the template is accommodated along the channel whilst it migrates along it, with only very slight changes in conformation

of both RNA oligonucleotides and protein. Less stabilizing interactions between the template and the protein in the intermediate position allow it to be easily displaced, either towards the S site or away from it, leading to the observed flexibility in the binding of the 5nt RNA oligonucleotide. The need for a longer oligonucleotide to produce a more stable complex demonstrates that interactions with amino acids distant from the active site stabilize the template binding (Fig. 3.2.A). In line with this, a number of those amino acid residues are conserved across the *Cystoviridae* polymerases, namely S149, R204, A272 and R291. Furthermore, S149 and R204 do not belong to the palm domain, where sequence conservation across the family are generally found, suggesting they are likely to play significant functional roles in template binding.

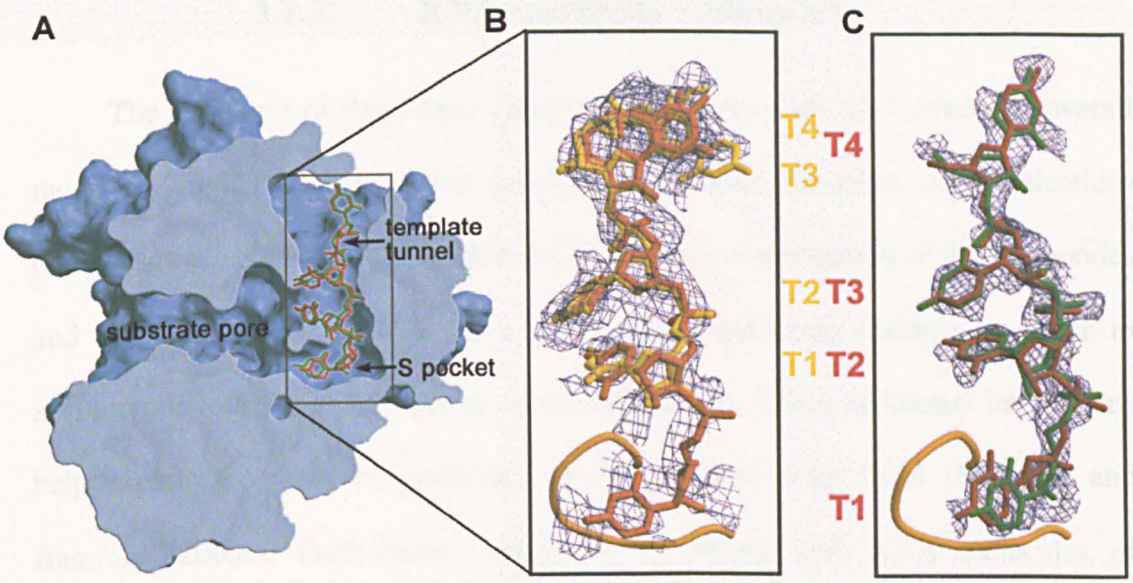


Figure 3.3. **5nt RNA oligonucleotides binding vs. longer RNA templates**

A. A slice through the surface representation of Φ6pol (as in Fig. 3.1)

B. Flexible binding of a 5nt RNA oligonucleotide. The predominant conformations are shown in yellow and deep orange. The polypeptide chain for the 629QYKW632 loop is shown in pale orange. Despite the striking differences at the 3′ and 5′ ends, the internal nucleotides are largely superimposable. An |Fo|-|Fc| map contoured at 1.5σ around the RNA oligonucleotide is drawn in blue (RNA was omitted from the phasing model).

C. 6nt RNA (red) and 7nt RNA (green) oligonucleotides binding in the template tunnel with 629QYKW632 loop as in central panel. |Fo|-|Fc| map contoured at 1.5σ around the RNA oligonucleotide (for which the RNA was omitted from the phasing model) is shown in blue chicken wire.

[From (Salgado *et al.*, 2004)]

3.1.3. RNA specificity - summary

The structures of these three complexes of Φ6pol with RNA reveal an overall mode of binding similar to that previously described for DNA oligonucleotides (Butcher *et al.*, 2001) with clear changes in the sugar conformation of the nucleotides and slight rearrangements in the neighbouring polymerase residues in order to accommodate the extra OH group on the ribose ring. These additional interactions help explain the observed preference of the polymerase for RNA (Makeyev and Bamford, 2000a). Furthermore, complexes of Φ6pol with RNA molecules of different lengths reveal multiple modes of binding and a plasticity of interaction, indicating that nucleotides exterior to the template tunnel contribute significantly to stable template binding. Therefore, we suggest that transport of RNA through the tunnel during the initiation of polymerization and elongation would be facilitated via interactions with residues at the tunnel surface, distant from the active site and the S pocket. These observations agree with the previously proposed mechanism for opening the dsRNA template (Butcher *et al.*, 2001) where the highly charged surface around the entrance of the template tunnel (plough) would separate the strands, feeding one directly into the tunnel and the other out of the viral capsid.

3.2. *In crystallo* polymerisation

Φ6pol-RNA-GTP-Mg^{2+} complex

In the attempt to confirm the nature of molecular interactions in an initiation competent complex when the preferred template of Φ6pol is present, Φ6pol-RNA co-crystals were soaked in GTP and Mg^{2+}, under similar experimental conditions to those used to visualise the DNA initiation complex (Butcher *et al.*, 2001). Surprisingly, in this case, reaction occurs within the crystal at a rate too high to allow the initiation complex to be captured crystallographically. Instead, a round of the polymerization reaction occurs, with production of guanylyl(3'-5')-guanosine-5'-triphosphate (PPP-G-P-G). This polymerisation proceeds via nucleophilic attack by the O3* of D1 on the oxygen atom of the α-phosphate group in D2 causing the formation of a new covalent bond between D1 and D2 and subsequent release of pyrophosphate (PPi) as a by-product. The polymerised product is located close to, but displaced from, the catalytic site and has presumably rearranged after the reaction. The di-nucleotide has a closed, collapsed, conformation with the equivalent faces of bases D1 and D2 stacked together (Fig. 3.4.A). The triphosphate group of the 5'-guanosine is stabilized by coordination to a Mg^{2+} ion (denoted #1, see Fig. 3.4.A & B) and the base of the guanosine D2, locking the product into the closed conformation. Mg^{2+}#1 is itself locked in position by coordination with the hydroxyl group of Y630 (Fig. 3.4.B). This conformation is also stabilized by interactions between the sugars of the two nucleosides and hydrogen bonds between the phosphate backbone of D1 and the base as well as the stacking of the side chain of Y630 against the guanosine base D1. The reaction by-product PPi is coordinated by

Mg^{2+}#2, presumably normally transiently, prior to release through the substrate/sub-products pore (see Fig. 3.4.B for detailed interactions). The electron density in the template tunnel is poor, probably due to the loss of the RNA template from the tunnel since there is a lack of further substrate (ATP) molecules to continue the reaction. It seems that two base-pairs are insufficient to stabilize the polymerised dsRNA product, allowing the di-nucleotide to fall away from the RNA template, to form a dead-end state. Elongation is also presumably prevented by the C-terminal sub-domain, which is locked in place by crystal contacts, blocking the exit pore.

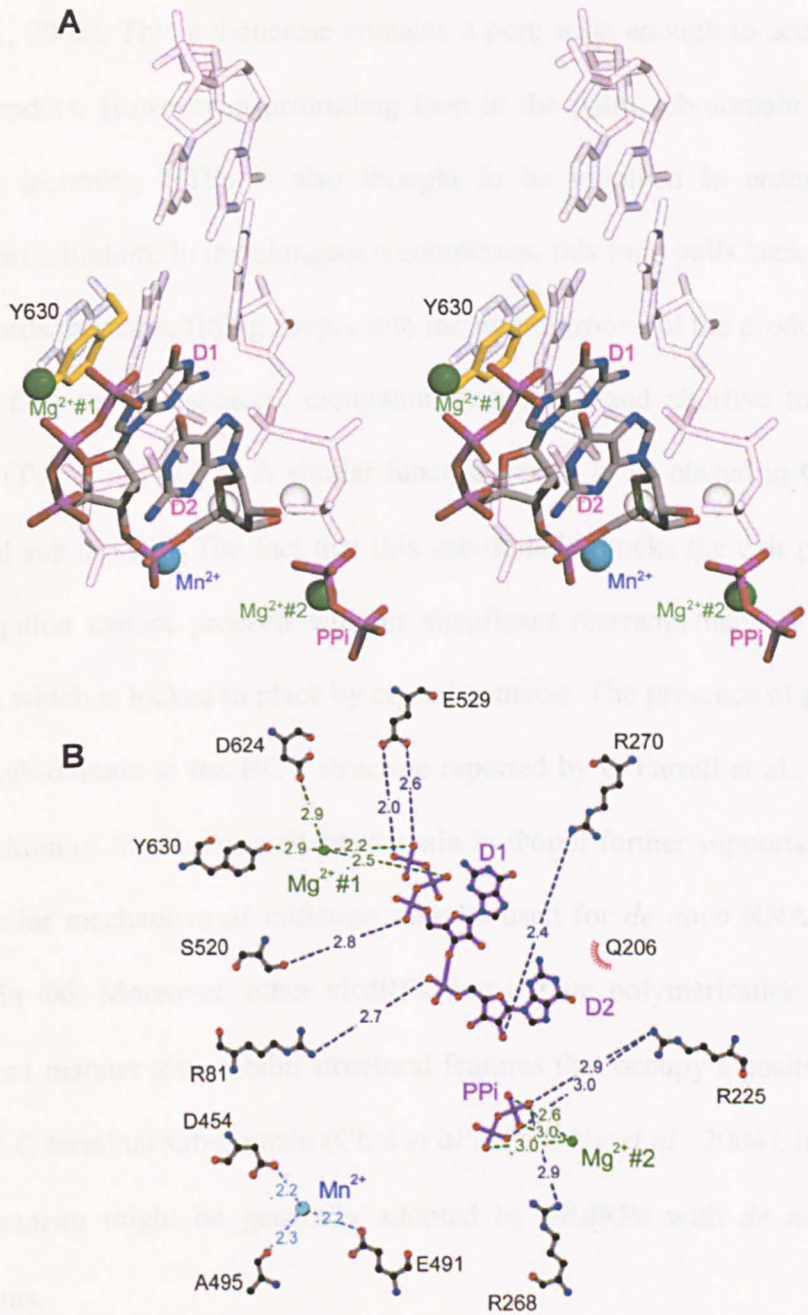


Figure 3.4. "Dead-end" vs. initiation complex

A. Stereo figure of the superposition of the dead-end complex (ball and stick representation, coloured by atom type: red – O, blue – N, magenta – P, grey – C, cyan – Mn^{2+}, green – Mg^{2+}) with the DNA initiation complex (equivalent representation but drawn as transparent) described by Butcher *et al.* (2001). Mg^{2+} coordinates the polymerised product, the by-product PPi and Y630 (shown in yellow).

B. LIGPLOT (Wallace et al., 1995) representation of the interactions between the reaction product PPP-G-P-G, the by-product PPi, the divalent cations Mg^{2+}, Mn^{2+} and Φ6pol.

[From (Salgado *et al.*, 2004)]

Elongation complexes for the reovirus polymerase have recently been reported (Tao *et al.*, 2002). This polymerase contains a pore wide enough to accommodate a dsRNA product. However, a protruding loop in the palm sub-domain proposed to prime the incoming NTPs is also thought to be involved in ensuring primer-independent initiation. In the elongation complexes, this loop pulls back by some 3Å away towards the palm, fitting snugly into the minor groove of the product duplex. If the loops fails to be displaced, elongation is blocked and abortive transcripts are produced (Tao *et al.*, 2002). A similar function seems to be played in Φ6pol by the C-terminal sub-domain. The fact that this sub-domain blocks the exit pore signifies that elongation cannot proceed without significant rearrangement of this domain (Fig. 3.5), which is locked in place by crystal contacts. The presence of part of the C-terminal sub-domain in the HCV structure reported by O'Farrell et al., (2003) close to the position of the C-terminal sub-domain in Φ6pol further supports the proposal that a similar mechanism of initiation may be used for *de novo* RNA synthesis in HCV as in Φ6. Moreover, other vRdRPs that initiate polymerisation in a primer-independent manner also exhibit structural features that occupy a position similar to the Φ6pol C-terminal sub-domain (Choi *et al.*, 2004; Ng *et al.*, 2004), indicating that this mechanism might be generally adopted by vRdRPs with *de novo* initiation mechanisms.

Furthermore, these studies have demonstrated Φ6pol *in crystallo* polymerase activity, validating the models for the mechanism of polymerisation derived from other studies of Φ6pol, previously carried out by Butcher *et al.* (2001) and those described in this thesis.

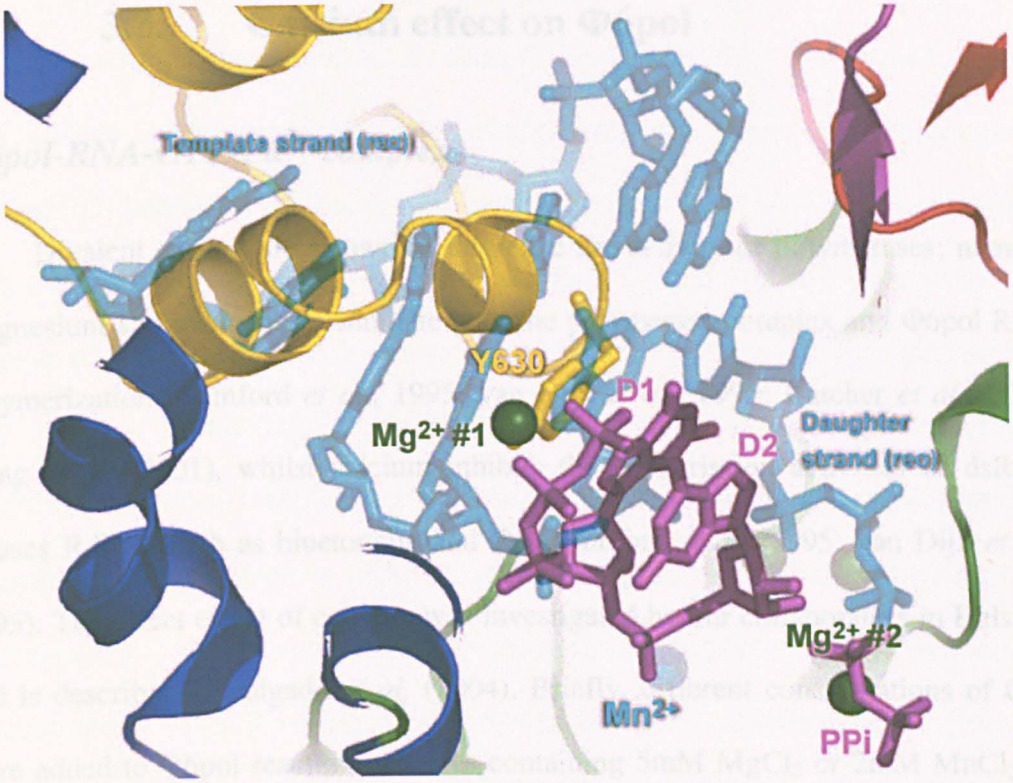


Figure 3.5. Comparison of structures of Φ6pol "dead-end" complex with reovirus λ3pol elongation complex

Superposition of the elongation complex of reovirus λ3pol (semi-transparent ball-and-stick representation, coloured cyan) onto the coordinates of the "dead-end" complex. PPPGPG, PPi (magenta) and Y630 (yellow) are shown as ball-and-stick representation. Manganese and magnesium ions are represented as in Fig. 3.4.A. Φ6pol is coloured according to sub-domains: palm - -green, fingers – red, thumb – blue and C-terminal – yellow. The figure clearly shows that the C-terminal in Φ6pol blocks the exit route of the dsRNA. Therefore, for elongation to occur, this sub-domain has to be displaced which, in this crystallographic system, would disrupt crystallographic contacts. Conversely, the reoviral RdRP has an opened dsRNA exit route and elongation complexes could be captured in the crystals (Tao *et al.*, 2002).

[Prepared with PyMol (DeLano, 2004)]

3.3. Calcium effect on Φ6pol

Φ6pol-RNA-GTP-Ca^{2+}-complex

Divalent cations are known to influence the activity of polymerases; namely, magnesium and manganese stimulate both the polymerase complex and Φ6pol RNA polymerization (Bamford *et al.*, 1995; van Dijk *et al.*, 1995; Butcher *et al.*, 2000; Yang *et al.*, 2001), whilst calcium inhibits the transcription activities of dsRNA viruses RdRPs such as bluetongue and Φ6 (Bamford *et al.*, 1995; van Dijk *et al.*, 1995). The direct effect of calcium was investigated by our collaborators in Helsinki and is described in Salgado *et al.* (2004). Briefly, different concentrations of Ca^{2+} were added to Φ6pol reaction mixtures containing 5mM $MgCl_2$ or 2mM $MnCl_2$. In both conditions, Ca^{2+} clearly inhibits RNA polymerisation, although half-inhibition is achieved at ~0.050mM in the presence of magnesium but only at 0.3mM in the presence of manganese, indicating that Mn^{2+} might have some protective effect.

Analysis of the maps obtained from crystals of the Φ6pol-RNA-Ca^{2+}-GTP complex, obtained as described in Chapter 2, shows two GTP molecules base paired to the T1 and T2 cytidines of the template, coordinated by two Ca^{2+} ions. The molecular details of calcium inhibition are better understood by comparing the structure of the complex with that of the initiation complex obtained by Butcher *et al.* (2001). Briefly, in that complex, the 3' cytidines T1 and T2 are positioned within the catalytic site, forming Watson-Crick base pairs with 2 incoming GTP molecules [denoted D1 and D2, according to the notation of Butcher *et al.*, (2001)]. D1 is further stabilized by base stacking with the ring from Y630. Two Mg^{2+} ions coordinate the phosphate backbones of the incoming GTPs. In the Φ6pol-RNA-Ca^{2+}-

GTP complex, the two catalytic Mg^{2+} ions are replaced by Ca^{2+}. One of the calcium ions, that would be equivalent to the catalytically crucial Mg^{2+} (denoted #1, see Fig. 3.6), is displaced by some 4.6Å away and no longer co-ordinates the triphosphates of D1 and D2. Instead, it coordinates the hydroxyl group of Y630 and the triphosphate of D1 (Fig. 3.6.B). Ca^{2+} ions have a more flexible coordination, with an average number of ligands of seven compared to the rather rigid octagonal coordination of Mg^{2+} with six ligands (Glusker, 1999; Saris *et al.*, 2000). As a result, calcium might establish the observed interactions to complete the favoured heptagonal coordination. Rotation of the side chain of Y630, which usually provides a stabilizing platform for the incoming GTP, D1, via aromatic interactions (Laurila *et al.*, 2002), renders the base pairing of D1 to T1 of the RNA template less stable, and displaces D1 from D2, preventing the nucleophilic attack by O3* of D1 to one of the oxygen atoms in the α-phosphate group. The second Ca^{2+} (denoted #2) is situated between the triphosphate of D2 and the protein, in the same position as one of the Mg^{2+} ions in the initiation complex. D2 is further stabilized by base pairing with T2 and interactions with the side chains of neighbouring residues (Fig. 3.6.B). Overall, calcium ions cause a reorganisation of the complex, inhibiting catalysis by subtle changes in the geometry of the initiation complex.

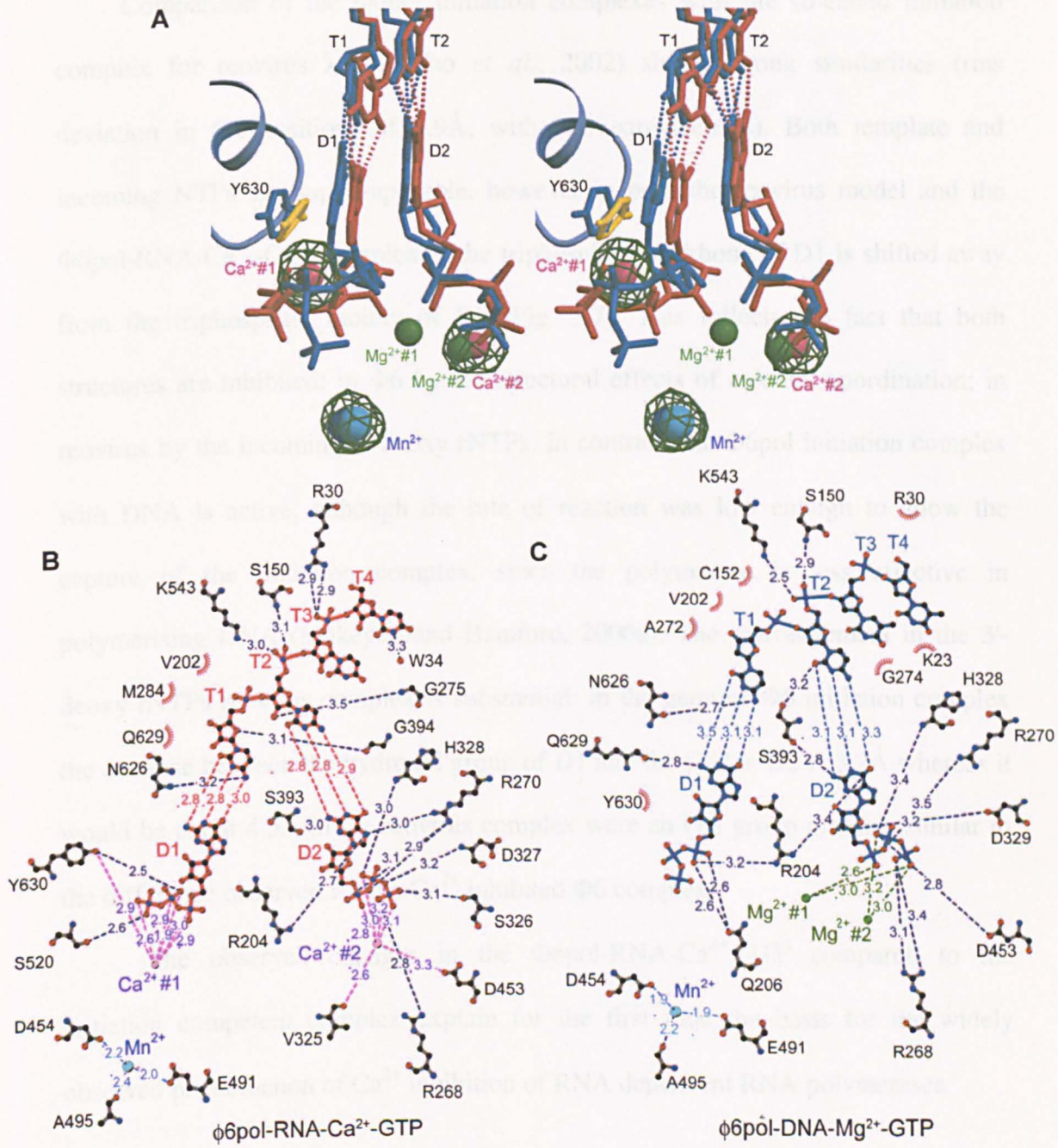


Figure 3.6. Ca^{2+}inhibition of the initiation competent state

A. Stereo figure of the superposition of the active initiation structure (Φ6pol-DNA-Mg^{2+}-GTP) (Mg^{2+} coloured green, other atoms light blue) onto the Φ6pol-RNA-Ca^{2+}-GTP structure (red). Base pairing interactions are shown as dotted lines. Ca^{2+} ions are shown in pink and Mn^{2+} ions in cyan. The anomalous difference Fourier map is shown as green chicken wire. Y630 is shown in yellow (ball-and-stick representation).

B. LIGPLOT representation of the interactions between RNA, GTP, Ca^{2+}, Mn^{2+} and Φ6pol.

C. LIGPLOT representation of the interactions between DNA, GTP, Mg^{2+}, Mn^{2+} and Φ6pol.

[From (Salgado et al., 2004)]

Comparison of the Φ6pol initiation complexes with the so-called initiation complex for reovirus λ3pol (Tao *et al.*, 2002) shows strong similarities (rms deviation in Cα positions of 2.9Å, with 280 equivalences). Both template and incoming NTPs are superimposable, however in both the reovirus model and the Φ6pol-RNA-Ca^{2+}-GTP complexes, the triphosphate backbone of D1 is shifted away from the triphosphate moiety of D2 (Fig. 3.7). This reflects the fact that both structures are inhibited: in Φ6 by the structural effects of calcium coordination; in reovirus by the incoming 3'-deoxy rNTPs. In contrast, the Φ6pol initiation complex with DNA is active, although the rate of reaction was low enough to allow the capture of the initiation complex, since the polymerase is less effective in polymerising DNA (Makeyev and Bamford, 2000a). The rearrangement in the 3'-deoxy rNTPs reovirus complex is substantial: in the genuine Φ6 initiation complex the distance between the hydroxyl group of D1 and the O-P in D2 is 3.2Å whereas it would be about 4.5Å in the reovirus complex were an OH group present (similar to the difference observed for the Ca^{2+} inhibited Φ6 complex).

The observed changes in the Φ6pol-RNA-Ca^{2+}-GTP compared to the initiation competent complex explain for the first time the basis for the widely observed phenomenon of Ca^{2+} inhibition of RNA dependent RNA polymerases.

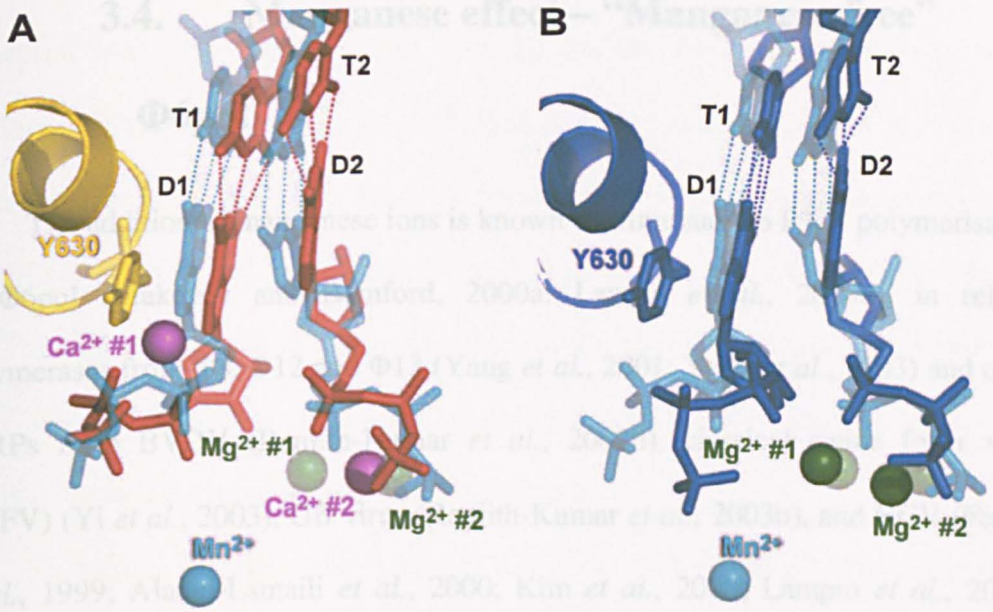


Figure 3.7. Comparison of Φ6pol initiation and calcium-inhibited complexes with λ3pol initiation complex structures

A. Superposition of the initiation complex of reovirus λ3pol (semi-transparent, cyan ball-and-stick representation; Mg – dark green spheres) onto the coordinates of the "calcium-inhibited" complex (red). 6nt RNA oligonucleotide, GTP molecules, manganese and magnesium ions are represented as in Fig. 3.6.A. Φ6pol Q629-W632 platform loop (yellow) is shown as cartoon representation. Y630 is drawn as ball-and-stick representation.

B. Superposition of the initiation complex of reovirus λ3pol (shown as in A) onto the coordinates of the DNA initiation complex described by Butcher *et al.* (2001). 6nt DNA oligonucleotide, GTP molecules, manganese and magnesium ions are represented as in Fig. 3.6.A. Φ6pol Q629-W632 platform loop (marine blue) is shown as cartoon representation. Y630 is drawn as ball-and-stick representation.

[Prepared with PyMol (DeLano, 2004)]

3.4. Manganese effect – "Manganese-free" Φ6pol

The addition of manganese ions is known to stimulate Φ6 RNA polymerisation in Φ6pol (Makeyev and Bamford, 2000a; Laurila *et al.*, 2005a), in related polymerases from Φ8, Φ12 and Φ13 (Yang *et al.*, 2001; Yang *et al.*, 2003) and other RdRPs from BVDV (Ranjith-Kumar *et al.*, 2002b), classical swine fever virus (CSFV) (Yi *et al.*, 2003), GB virus (Ranjith-Kumar *et al.*, 2003b), and HCV (Ferrari *et al.*, 1999; Alaoui-Lsmaili *et al.*, 2000; Kim *et al.*, 2000; Lampio *et al.*, 2000). Previous studies identified a high affinity manganese binding site in Φ6pol (Butcher *et al.*, 2001) in the palm domain, coordinated by D454, E491 and A495. Mn^{2+} ions have also been detected in crystal structures of rabbit hemorrhagic disease virus (Ng *et al.*, 2002), reovirus λ3 (Tao *et al.*, 2002) and HCV RdRPs (Bressanelli *et al.*, 2002; O'Farrell *et al.*, 2003). Due to its location in Φ6pol, some distance from the active site (~6Å), it was hypothesised that the ion could have a structural role, in stabilizing the overall structure of Φ6 polymerase. To investigate this hypothesis, the structures of Φ6pol without bound manganese ions were determined.

The data collected at the manganese *K* edge for the Φ6pol-EDTA crystals allowed for unambiguous identification of a water molecule in the "manganese binding site" through analysis of anomalous difference Fourier maps (see Chapter 2 for details). In the crystal structure, replacement of the Mn^{2+} seems to have no apparent structural consequences either in surrounding residues (Fig. 3.8.A & C)or the overall structure of the protein.

Previous studies (Butcher *et al.*, 2001) reported a structure of Φ6pol with the structural Mn^{2+} replaced by a Mg^{2+} ion. To confirm this result, crystals of Φ6pol were crystallised in the presence of EDTA and subsequently soaked in a solution containing an excess of Mg^{2+}. Electron density maps analysis allowed unambiguous identification of bound Mg^{2+} in the "manganese binding site". Overall, this Φ6pol-Mg^{2+} complex structure is superimposable on the model previously described with Mn^{2+} with no significant structural changes in either the cation binding region (Fig. 3.8.B & C) or the overall structure of the polymerase.

Therefore, the molecular basis for the stimulation of Φ6pol activity by Mn^{2+} ions (Butcher *et al.*, 2000; Laurila *et al.*, 2002) remains unclear although it is possible that Mn^{2+} can act as the catalytic ion replacing Mg^{2+} as reported for other polymerases (Arnold *et al.*, 1999; Baulcombe, 2002; Tao *et al.*, 2002; O'Farrell *et al.*, 2003). Other possibilities include effects in the early stages of initiation or at the switch from initiation to elongation, not detectable by these experiments.

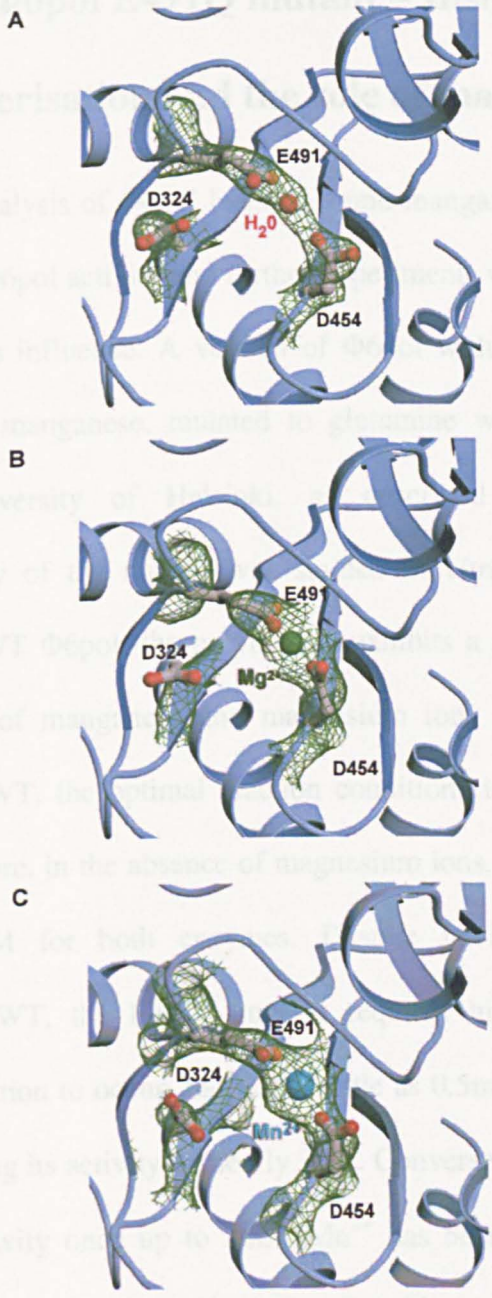


Figure 3.8. Binding at the "manganese site"

The interacting residues D454, E491 and D324 (the latter interaction is via a water molecule which is not shown), are shown in ball and stick representation (blue – N atoms, red – O atoms, grey – C atoms). The green chicken wire shows a SIGMAA 2|Fo|-|Fc| map contoured at 1.5σ levels.

A. Water molecule (red) occupying the binding site in the Φ6pol-EDTA structure.

B. Mg^{2+} (green) occupying the site in the Φ6pol-Mg^{2+} structure.

C. Mn^{2+} (cyan) binding as previously described (Butcher *et al.*, 2001).

3.5. Φ6pol E491Q mutant – Insights into polymerisation and the role of manganese

The structural analysis of Φ6pol lacking bound manganese did not explain its stimulatory effect on Φ6pol activity and further experiments were therefore designed to try to elucidate this influence. A version of Φ6pol with E491, one of the key residues coordinating manganese, mutated to glutamine was constructed by our collaborators in University of Helsinki, as described in Chapter 2. The polymerisation activity of the mutant was studied *in vitro* (Koivunen *et al.*, in preparation). As for WT Φ6pol, the polymerase exhibits a preference for a 3' end cytidine. The effects of manganese and magnesium ions on the mutant activity revealed that, as for WT, the optimal reaction conditions include 5mM Mg^{2+} and 2mM Mn^{2+}. Furthermore, in the absence of magnesium ions, the optimal manganese concentration is 5mM for both enzymes. Despite having the same optimal requirements as the WT, the E491Q mutant requires higher concentrations of manganese for stimulation to occur. Indeed, as little as 0.5mM Mn^{2+} has a dramatic effect in WT, increasing its activity by nearly 30%. Conversely, E491Q only exhibits the same level of activity once up to 2mM Mn^{2+} has been added to the reaction mixture. Furthermore, in the absence of Mn^{2+}, reduced levels of RNA synthesis were detected for the mutant. The rate of elongation of the WT and mutant proteins was also determined. Surprisingly, E491Q exhibits an elongation rate ~7.5 times lower than the WT protein. Furthermore, whilst addition of manganese increased the elongation rate of WT protein, no effect of the ion was detected in E491Q elongation rate. In parallel to the biochemical characterisation carried out at the University of

Helsinki by our collaborators, diffraction experiments were performed on crystals, prepared in Oxford, of E491Q protein alone and in complex with templates and/or substrates.

3.5.1. E491Q Φ6pol mutant

The apo protein crystal is isomorphous to previously solved WT models (Butcher *et al.*, 2001). Overall, the model can be superimposed on the WT structure (rms deviation in Cα positions of 0.4Å), with no significant differences detectable. However, unlike WT, where manganese is present even when absent from the crystallization buffer and/or precipitant solution (unpublished data), no Mn^{2+} was identified in the E491Q crystal. As a consequence, Q491 assumes a slightly different conformation, establishing direct interactions with A495 and D454, whereas these are mediated by the Mn^{2+} ion in the WT protein (Fig. 3.9). Whilst these changes seem to have no effect in the overall structure of the protein, the structure of the mutant protein seems to exhibit greater flexibility, as discussed below.

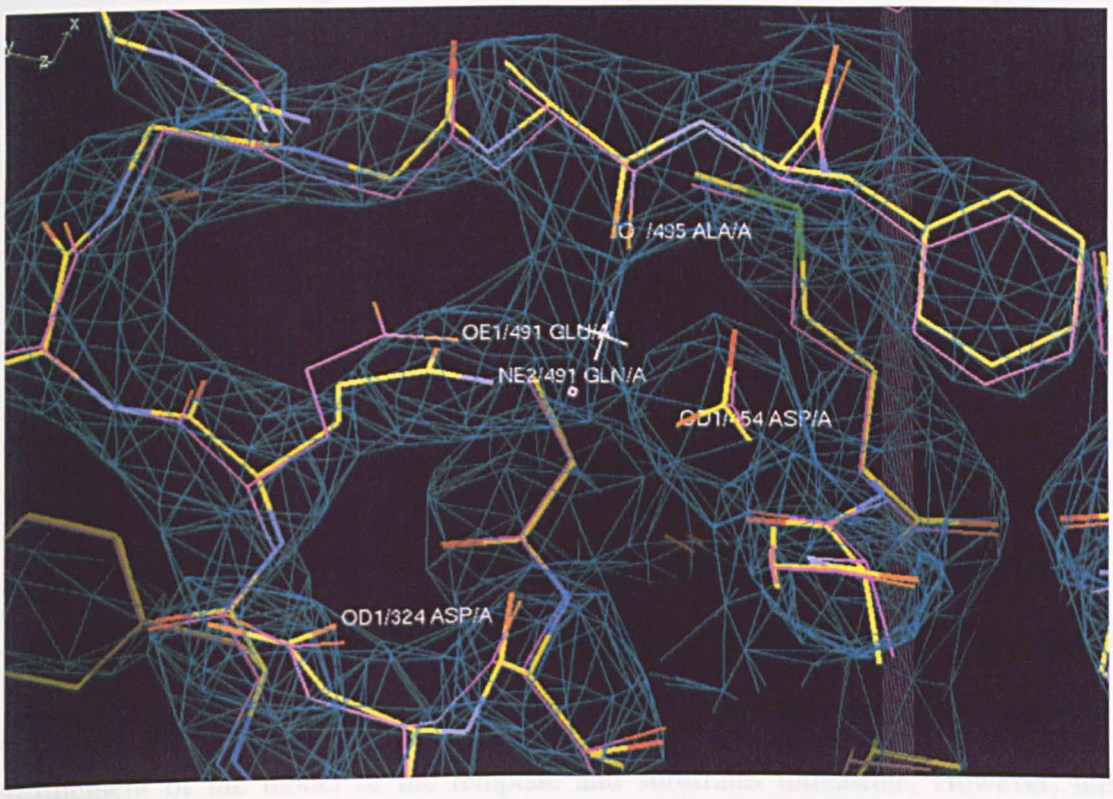


Figure 3.9. Φ6pol E491Q structure

Superposition of Φ6pol E491Q (C – yellow; O – red; N – blue) and WT (C – purple; O – red; N – blue) structural models at the "manganese binding site". SIGMAA 2|Fo|-|Fc| electron density map after rigid body refinement in REFMAC5, contoured at 1.5σ, are shown as blue chicken-wire.

[Figure drawn from COOT (Emsley and Cowtan, 2004)]

To further investigate the effect of the mutation in connection to template and substrate binding, complexes of the mutated protein E491Q Φ6pol with different templates and substrates were determined. To simplify the discussion, each complex is described individually and overall conclusions are discussed at the end of this subchapter.

3.5.2. E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} complex

E491Q co-crystals with an RNA oligonucleotide, grown in the presence of manganese and soaked in a solution containing 25mM GTP and 5mM Mg^{2+} diffracted to poor resolution (3.8Å, see table 2.2 and section 2.4.8), rendering refinement of the model of the template and substrates unfeasible. However, the electron density map calculated using phases from rigid body refined coordinates, revealed several interesting aspects of this complex. Surprisingly, despite the fact that diffracting E491Q-RNA co-crystals only grew in the presence of manganese ion, no Mn^{2+} was detectable in the electron density maps. Clear differences in electron density maps in the template and substrate tunnels are observed in the three molecules present in the asymmetric unit. Two of the molecules (denoted "A" and "B", as defined for WT $P2_1$ molecules in the a.u., Fig. 2.7) show similar electron density features with no identifiable manganese ion bound. This results in weak electron density for Q491 side-chain, presumably due to weaker interactions to surrounding residues in the absence of the ion. In the substrate tunnel, a clear electron density peak is observable, probably corresponding to the phosphate backbone of a GTP molecule bound in a position similar to ATP in a complex in the "NTP interrogation" state previously described by Butcher *et al.* (2001) (Fig.

3.10.A). Density for RNA is very weak in both molecules, particularly deep into the tunnel, indicating that the template is not present at full occupancies and may not completely penetrate the tunnel (Fig. 3.10.A). In molecule C, weak electron density is detectable further into the tunnel (Fig. 3.10.B). However, the template does not seem to bind into the S pocket, assuming an intermediate conformation within the tunnel, in a situation similar to that observed for the 5nt RNA oligo bound to WT Φ6pol. In the NTP substrate channel, electron density features in the C molecule are equivalent to those observed in molecules A and B (Fig. 3.10.B). As with WT complexed with 6nt and 7nt RNA oligonucleotides, differences in electron density in the three molecules are likely to be related to the accessibility of the template tunnel due to crystallographic contacts (see Fig. 2.7).

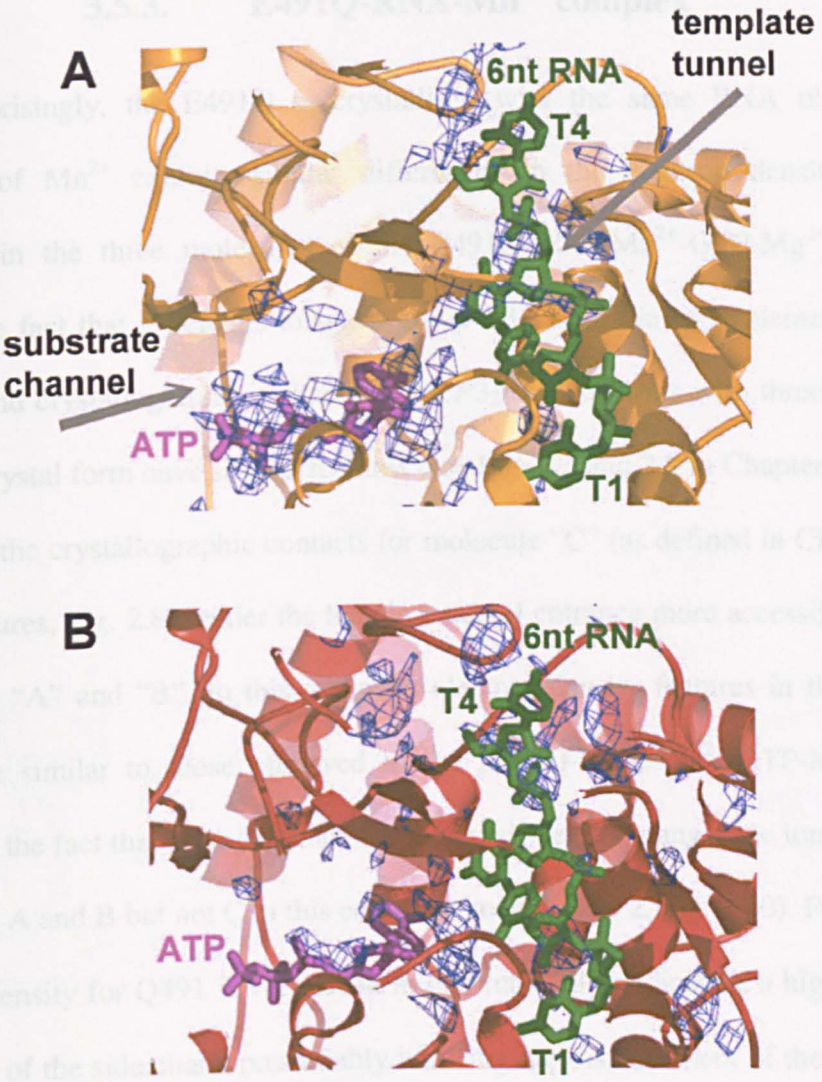


Figure 3.10. E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} structure

A. Electron density features (1.5σ contour level) in the substrate channel and template tunnel in molecules A and B.

B. Electron density features (1.5σ contour level) in the substrate channel and template tunnel in molecule C.

WT Φ6pol ATP-bound complex previously described by Butcher et al. (2001) was superimposed onto the E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} coordinates using SHP (Stuart *et al.*, 1979) and the ATP molecule is drawn here in ball-and-stick representation. The 6nt Φ6pol-RNA complex described in section 3.1 was also superimposed onto the coordinates of the mutant complex and that template model is drawn in green ball-and-stick representation.

[Prepared with PyMol (DeLano, 2004)]

3.5.3. E491Q-RNA-Mn^{2+} complex

Surprisingly, the E491Q co-crystallised with the same RNA oligo in the presence of Mn^{2+} exhibits similar differences in the electron density features observed in the three molecules of the E491Q-RNA-Mn^{2+}-GTP-Mg^{2+} complex, despite the fact that it belongs to space group $P3_2$. This can be explained since the packing and crystallographic contacts in the $P3_2$ crystal forms with three molecules and $P2_1$ crystal form have similar features (see Fig. 2.7 and 2.8 in Chapter 2). In fact, as in $P2_1$, the crystallographic contacts for molecule "C" (as defined in Chapter 2 for $P3_2$ structures, Fig. 2.8) render the template tunnel entrance more accessible than for molecules "A" and "B". In this complex, electron density features in the template tunnel are similar to those observed in the E491Q-RNA-Mn^{2+}-GTP-Mg^{2+}. More striking is the fact that clear evidence for the presence of a manganese ion is found in molecules A and B but not C in this complex (see Chapter 2, Fig. 2.10). Furthermore, electron density for Q491 is very weak in this molecule, indicating a high degree of flexibility of the side chain, presumably inducing the displacement of the 66-77 loop in the thumb domain away from the 489-493 loop by 1.5Å (Fig. 3.11.B). This rearrangement might lead to further slight changes in the overall structure of the protein, namely in the thumb and finger domains (residues 37-100, 503-551 and 567-617 - thumb, 204-268 - fingers and fingertips) (Fig. 3.11.A). Furthermore, the 210-223 loop is displaced by ~4Å towards the protein core, assuming a more "closed" conformation than in the WT protein structure (Fig. 3.11.C). These differences in the protein structure might indicate greater intrinsic flexibility in the mutant, more manifest in the structure of molecule C due to less stringent crystallographic contacts (as seen in Fig. 2.8, Chapter 2).

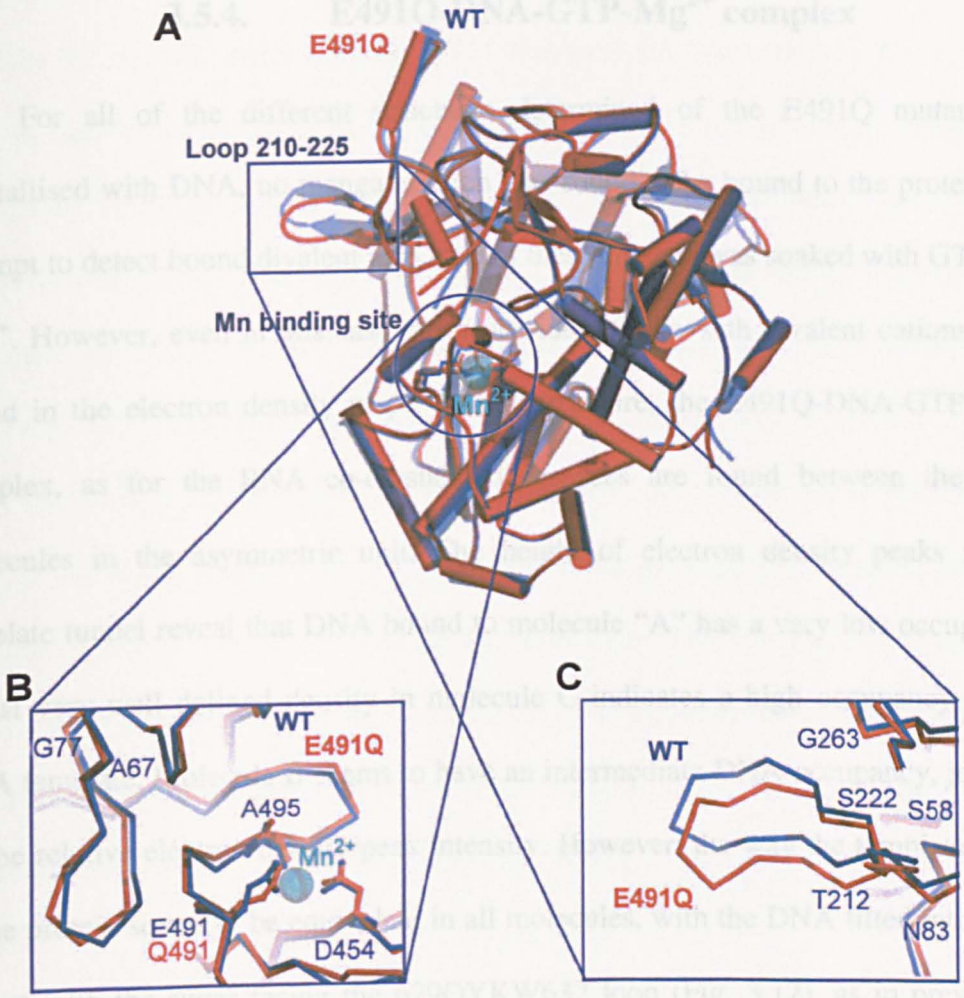


Figure 3.11. E491Q-RNA-Mn^{2+} structure

A. Cartoon representation of the overall superposition of molecule C in the E491Q-RNA-Mn^{2+} (red) structural model with the WT model (marine blue). Manganese ion from the WT model is shown as cyan sphere.

B. Cα trace of E491Q-RNA-Mn^{2+} (red) and WT (marine blue) models, zoomed in on the "manganese binding site" and loop 66-78. D454, A495 and E/Q491 that coordinate the manganese ion (WT model, cyan sphere) are shown in ball-an-stick representation, coloured according to atom type (O – light red; N – blue; C – red in mutant, marine blue in WT). Q491 has a slightly different conformation that has a cascade effect, causing the displacement of loop 66-78 away from the protein core by 1.5 Å.

C. Cα trace of E491Q-RNA-Mn^{2+}(red) and WT (marine blue) models, zoomed in on loop 210-223. In the mutant, the loop is displaced towards the core of the protein by ~4Å.

[Prepared with PyMol (DeLano, 2004)]

3.5.4. E491Q-DNA-GTP-Mg^{2+} complex

For all of the different structures determined of the E491Q mutant co-crystallised with DNA, no manganese ion was found to be bound to the protein. To attempt to detect bound divalent ions, one of these crystals was soaked with GTP and Mg^{2+}. However, even in this case, no peaks identifiable with divalent cations were found in the electron density maps. In this structure, the E491Q-DNA-GTP-Mg^{2+} complex, as for the RNA co-crystals, differences are found between the three molecules in the asymmetric unit. The height of electron density peaks in the template tunnel reveal that DNA bound to molecule "A" has a very low occupancy, whilst very well defined density in molecule C indicates a high occupancy of the DNA template. Molecule B seems to have an intermediate DNA occupancy, judging by the relative electron density peak intensity. However, the way the template binds to the protein seems to be equivalent in all molecules, with the DNA fitted into the S pocket with the sugar facing the 629QYKW632 loop (Fig. 3.12), as in previously described models (Butcher *et al.*, 2001).

Whilst molecule A and B are readily superimposable to the WT model, molecule C exhibits clear differences. In molecules A and B, only small differences are found in the conformation of Q491 that establishes direct interactions with A495 and D454, as in the apo mutant model. As with the E491Q-RNA-Mn^{2+} model, molecule C assumes a somewhat more closed conformation, due to the displacement of the loop 210-223 by ~4Å towards the protein core.

The more significant differences observed between the three molecules in the asymmetric unit are found in the GTP binding position. In all molecules, the phosphate backbone is found in a similar position as that of ATP in the "NTP

interrogating" complex (Butcher *et al.*, 2001), establishing hydrogen bonds with key residues R225, R268 and R270 (Fig. 3.12). However, the orientation of the sugar and base in each molecule is different. In molecule A, the sugar of the incoming GTP establishes hydrogen bonds with R225 and D209, whilst the base interacts with thumb residues N517 and N523. In contrast, in molecules B and C, the sugar and base face residues S452 and D453 and establish interactions with residues R270, D329 and T398. Moreover, the base is close to, but not interacting with, the bases of DNA nucleotides T2 and T3 (Fig. 3.12.B). This position seems to be an intermediate state, where the incoming GTP base is already positioned ready to establish Watson-Crick base-pairing interactions with the DNA template. The fact that, in this position, R270 establishes interactions with the base of the incoming GTP corroborates the idea that this residue plays a central role in the identification of the correct NTP by the polymerase, as proposed by Butcher *et al.* (2001).

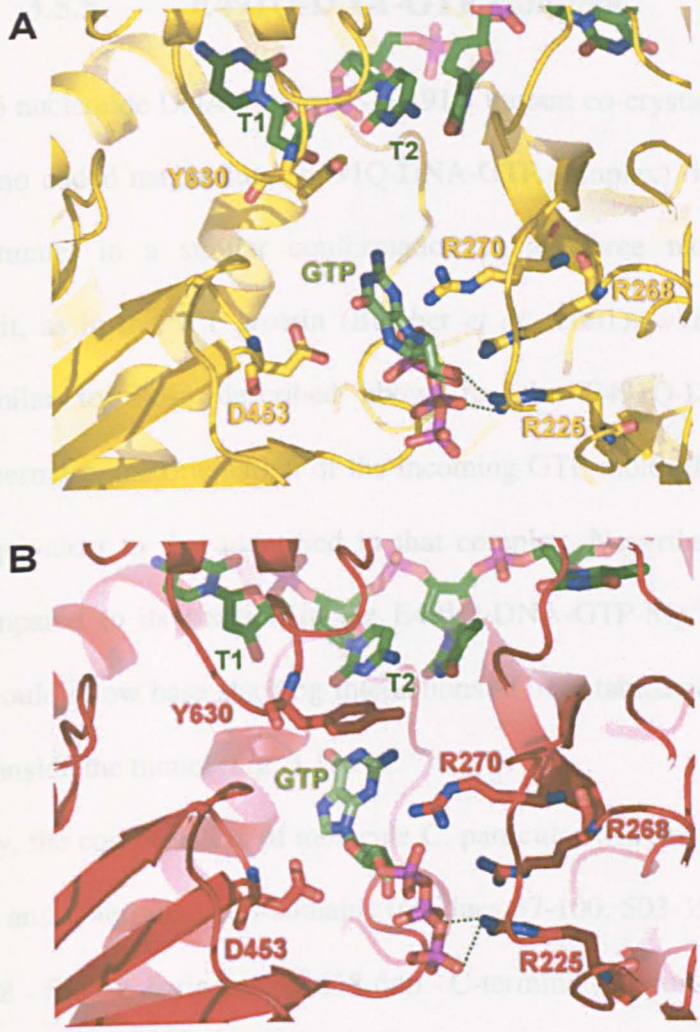


Figure 3.12. Structure of the E491Q-DNA-GTP-Mg^{2+} complex

A. Cartoon representation of the conformation of the complex in molecule A. 6nt DNA (ball-and-stick representation, coloured by atom type: O – red; N – blue; C – green) binds in the template tunnel, with the 3′ cytidine T1 fitted snugly in the S pocket. GTP is shown as ball-and-stick, (coloured by atom type with C – lime green). It establishes hydrogen-bonds (as green dotted lines) with R225 and R270 (ball-and-stick, coloured by atom type with C – yellow). "Priming" R268, catalytic D453 and "initiation platform" Y630 are represented in a similar way.

B. Cartoon representation of 6nt DNA (shown as in A) and GTP (shown as in A) binding mode in molecule C (red) (equivalent to that observed in molecule B). Residues R225, R268, R270, D453 and Y630 are represented as above, with C atoms coloured dark red. The hydrogen bonds between GTP and the protein (red) are represented as green dotted lines.

[Prepared with PyMol (DeLano, 2004)]

3.5.5. E491Q-DNA-GTP complex

When a 6 nucleotide DNA oligomer – E491Q mutant co-crystal was soaked in GTP but with no added magnesium (E491Q-DNA-GTP complex), DNA binds into the template tunnel in a similar conformation in all three molecules in the asymmetric unit, as in the WT protein (Butcher *et al.*, 2001), with differences in occupancy similar to those described above for the E491Q-DNA-GTP-Mg^{2+} complex. Furthermore, the orientation of the incoming GTP molecules in molecules A and B is equivalent to that identified in that complex. Nevertheless, Y630 has rearranged compared to its position in the E491Q-DNA-GTP-Mg^{2+} complex to a position that would allow base stacking interactions to be established once GTP has moved further inside the tunnel (Fig. 3.13.A).

Strikingly, the conformation of molecule C, particularly in some regions of the thumb, fingers and C-terminal sub-domains (residues 37-100, 503-551 & 567-617 – thumb; 204-268 - fingers & fingertips; 618-646 - C-terminal), is considerably altered (see Chapter 2, Fig. 2.9). In this molecule, the incoming GTP assumes a "post-interrogation" conformation, with the base establishing interactions with the bases of nucleotides T2 and T3 in the template tunnel (Fig. 3.13.B). Moreover, analysis of the electron density maps reveals weak density that could be interpreted as a low occupancy second GTP molecule bound into the tunnel, ready to establish base stacking interactions with Y630 and trigger rearrangements of the template to form a pre-initiation competent state. Changes in the overall structure of the protein might be related to the fact that interactions between the template and GTP have triggered some of the conformational rearrangements that will allow displacement of the template away from the S pocket to form the initiation competent complex.

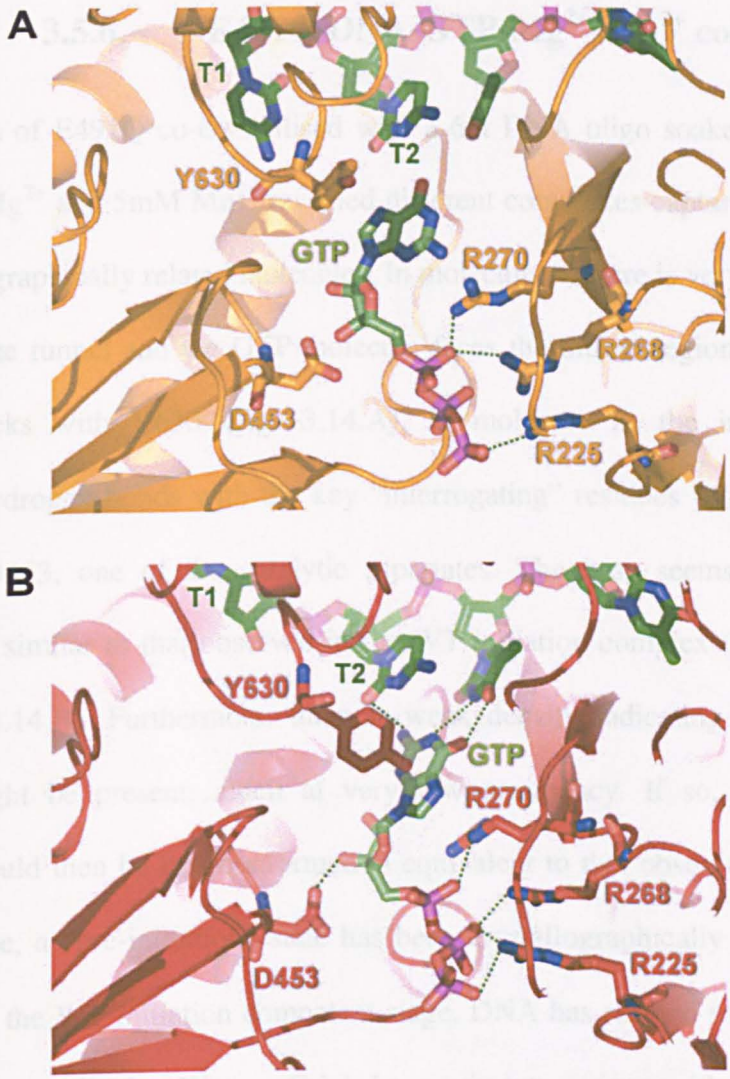


Figure 3.13. E491Q-DNA-GTP structure

A. Cartoon representation of the conformation of the complex in molecule B (equivalent to conformation of molecule A). 6nt DNA (ball-and-stick representation, coloured by atom type: O – red; N – blue; C – green) binds in the template tunnel, with the 3′ cytidine T1 fitted snugly in the S pocket. GTP, represented by ball-and-sticks (coloured by atom type with C – lime green) is further inside the tunnel. It establishes hydrogen-bonds (as green dotted lines) with R225, R268, R270 and D453 (ball-and-stick, coloured by atom type with C – orange). The base of the GTP faces T2 and T3 and is close to Y630 (shown in ball-and-sticks, with C atoms in orange).

B. "Post-interrogation" complex. Cartoon representation of 6nt DNA (shown as in A) and GTP (shown as in A) binding mode in molecule C (red). Residues R225, R268, R270, D453 and Y630 are represented as above, with C atoms coloured dark red. The hydrogen bonds between GTP and the protein (red) are represented as green dotted lines. GTP hydrogen-bonds to D453 and interacts with T2 and T3 nucleotides from the DNA template.

[Prepared with PyMol (DeLano, 2004)]

3.5.6. E491Q-DNA-GTP-Mg^{2+}-Mn^{2+} complex

Crystals of E491Q co-crystallised with a 6nt DNA oligo soaked with 25mM GTP, 5mM Mg^{2+} and 5mM Mn^{2+} revealed different complexes captured in the three non-crystallographically related molecules. In molecule A, there is very weak density in the template tunnel and the GTP molecule faces the thumb region around N523 and base-stacks with Y630 (Fig. 3.14.A). In molecule B, the incoming GTP establishes hydrogen bonds with the key "interrogating" residues (R225, R268 and R270) and D453, one of the catalytic aspartates. The base seems to assume a conformation similar to that observed in the WT initiation complex (Butcher *et al.*, 2001) (Fig. 3.14.B). Furthermore, there is weak density indicating that a second molecule might be present, albeit at very low occupancy. If so, the two GTP molecules would then be in a conformation equivalent to that observed in molecule C. In this case, a "pre-initiation" state has been crystallographically captured (Fig. 3.15.A). Like the WT initiation competent stage, DNA has ratched back away from the S pocket, establishing Watson-Crick base-pair interactions with two incoming GTP molecules and base-stacking with Y630. However, the phosphate backbone of D1 [as defined by Butcher *et al.* (2001)] is displaced towards the protein, some 6Å away from the phosphate backbone of D2. In this conformation, reaction cannot occur, in a situation similar to that observed in the calcium inhibited complex described above and in the reovirus initiation complex (Tao *et al.*, 2002). Moreover, only one divalent cation with low occupancy seems to be found between the two phosphate backbones, indicating that the coordination by the cation that facilitates the nucleophilic attack by O3* of D1 to the O-P bond is not yet established (Fig. 3.15).

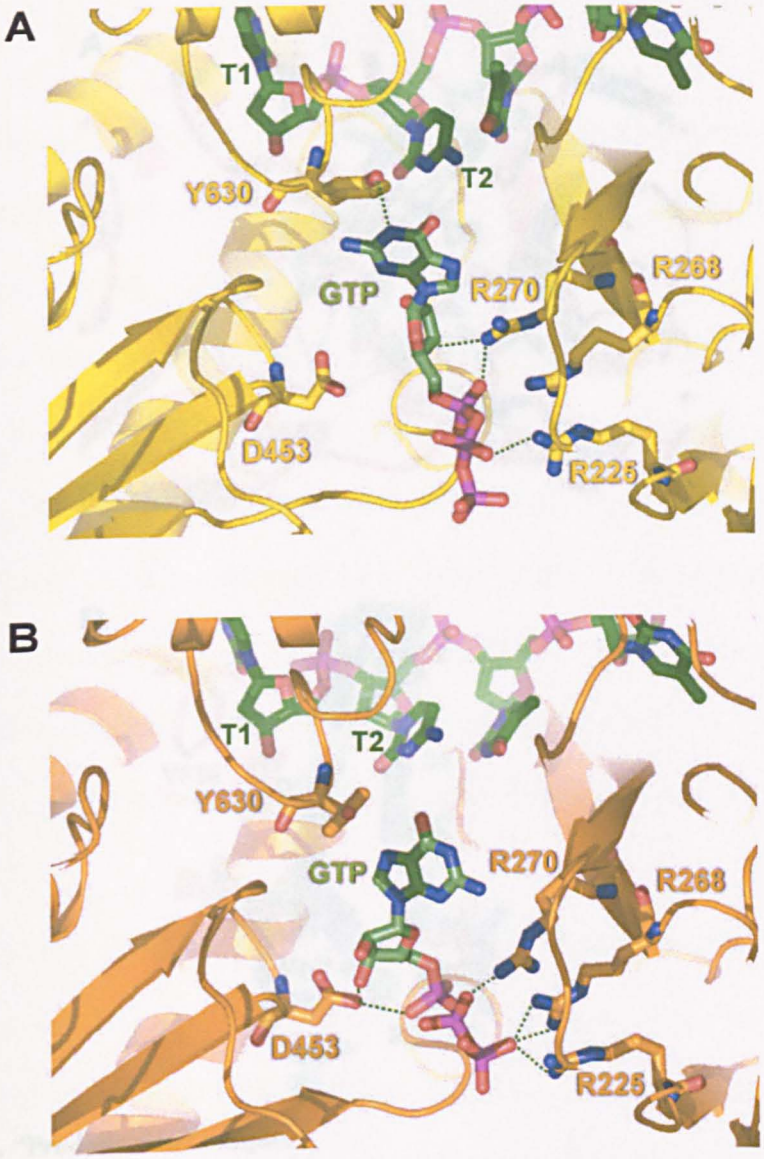


Figure 3.14. **E491Q-DNA-GTP-Mg^{2+}-Mn^{2+} structure**

A. Cartoon representation of the conformation of the complex in molecule A. 6nt DNA (ball-and-stick representation, coloured by atom type: O – red; N – blue; C – green) binds in the template tunnel, with the 3' cytidine T1 fitted snugly in the S pocket. GTP, represented by ball-and-sticks (coloured by atom type with C – lime green) is further inside the tunnel. It establishes hydrogen-bonds (as green dotted lines) with R225, R270 and Y630 (ball-and-stick, coloured by atom type with C – yellow). The base of the GTP molecule faces Y630 that is rotated towards the bases of T2 and T3.

B. Cartoon representation of 6nt DNA (shown as in A) and GTP (shown as in A) binding mode in molecule B (red). Residues R225, R268, R270, D453 and Y630 are represented as above, with C atoms coloured orange. The hydrogen bonds between GTP and the protein (red) are represented as green dotted lines. The base of GTP faces T2 and T3 and is displaced away from Y630. The phosphate backbone and sugar rearrange towards the catalytic D453. View as in Fig. 3.6.A.

[Prepared with PyMol (DeLano, 2004)]

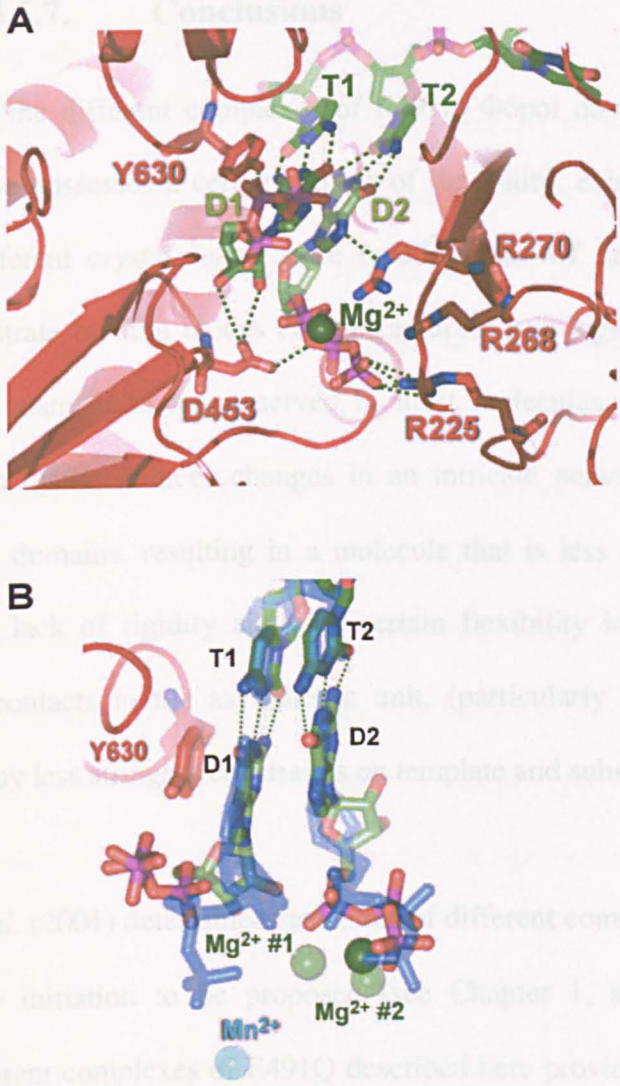


Figure 3.15. "Pre-initiation" complex.

A. Cartoon representation of a pre-initiation stage complex. 6nt DNA (ball-and-sticks representation, coloured by atom type: O – red; N – blue; C – green) has racheted away from site S and base-pairs GTP molecules D1 and D2 (C atoms are coloured lime green). Residues R225, R268, R270, D453 and Y630 are represented as ball-and-sticks, with C atoms coloured dark red. The hydrogen bonds established by D1 and D2 and the protein (dark red) are represented as green dotted lines. D1 hydrogen-bonds to D453 and Y630 and establishes Watson-Crick base-pair interactions with T1. D2 base-pairs with T2 and hydrogen-bonds with R225 and R270. A magnesium ion (dark green sphere) coordinates the phosphate backbone of D2.

B. Superposition of the active initiation structure (Butcher *et al.*, 2001) (semi-transparent; Mg^{2+} coloured green, other atoms marine blue) onto the pre-initiation structure (as in A). Base pairing interactions are shown as dotted lines. Mn^{2+} ion is shown in cyan.

[Prepared with PyMol (DeLano, 2004)]

3.5.7. Conclusions

Analysis of the different complexes of E491Q Φ6pol demonstrates that this mutant polymerase possesses a certain degree of flexibility, evidenced by the fact that it adopts different crystal forms more readily than WT and that a range of template and substrate binding modes can be captured. It is logical to propose that the absence of a manganese ion observed in most molecules, due to the poorer coordination from Q491, induces changes in an intricate network of interactions between different domains, resulting in a molecule that is less rigid than the WT polymerase. This lack of rigidity allows a certain flexibility in the formation of crystallographic contacts in the asymmetric unit, (particularly in the $P3_2$ crystal form) that may play less stringent constraints on template and substrate binding.

Butcher *et al.* (2001) determined structures of different complexes that allowed a mechanism for initiation to be proposed (see Chapter 1, section 1.4.2). The structures of different complexes of E491Q described here provide further structural insight to the several stages leading to the formation of an initiation competent complex of the polymerase. The fact that these intermediate steps were captured with the mutant protein might be due to the more flexible E491Q structure, which allows stabilization of intermediates that in the WT polymerase are quickly directed towards the initiation state. If so, the role of manganese might be to lock the Φ6pol in a more rigid structure that ensures efficient initiation, hence stimulating polymerisation. It must be noted, however, that shorter soaks (less than 2min) were used to obtain E491Q complexes than in the WT experiments (~5min). It is therefore possible that a

combination of flexible structure and short soaks could be responsible for the determination of the intermediate steps.

The stimulatory effect of manganese might also be related to the early stages of the mechanism, namely the rearrangements involved from interrogation to initiation competent states. In the absence of the ion, we were able to obtain post-interrogation and pre-initiation complexes of GTP and template, together with other intermediate stages. One possibility is that manganese accompanies the incoming GTP, facilitating the conformational changes of the base that are necessary to occur for an initiation competent state to be assembled, as has been hypothesised for other polymerases (Arnold *et al.*, 1999; Baulcombe, 2002; Tao *et al.*, 2002; O'Farrell *et al.*, 2003). This would provide a possible general mechanism for the stimulatory effect of manganese in polymerisation. Further studies of Φ6pol and other vRdRPs are necessary to provide a definite mechanism for the manganese stimulation of RNA polymerisation.

3.6. Φ6pol Initiation platform

SG mutant

The mechanism of primer-independent (*de novo*) RNA template polymerization, initiated at the 3' end, has been studied biochemically and structurally for several vRdRPs (Kao *et al.*, 2001; Choi *et al.*, 2004; van Dijk *et al.*, 2004). HCVpol and BVDVpol are capable of *de novo* initiation of RNA synthesis (Hoogstraten *et al.*, 2000; Kim *et al.*, 2000; Ranjith-Kumar *et al.*, 2002a; Ranjith-Kumar *et al.*, 2002b), although, *in vitro,* they preferentially utilize a back-priming initiation mode (Behrens *et al.*, 1996; Zhong *et al.*, 1998; Kim *et al.*, 2000; Lampio *et al.*, 2000). This type of initiation is deleterious *in vivo*, since the newly produced daughter strand remains covalently bound to the template strand (Kao *et al.*, 2001). The stabilisation of the initiation complex (Butcher *et al.*, 2001) by base stacking interaction of the incoming GTP, D1, with Y630, part of the stabilizing platform loop. This primer-mimicking loop, as well as high concentrations of initiatory nucleotides, has been proposed to prevent back-primed initiation in Φ6pol (Laurila *et al.*, 2002). Similar factors have been found to allow HCV polymerase to initiate *in vitro* in a primer-independent mode (Ranjith-Kumar *et al.*, 2003a). In HCV polymerase, a β-hairpin protruding from the thumb domain and a C-terminal hydrophobic pocket form the equivalent proposed stabilizing platform (Hong *et al.*, 2001; Ranjith-Kumar *et al.*, 2003a). A Φ6pol protein form where the "initiation platform" loop 629QYKW632 had been mutated to GSG, was found to be prone to back-primed initiation (Laurila *et al.*, 2002).

Due to low yields of the GSG mutant, a functionally similar mutant was produced for crystallization trials by our collaborators at the University of Helsinki. In this case, the four bulky amino acids 629QYKW632, were changed to small residues, SG, resulting in a crystallisable mutant (as described in Chapter 2).

Overall, the structure of the SG mutant form of the polymerase is similar to the WT structures previously described (Butcher *et al.*, 2001). Furthermore, the catalytic site defined by aspartic acid residues 324, 453 and 454 does not appear to be perturbed, in accordance with the biochemical data (Laurila *et al.*, 2005b), showing that the mode of the initiation, but not catalysis, is affected by the mutation.

Poorly defined density for the residues upstream and downstream of the mutated loop, together with the absence of electron density in the short SG loop region, indicate that disruption of the loop (Fig. 3.16.A) has consequences for the overall structure of the C-terminal sub-domain. Indeed, the ratio of average crystallographic B-factors (a measure of the relative disorder of the protein) between the C-terminal sub-domain and the rest of the protein is 2.3 for SG compared to 1.5 for WT. Furthermore, a calculation of volumes of internal cavities (VOLUMES, Esnouf, unpublished computer program) indicates an expanded cavity in the mutant enzyme, due to a loss of stabilizing interactions. In the WT polymerase, residues 616-619 from the C-terminal sub-domain pack against residues 305-308 from the palm domain, whereas in the mutant these regions shift apart by 1.4Å, creating a slight gap, and destabilizing the whole C-terminal sub-domain. The effect extends to side-chains of contact residues which are rearranged and generally less well defined in the SG mutant (Fig. 3.16.B). It is likely that a major reason for the destabilization

is the loss of the bulky side-chain of W632, which in the WT makes bridging contacts with the 305-308 loop region of the palm domain (Fig. 3.16.B).

Therefore, deletion of the initiation platform has effects on the overall stability of the C-terminal sub-domain. Hence, the initiation platform loop makes important interactions not only with RNA template but also establishes important contacts with the surrounding protein residues. These contacts may help modulate the switch from initiation to elongation.

These results provide evidence to support the view that great care is required when translating the results of studies of back-priming initiation *in vitro* to explain the *in vivo* mechanism of polymerization in other RNA-dependent RNA polymerases. Indeed, observation of this back-priming initiation mode is an artefact in Φ6pol, arising from subtle changes in the polymerase structure by *in vitro* conditions. Furthermore, *in vitro* studies were carried out with a truncated version of HCV polymerase where part of C-terminal sub-domain had been deleted (Behrens *et al.*, 1996; Zhong *et al.*, 1998; Kim *et al.*, 2000; Lampio *et al.*, 2000). This truncation could produce similar disruption of the initiation platform as seen for the SG mutation, explaining its preference for back-primed initiation *in vitro*.

The SG protein structure strengthens the view of the polymerase structure as a switchable enzyme. The primer-independent initiation form is stabilized by a discreet set of contacts, controlling the switch from the initiation to elongation phase. Relatively modest changes in those contacts can have long range effects and result in a premature conformational switching that produces a structure with a preference for back-priming initiation mode.

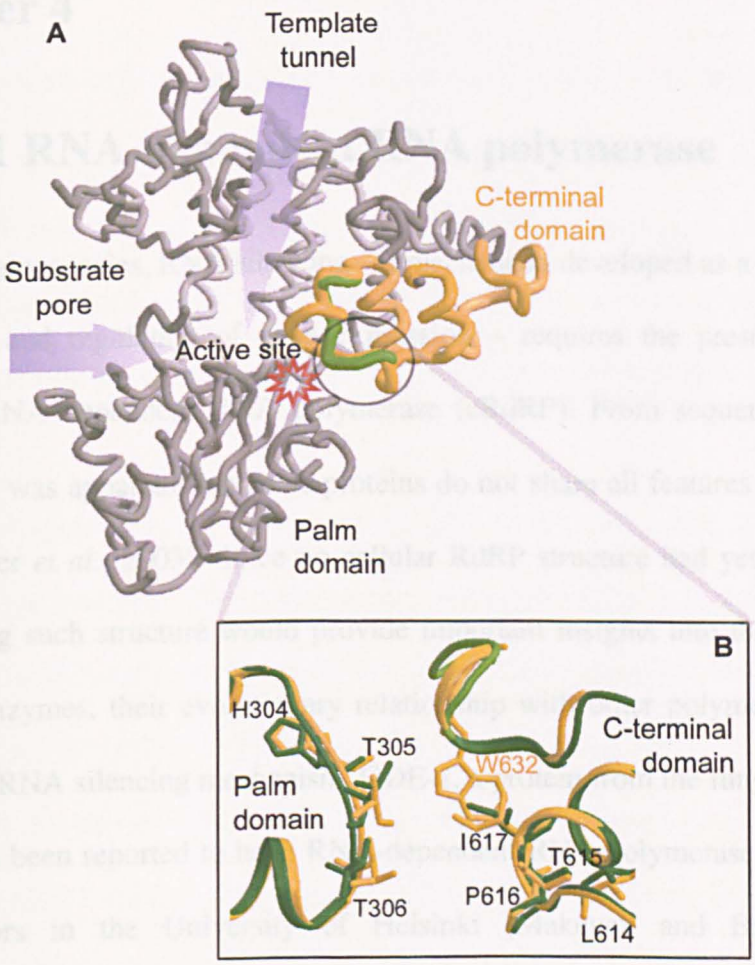


Figure 3.16. **SG mutant structure**

A. Section through the Φ6 polymerase showing the relative position of the C-terminal sub-domain (orange) and the mutated loop 629-632 (lime-green), with respect to the RNA template tunnel and substrate pore. The active site is marked as a red star.

B. A view of key changes at the interface between the C-terminal sub-domain and a loop in the palm domain. Residues in the C-terminal sub-domain of the mutant polymerase (green) are less ordered and the complementarity of hydrophobic surface interactions is less than that observed in the WT polymerase (orange). This destabilization results in a relative movement of 1.4Å between the main chain of residues 305 to 308 in the palm sub-domain and the C-terminal sub-domain.

[From (Laurila *et al.*, 2005b)]

Chapter 4

QDE-1 RNA-dependent RNA polymerase

In many species, RNA silencing – a mechanism developed as a defence against pathogens and regulation of gene expression – requires the presence of a cell-encoded RNA dependent RNA polymerase (cRdRP). From sequence comparison analysis, it was apparent that these proteins do not share all features present in viral RdRPs (Iyer *et al.*, 2003). Since no cellular RdRP structure had yet been reported, determining such structure would provide important insights into the particularities of these enzymes, their evolutionary relationship with other polymerases and their role in the RNA silencing mechanism. QDE-1, a protein from the fungus *Neurospora crassa* had been reported to have RNA-dependent RNA polymerase activity by our collaborators in the University of Helsinki (Makeyev and Bamford, 2002). Furthermore, its catalytic activity was shown to reside in the C-terminal domain (377-1402 aa) – QDE-1 ΔN. The polymerase catalyzes two reactions *in vitro*: production of full-length copies or short 9-21nt copies scattered throughout the entire length of input ssRNA templates. QDE-1 can also extend complementary primers in a template-dependent manner, but this activity was very inefficient under *in vitro* conditions. In this chapter, the determination and analysis of QDE-1 ΔN structure is described.

4.1. Protein expression and purification

QDE-1 ΔN expression and purification were carried out by Minni Koivonen, as part of the collaboration with Prof. Dennis Bamford's group at the University of Helsinki. A summary of this work is included for completeness and details of all procedures are described in Laurila *et al.* (2005). The structural studies were then carried out in Oxford as my contribution to this collaborative project.

To produce QDE-1 ΔN (residues 377-1402), an initial plasmid pEM55 (Makeyev and Bamford, 2002) was constructed from a QDE-1 fragment obtained by PCR-amplification of a wild-type QDE-1 gene from genomic DNA of *Neurospora crassa* DSM 1257 (FGSC 987) and cut with *Hind*III-*Bst*EII and ligated with the large fragment of the similarly cut pEM41. The PCR fragment was then inserted into the vector pYES2/CT (Invitrogen) at the H*ind*III-E*co*RI sites. In an attempt to optimize yields and protein quality, a modified QDE-1 E*co*RI-P*me*I fragment of pEM55 (plasmid pEM69), expressing a recombinant protein that contains a shorter linker (EFGS), tethered to a C-terminal hexahistidine tag, was constructed and introduced into *S. cerevisiae* strain INVSc1 (Invitrogen). Native protein expression was carried out as described by Makeyev and Bamford (2002), according to the manufacturers recommendations.

Since no molecular replacement model was available and all attempts to use heavy atoms to solve the structure were unsuccessful, production of selenomethionated protein was crucial. To express the labelled protein, some modifications to the procedures recommended by the manufacturer (Invitrogen) were necessary, as described by Laurila *et al.* (2005). Due to the toxic effect of SeMet, no

significant increase in turbidity was observed during expression in SeMet containing media, as opposed to the situation observed in the presence of Met, where a rapid increase in turbidity was detected (Fig. 4.1.A).

Protein purification of both native and SeMet proteins was carried at 4°C, as described previously (Makeyev and Bamford, 2002). Storage overnight at 4°C after a Ni-NTA column (Qiagen) caused some precipitation, so additional centrifugation for 10 min at 4,300g and filtration was necessary. As part of the optimization process to obtain crystallisable protein, elution of pooled, protein-containing fractions in a Mono Q column (Amersham Biosciences; RT) was introduced as a final step in the purification of both native and SeMet protein (Laurila *et al.*, 2005). Eluted fractions were analyzed by SDS-PAGE (15% acrylamide) and the protein concentration was determined either by comparing the QDE-1 ΔN band intensities with bands of known amounts of the same protein in SDS-PAGE (15% acrylamide) or by A_{280} measurements (ε_{280}=1.512$M^{-1}cm^{-1}$) in 10mM Tris-HCl (pH 8.0), 100mM NaCl (Fig. 4.1.B). The yield of expressed SeMet QDE-1 ΔN (~0.4 mg of purified protein/L of yeast culture) was comparable to that of the native protein (~2.8 mg of purified protein/L of culture), taking into account that much of the SeMet protein was precipitated after overnight storage.

Replicase activity of labelled and native QDE-1 ΔN protein was assayed as described by Laurila *et al.* (2005). In brief, 10μl reaction mixtures containing 50mM HEPES-KOH (pH 7.8), 20mM NH_4OAc, 6% (w/v) PEG4000, 5mM $MgCl_2$, 0.1mM EDTA, 0.1% Triton X-100, 1mM each of ATP and GTP, 0.2mM each of CTP and UTP, 0.1–0.2mCi/ml of [$^{\alpha 32}$P]UTP (~3000 Ci/mmol) and 0.8 unit/μl Rnasin (ribonuclease inhibitor) and a final concentration of RNA substrate (positive strand

of Φ6 virus small segment - s^+) of 90μg/ml were used. Reactions were initiated by the addition of a polymerase preparation to a final concentration of 4–40μg/ml followed by incubation at 30°C for 1h. The reaction products were then analyzed by native agarose gel-electrophoresis (Fig. 4.1.C). Synthetic ssRNA substrate was prepared by in vitro run-off transcription from pLM659 (Gottlieb *et al.*, 1992) cut with XbaI, as described in Frilander and Turunen (2004).

To estimate the levels of incorporation of selenomethionine in QDE-1 ΔN, a mass spectrometry analysis of both the derivatised and native concentrated proteins was carried out at the Oxford Protein Production Facility by LC-ESI-MS (HPLC: Dionex, Sunnyvale, California, USA, and electrospray: ionization mass-spectroscopy Q-Tof micro, Waters, Milford, Massachusetts, USA) as described by Laurila *et al.* (2005). Due to the considerable molecular weight of QDE-1 ΔN, overnight tryptic digestion was necessary to produce smaller, analyzable fragments. QDE-1 ΔN contains 27 methionine residues spread along 1026 residues, and six methionine-containing digestion products were found scattered throughout the protein. From the average peak heights, the selenomethionine content was estimated to be ~98%. Prior to data collection from a SeMet labelled crystal at beam-line BM14 (ESRF, Grenoble) an X-ray fluorescence analysis was performed around the Se *K* absorption edge. This showed a strong white line (f" = 6.2 e^-) at the absorption peak (Fig. 4.1.D), conforming that high SeMet incorporation had been achieved.

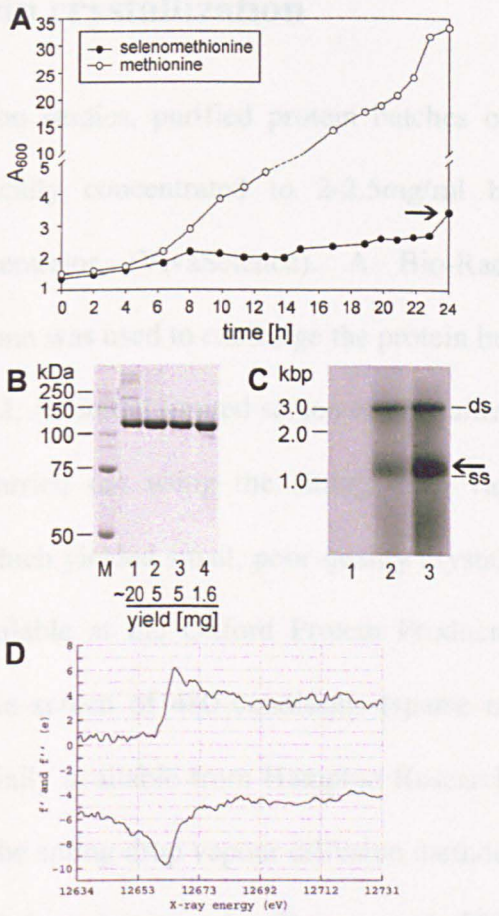


Figure 4.1. QDE-1 ΔN SeMet production

A. Growth curves of *S. cerevisiae* INVSc1/pEM69 in SeMet and in Met containing media after induction and addition of SeMet/Met. Cells harvesting time point indicated by arrow.

B. SDS-PAGE of the QDE-1 ΔN purification steps. 1 - Ni-NTA, 2 - Heparin agarose, 3 - gel filtration, 4 - Mono Q columns. Molecular masses (kDa) markers are shown on the left. Total yield of SeMet QDE-1 ΔN after each purification step is indicated at the bottom.

C. Native agarose gel-electrophoresis analysis of purified SeMet and native QDE-1 ΔN activities. dsDNA markers (kbp) sizes are marked on the left. Lanes: 1 - no QDE-1 ΔN; 2 - QDE-1 ΔN; 3 - SeMet QDE-1 ΔN. ss and dsRNA are shown on the right.

D. Se *K* edge fluorescence scan of selenomethionine QDE-1 ΔN crystal performed at BM14. The plot shown is the result of CHOOCH (Evans and Pettifer, 2001) analysis of the raw data.

[From (Laurila *et al.*, 2005)]

4.2. Protein crystallization

For crystallization studies, purified protein batches of QDE-1 ΔN (residues 377-1401) were typically concentrated to 2-2.5mg/ml by centrifugation in a VivaSpin 4ml concentrator (VivaScience). A Bio-Rad Micro Bio-Spin 6 Chromatography Column was used to exchange the protein buffer to 10mM Tris-HCl (pH 8.0), 100mM NaCl. An initial limited screen of crystallization conditions for the native protein was carried out using the sitting drop vapour diffusion method (McPherson, 1982), which yielded small, poor quality crystals. Taking advantage of a Cartesian robot available at the Oxford Protein Production Facility (OPPF), a wider, more exhaustive screen of 480 conditions (sparse matrix and grid screens formulation, commercially available from Hampton Research and Jena BioScience) was carried out using the sitting drop vapour diffusion method with a 200nl drop size (100nl protein plus 100nl precipitant ratio) (Brown *et al.*, 2003; Walter *et al.*, 2003). However, this screen did not yield crystals of better quality than those obtained previously in the limited screen. Therefore, fine screens around the original conditions were carried out, using an extended, more exhaustive version of the optimization set up developed at the OPPF (Walter *et al.*, 2005) (Fig. 4.2.A). In the optimized conditions (100mM Tris-HCl [pH 7.5], 100-200mM NaCl, 6-14% PEG 6000, 5mM $MgCl_2$) crystals grew to 100x40x40 μm within 1-3 days (Fig. 4.2.B). Initial diffraction tests revealed that these crystals were highly anisotropic and diffracted poorly, so further optimization included an additive screen. The addition of 10mM spermine to the protein solution gave the best results, with crystal size of 195x85x75 μm (Fig. 4.2.C) when the set-up was scaled to 2μl - 3μl drops (1:1, 1:2 or

2:1 protein/precipitant ratio) in 24-well plates. An initial wide screen was also performed for the selenomethionated QDE-1 ΔN with a few conditions allowing crystal growth. Attempts to optimize them with a fine screen grid (Fig. 4.2.A) were then carried out, but no major improvements were observed. Simultaneously, fine screens around the native protein crystallization conditions were also performed and the optimal crystal growth conditions were found to be the same, with 1mM DTT also added to the solution prior to crystallization to avoid selenium oxidation. Crystal size and morphology were similar than those observed for the best native crystals (Fig. 4.2.D). All crystallisation trials were carried out at room temperature (293K). For data collection at 100K temperatures, the crystals were briefly washed in a crystallization solution to which 25% glycerol was added.

A

		-1 pH				pH				+1 pH			
		1	2	3	4	5	6	7	8	9	10	11	12
1:1	F	100%	90%	75%	60%	100%	90%	75%	60%	100%	90%	75%	60%
2:1	G												
1:2	H												

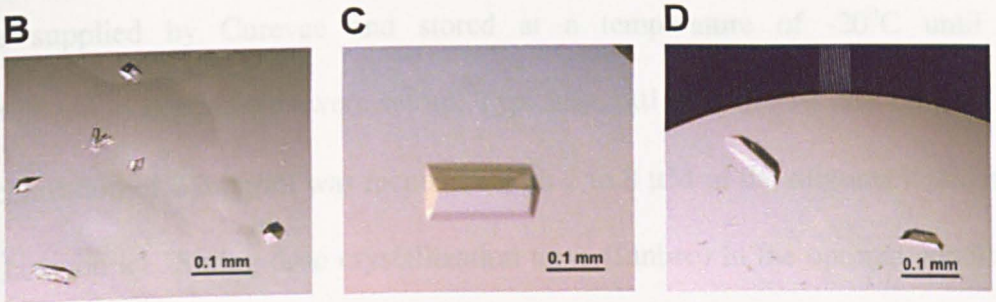


Figure 4.2. QDE-1 ΔN crystallization

A. Schematic optimization grid. The three bottom rows in a 96-well Greiner plate are used, with different protein:precipitant solution ratios (1:1, 2:1, 1:2). Original hit conditions are used at 3 different pHs (-1 pH unit, original pH and +1 pH unit) with a set of dilutions for each pH tested (100%, 90%, 75%, 60%).

B. QDE-1 ΔN Crystal form 1

C. QDE-1 ΔN Crystal form 2

D. QDE-1 ΔN SeMet optimized crystals

4.2.1. Co-crystallization and soaking experiments

Once optimal crystallisation conditions had been established, attempts to co-crystallise QDE-1 ΔN with different RNA oligonucleotides, both single and double stranded were carried out (Table 4.1). Oligonucleotides with the chosen sequence were supplied by Curevac and stored at a temperature of -20°C until co-crystallization experiments were set up. Typically, 6µl of a QDE-1 ΔN solution at a concentration of 2-3mg/ml was incubated with 2 to 8 µM of the oligonucleotides for one hour on ice. Sitting drop crystallization trays (Linbro) in the optimal conditions were then prepared. These experiments yielded either no crystals or very small crystals that, when tested revealed no RNA bound or very low occupancies of the oligonucleotides (as described in section 4.3). In separate experiments, the protein crystals were soaked in solutions containing 10mM RNA oligonucleotides for ~10-15min or until cracking appeared at the surface of the crystals.

The small co-crystals were also used for soaking experiments with a non-hydrolysable GTP analogue [guanosine-5'-(β,γ-methyl)triphosphate] and $MgCl_2$. The crystals were soaked in a droplet of the crystallisation mother liquor to which $MgCl_2$ and the GTP analogue had been added to a final concentration of 10 and 50 mM, respectively. In separate experiments, 10mM RNA oligonucleotide was added to that solution prior to soaking native protein crystals. Total soaking times were around 8-10min or until cracking became apparent in the surface of the crystals.

Table 4.1. RNA oligos used in co-crystallization and/or soaking experiments

	Sequence	Crystals
ssRNA 1	5'-UUUUUUUUUCCC-3'	yes
ssRNA 2	5'-UCUCUCUCUCCC-3'	no
dsRNA 12nt (2nt overhang)	5'-CAAGACGUGUCC-3' 3'-GUUCUGCACAGGCC-3'	no
dsRNA 16nt	5'-GCCGGUUCUGCACAGG-3' 3'-CGGCCAAGACGUGUCC-3'	no
dsRNA 16nt (6nt overhang)	5'-GCCGGUUCUGCACAGG-3' 3'-CGGCCAAGACGUGUCCUUCCCC-3'	yes
dsRNA 16nt (10nt overhang)	5'-GCCGGUUCUGCACAGG-3' 3'-CGGCCAAGACGUGUCCCCUUCCUCCC-3'	no
dsRNA 8nt iodinated	5'-ACG (I)U (I)C (I)U CC-3' 3'-UGC A G A GG-5'	only soaking
dsRNA 8nt iodinated (5nt overhang)	5'-ACG (I)U (I)C (I)U CC-3' 3'-UGC A G A GGCUCCC-5'	only soaking

4.3. Data collection and processing

From the various crystallization studies carried out, two different crystal forms have been characterised. A data set to 3.5Å resolution was collected from initial crystals obtained after fine optimization screen. 256 images were collected at 100K on station BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble at a wavelength 0.97Å, with a 1° oscillation between frames over a continuous rotation range and exposure time of 30s. Images were recorded on a MAR 225 CCD detector (MarResearch) set at 240mm from the crystal. These crystals belonged to space group *C*2, with unit cell dimensions a=114.6, b=124.0, c=101.9 Å, β=108.9°. Data were processed and scaled using the HKL2000 suite of programs (Otwinowski, 1997). A summary of data collection statistics is given in Table 4.3. Cell content analysis using the CCP4 program suite (Collaborative Computational Project, 1994) revealed a Matthews coefficient (V_m) (Matthews, 1968) of 3.5Å^3/Da, consistent with one QDE-1 ΔN monomer in the asymmetric unit and a solvent content of 64%.

A second crystal form was obtained after improving the crystal quality by addition of 10mM of spermine. X-ray data collection of 312 images to 2.3Å resolution at 100K was carried out on station BM14, ESRF, Grenoble using a Marmosaic 225 CCD detector, at a wavelength of 1.00Å (Fig 4.3.A). Oscillation ranges, detector to crystal distance and exposure times were the same as used to collect the *C*2 data set. Improved quality crystals belonged to space group $P2_1$ with unit cell dimension a=101.0, b=122.5, c=114.7 Å, β=108.9°. Data were scaled and processed with HKL2000 (Otwinowski, 1997) and details are presented in Table 4.1. These unit cell dimensions are consistent with the presence of two monomeric

subunits of QDE-1 ΔN in the asymmetric unit, with a solvent content of 55% (V_m=2.8 Å^3/Da). Native Patterson maps calculated using the CCP4 program suite (Collaborative Computational Project, 1994) show a peak with 1/4 of the height of the origin peak at u=0.07, v=0.50, w=0.54 (Fig. 4.3.B). This indicates that the two molecules present in the asymmetric unit are related by a translation. This observation is also consistent with the presence of a local molecular two-fold axis slightly misaligned from the crystallographic 2_1 screw axis, passing through the point x=0.035, y=0.000, z=0.270.

Since the protein is expressed in yeast and, at this point in the project, SeMet incorporation was known to be a difficult procedure (Bushnell *et al.*, 2001; Larsson *et al.*, 2002; 2003), yielding very low incorporation levels, attempts to solve the structure of QDE-1 ΔN using Multiple Isomorphous Replacement (MIR) were carried out. Co-crystallization and soaking experiments with most of the usual and some more unusual heavy atom compounds (several mercury, gold, platinum, osmium, lead, and tantalum complexes, iodine, bromide, manganese) were carried out with no success (Table 4.2). A data set to approximately 4Å was collected from a tantalum containing crystal but low resolution combined with the pseudo-symmetry meant the derived phases were poor. For all other compounds, either there was no heavy atom in the crystal or the diffusion of the heavy atom destroyed the crystal lattice. It was, therefore, necessary to produce selenomethionated protein.

Table 4.2. Heavy-atom derivatives experiments

Heavy-Atom (HA) derivative	Conc.	Soak time	Diffraction/ Resl. (Å)	Data collected	Analysis
EMP (Hg)	1mM	o/n	no	-	-
EMP (Hg)	1mM	6h	~6-7	180 images	no HA present
EMP (Hg)	1mM	1h	~8	no	-
TAMM (Hg)	1mM	o/n	no	-	-
TAMM (Hg)	1mM	1h	no	-	-
Thimerosal (Hg)	2mM	o/n	no	-	-
Thimerosal (Hg)	2mM	6h	no	-	-
Tantalum cluster	50mM	2 days	~6.2	-	-
Tantalum cluster	50mM	o/n	~6.0	50 images	HA present, resl. too weak; incomplete data
Tantalum cluster	50mM	1h	~3.8	166 images	signal/noise ratio <1.5σ no solution due to pseudo-symmetry and low resolution
Bromide	1M	30-40s	~3.2	299 images	no HA present
Osmium cluster	50mM	2 days	no	-	-
Osmium cluster	50mM	o/n	no	-	-
Lead	5mM	o/nt	no	-	-
Gold	5mM	8h	no	-	-
Platinum	2mM	8h	no	-	-

EMP – Ethyl Mercury Phosphate [$(C_2H_5HgO)HPO_2$]; **TAMM** – Tetrakis(acetoximercuri)methane [$C(HgOOCH_3)_4$]; **Thimerosal** - Ethylmercurichiosalicylic acid, sodium salt; **Tantalum cluster** – tantalum bromide[Ta_6Br_{12}] ; **Bromide** – sodium bromide [NaBr]; **Osmium cluster** – Potassium hexachloroosmate (IV) [K_2OsCl_3]; **Lead** – Lead (II) acetate trihydrate [$Pb(CH_3COO)_2$]; **Gold** – Gold cyanide [AuCN]; **Platinum** – Platinum chloride [$PtCl_2$].

o/n - overnight

By careful experimental design it was possible, as described above, to obtain material with a very high incorporation level of SeMet which was successfully crystallised and a single SeMet QDE-1 ΔN crystal (Fig. 4.3.C), belonging to space group $P2_1$ was analysed on station BM14, ESRF, Grenoble. A three wavelength (λ_{peak} = 0.9789, $\lambda_{inflection}$ = 0.9791, λ_{remote} = 0.9078) MAD data set to a resolution better than 3.2Å was collected. 812 images were collected for the peak data and 360 for each of the other wavelengths, with an oscillation of φ=1° and exposure time of 60 seconds per image (Fig. 4.4.D). Images were recorded on a Marmosaic 225 CCD detector, set at 350mm for collection of the peak and the inflection data and 380mm for the remote data set. Each data set was processed and scaled independently using the HKL2000 suite of programs (Otwinowski and Minor, 1997). A summary of the data collection and processing statistics is given in Table 4.3. The anomalous signal-to-noise ratio, as calculated by XPREP (Bruker, 2001), is 2.4, 1.3 and 1.2 in the low resolution shell (30.00 - 8.00 Å) and 1.3, 1.1 and 1.1 to 3.2 Å for the peak data, high-energy remote and inflection data, respectively (Table 4.4).

X-ray data were collected from different QDE-1 ΔN co-crystallised with RNA oligonucleotides and native crystals soaked with RNA oligos and/or GTP analogue and Mg^{2+} (details in Table 4.5). Some of these crystals diffracted to very low resolution and exhibited high anisotropy and mosaicity, not allowing for indexing and scaling of the data collected. Although in some of the datasets there are features in the |Fo|-|Fc| electron density maps at the RNA binding regions and NTP tunnel, these peaks have very low intensity, indicating a low occupancy of the template and substrate. Combined with the low resolution of the data sets, no modelling of the RNA oligonucleotides or GTP analogue was therefore possible.

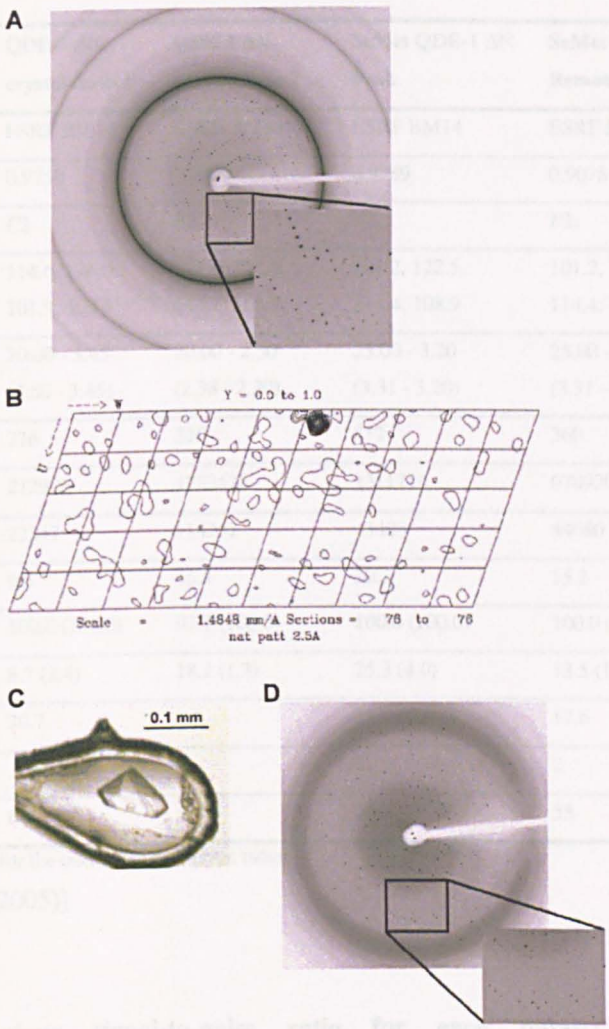


Figure 4.3. QDE-1 ΔN data collection

A. Representative 1° oscillation image data collected from a native QDE-1 ΔN $P2_1$ crystal on a Marmosaic 225 CCD detector on BM14, ESRF, Grenoble. Resolution at the edge ~ 2.3 Å. Bottom right corner: zoom of the diffraction image.

B. Native Patterson map for crystal form 2 calculated to 2.5 Å. The first contour level is at 1 σ and successive contour levels are at 3,5,7,9,...σ. The peak corresponding to the non-crystallographic translation has a height of 41 σ (~1/4 of the origin peak).

C. SeMet QDE-1 ΔN $P2_1$ crystal mounted in a cryo loop, prepared for data collection on BM14, ESRF, Grenoble.

D. Representative 1° oscillation image data collected from a SeMet QDE-1 ΔN $P2_1$ crystal on a Marmosaic 225 CCD detector on BM14, ESRF, Grenoble. Resolution at the edge ~2.8Å. Bottom right corner: zoom of the diffraction image.

[From (Laurila *et al.*, 2005)]

Table 4.3. Data collection statistics for native and selenomethionine QDE-1 ΔN

	QDE-1 ΔN crystal form 1	QDE-1 ΔN crystal form 2	SeMet QDE-1 ΔN Peak	SeMet QDE-1 ΔN Remote	SeMet QDE-1 ΔN Inflection
X-ray source	ESRF BM14	ESRF BM14	ESRF BM14	ESRF BM14	ESRF BM14
Wavelength (Å)	0.9750	1.0000	0.9789	0.9078	0.9791
Space Group	*C*2	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit Cell (a,b,c [Å]; β[°])	114.6, 124.0, 101.9; 108.9	101.0, 122.6, 114.7; 108.9	101.2, 122.5, 114.4; 108.9	101.2, 122.5, 114.4; 108.9	101.2, 122.5, 114.4; 108.9
Resolution range (Å) *	20.00 - 3.45 (3.60 - 3.45)	20.00 - 2.30 (2.38 - 2.30)	25.00 - 3.20 (3.31 - 3.20)	25.00 - 3.20 (3.31 – 3.20)	25.00 - 3.20 (3.31 - 3.20)
N. of images	276	326	812	360	360
Observations	212925	1752528	1311716	670920	605559
Unique reflections	22247	114252	43488	44080	43716
Redundancy	9.6	15.3	30.2	15.2	13.8
Completeness (%)*	100.0 (100.0)	97.9 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
I/σ(I)*	8.7 (2.4)	18.2 (1.3)	25.3 (4.0)	13.5 (1.8)	14.7 (1.8)
R_{merge} (%)	20.7	9.8	19.1	17.6	16.7
Molec. in A.U.	1	2	2	2	2
Solvent Content (%)	64	55	55	55	55

* Values in parenthesis are for the outermost resolution range

[From (Laurila *et al.*, 2005)]

Table 4.4. Anomalous signal-to-noise ratio for each dataset collected from a selenomethionine QDE-1 ΔN crystal

	Resolution range (Å)											
	30.0- 8.0	8.0 - 6.0	6.0 - 5.0	5.0 - 4.8	4.8 - 4.6	4.6 - 4.4	4.4 - 4.2	4.2 - 4.0	4.0 - 3.8	3.8 - 3.6	3.6 - 3.4	3.4 - 3.2
SeMet QDE-1 ΔN Peak	2.4	1. 8	1.4	1.3	1.3	1.2	1.2	1.2	1.2	1.2	1.3	1.3
SeMet QDE-1 ΔN Remote	1.3	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.1	1.1
SeMet QDE-1 ΔN Inflection	1.2	1.0	1.0	1.0	0.9	1.0	1.0	1.0	1.0	1.1	1.1	1.1

Table 4.5. **Data sets collected from QDE-1 ΔN co-crystallisation and soaking experiments with RNA oligonucleotides**

RNA oligo	GTP anal.	Mg^{2+}	Beam line	Resl. (Å)	Space grp.	Features in e.d. maps
ssRNA 1 (co-crystal)	-	-	ID 14.3			Not processable
ssRNA1 (co-crystal)	50mM	10mM	ID 14.3	4.1	*C*2	Very weak density in RNA tunnel
ssRNA 1 (soak)	-	10mM	ID 14.1	4.0	$P2_1$	Not clear
dsRNA 12nt (2nt overhang) (soak)	50mM	10mM	ID 14.1	2.7	$P2_1$	Not clear
dsRNA 12nt (2nt o/h) (soak)	50mM	10mM	In-house	4.0	$P2_1$	Very weak density in RNA tunnel
dsRNA 16nt (10nt overhang) (soak)	50mM	10mM	ID 14.2	4.0	*C*2	Very weak density in RNA tunnel
dsRNA 16nt (10nt overhang) (soak)	50mM	10mM	BM14			Not processable
dsRNA 16nt (6nt overhang) (soak) (co-crystal)	-	-	BM14			Not processable
dsRNA 8nt iodinated (5nt overhang) (soak)	50mM	10mM	BM14			Not processable

4.4. Structure determination

The three MAD data sets were scaled and merged together using SCALEPACK from the HKL2000 suit of programs (Otwinowski and Minor, 1997) and used by the dual-space direct-methods procedure Shake'n'Bake (SnB) (Weeks and Miller, 1999) to solve the selenium substructure. Despite the fact that SnB clearly found a selenium substructure solution as indicated by the R_{min} values for the best trial (see histogram of distribution of trials vs. R_{min}, Fig. 4.4.A), there was not a clear differentiation between correct Se positions and those due to noise, as evaluated from the peak height of each site found. Therefore, identification of the correct Se sites involved a more elaborate strategy. QDE-1 ΔN has 27 methionines, so 54 Se sites were expected in the asymmetric unit. Initially, the top 54 peaks identified by SnB were used in SOLVE (Terwilliger and Berendzen, 1999; Terwilliger, 2000) to calculate experimental electron density maps which showed clear features consistent with protein structure, confirming a correct substructure had been found (Fig. 4.4.B). Since a non-crystallographic two-fold axis relating the two molecules in the asymmetric unit had been identified (Fig. 4.3.B), a subset of 27 sites was identified as symmetry related by visual inspection of the refined 54 Se sites and the electron density maps calculated by SOLVE (Terwilliger and Berendzen, 1999). The non-crystallographic operation was then applied to the selected 27 sites in order to generate all 54 expected Se sites. To verify that all these coordinates corresponded to correct Se positions, related by the two-fold axis, anomalous difference Fourier maps were calculated using the peak and the native 2.3Å data. Visual inspection of the maps allowed the identification of a set of 46 Se atoms. Heavy atom refinement of

these positions was done in SHARP (de la Fortelle and Bricogne, 1997) and experimental phases calculated. Density modification procedures implemented in RESOLVE (Terwilliger, 2000) were then used to refine those experimental phases and build an initial Cα model. RESOLVE (Terwilliger, 2000) built 1128 residues out of the 2052 in the asymmetric unit, mostly as alanines (740). Despite only half the residues being built, the quality of the electron density maps was considerably improved and protein structural features well defined. Furthermore, at least in the well ordered regions, the model built by RESOLVE (Terwilliger, 2000) correctly followed those features (Fig. 4.4.C) and was a good starting point for manual model building. Analysis of the electron density maps and Se positions, together with the poly-alanine model built by RESOLVE (Terwilliger, 2000) revealed two distinct subsets of atoms in each molecule in the asymmetric unit, related by two slightly shifted two-fold axes. This indicates that different domains from each monomer are related by slightly different non-crystallographic symmetry operations which meant no averaging was possible and model building had to be carried for the complete dimer.

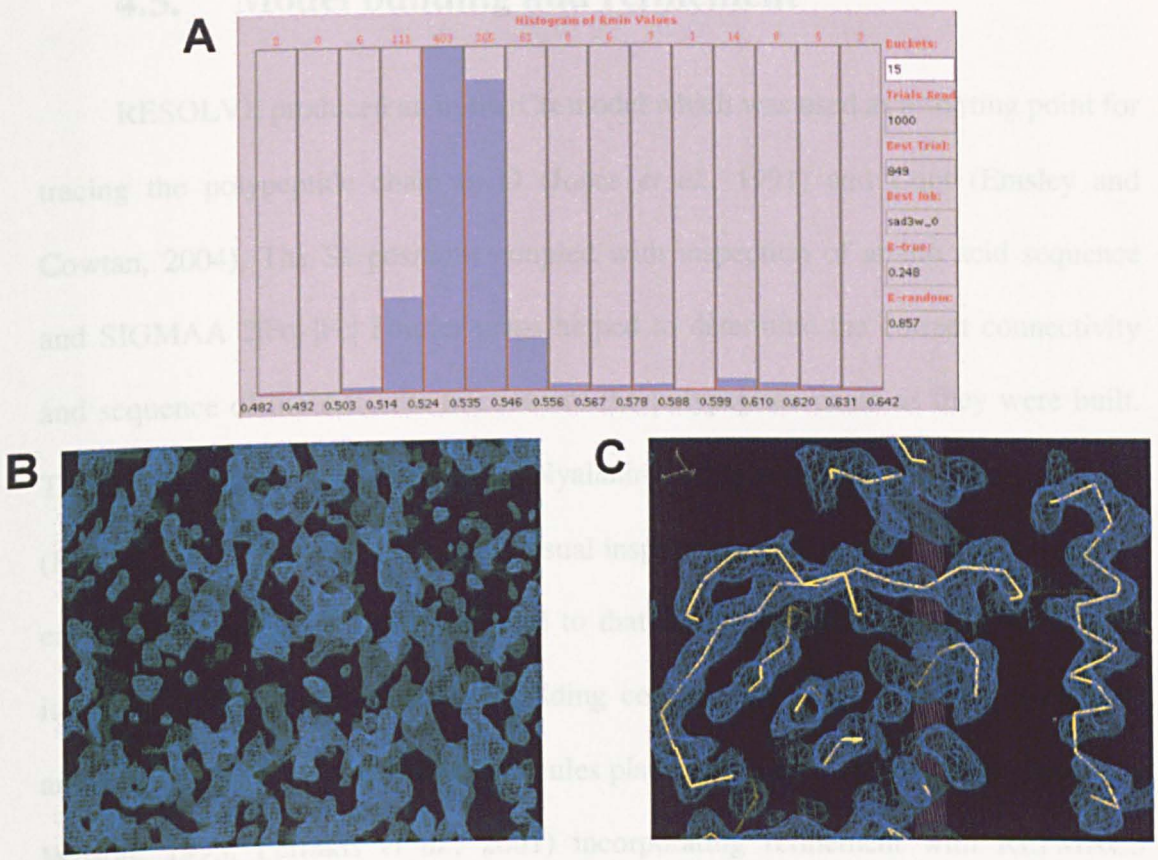


Figure 4.4. Structure determination of QDE-1 ΔN

A. Histogram of SnB trials vs R_{min}, showing a set of correct solutions at R_{min}=0.482.

B. Electron density maps calculated by SOLVE (Terwilliger and Berendzen, 1999; Terwilliger, 2000) after initial phase calculation (blue chicken wire), contoured at 1.5σ. Clear protein features can be identified, as well as the overall shape of the molecule. Snapshot from Coot (Emsley and Cowtan, 2004).

C. Initial RESOLVE (Terwilliger, 2000) model (yellow lines) and electron density map, contoured at 1.5σ (blue chicken wire). An α-helix and several β-strands have been correctly traced in the model built by RESOLVE (Terwilliger, 2000). Snapshot from Coot (Emsley and Cowtan, 2004).

4.5. Model building and refinement

RESOLVE produced an initial Cα model which was used as a starting point for tracing the polypeptide chain in O (Jones *et al.*, 1991) and Coot (Emsley and Cowtan, 2004). The Se positions coupled with inspection of amino acid sequence and SIGMAA 2|Fo|-|Fc| Fourier maps helped to determine the correct connectivity and sequence of the different fractions of the polypeptide chain as they were built. The Cα model was converted to a polyalanine chain using the program CALPHA (Esnouf, 1997) (Fig. 4.5.A,B). After visual inspection and correction of any observed errors, side-chains were then assigned to that model also using CALPHA. Several iterative cycles of manual model building coupled with sequence assignment and automated model building/water molecules placement with ARP/wARP (Lamzin and Wilson, 1993; Perrakis *et al.*, 2001) incorporating refinement with REFMAC5 (Collaborative Computational Project, 1994; Murshudov *et al.*, 1997), allowed gradual construction of a model accounting for most of the observed electron density. Initially, positional and isotropic individual B-factor refinement with CNS (Brunger *et al.*, 1998) was used between automated and manual model building cycles to refine the model (Fig. 4.5.C), using all the native data to 2.3Å resolution and imposing restraints to the identified domains in each monomer related by the two misaligned two-fold axis. As the model become more complete, differences between the domains in each monomer in the asymmetric unit became more apparent and ESCET (Schneider, 2004) was used to define precise limits for each domain (Fig. 4.5.E). The final refinement stages were carried out with REFMAC5, introducing TLS refinement (Winn *et al.*, 2001) and using non-crystallographic restraints for the

different domains. The final model with 933 residues in monomer A and 930 residues in monomer B (out of 1026 amino acids in the truncated QDE-1 protein monomer) with 922 water molecules, 2 magnesium ions and a glycerol molecule was refined to an R_{factor} of 21.7% and R_{free} of 26.4% with good stereochemistry (rms deviation bond = 0.013Å, rms deviation angle = 1.5Å, 2.3% of residues in disallowed regions of the Ramachandran plot) (Table 4.6; Fig. 4.5.D). Several regions were found to be disordered: 10 residues at the N-terminus, 30 residues at the C-terminal, 45 residues in monomer A and 48 in monomer B belonging to 4 loops (monomer A: residues 590-603, 628-640, 1241-1251, 1271-1281; monomer B: 591-606, 627-640, 1241-1251, 1271-1281) could not be traced in the electron density maps.

The *C*2 crystal form structure was solved by molecular replacement using the $P2_1$ final model with AMORE (Navaza, 1994) with one subunit in the asymmetric unit. Due to the poor resolution and data quality, only rigid body, followed by B-factor refinement with CNS was carried out, resulting in a model with R_{factor} of 36.8% and R_{free} of 38.4%. As expected, in this crystal form, the non-crystallographic two-fold axis aligns with the crystallographic 2_1 screw axis due to rearrangements of the domains in each monomer.

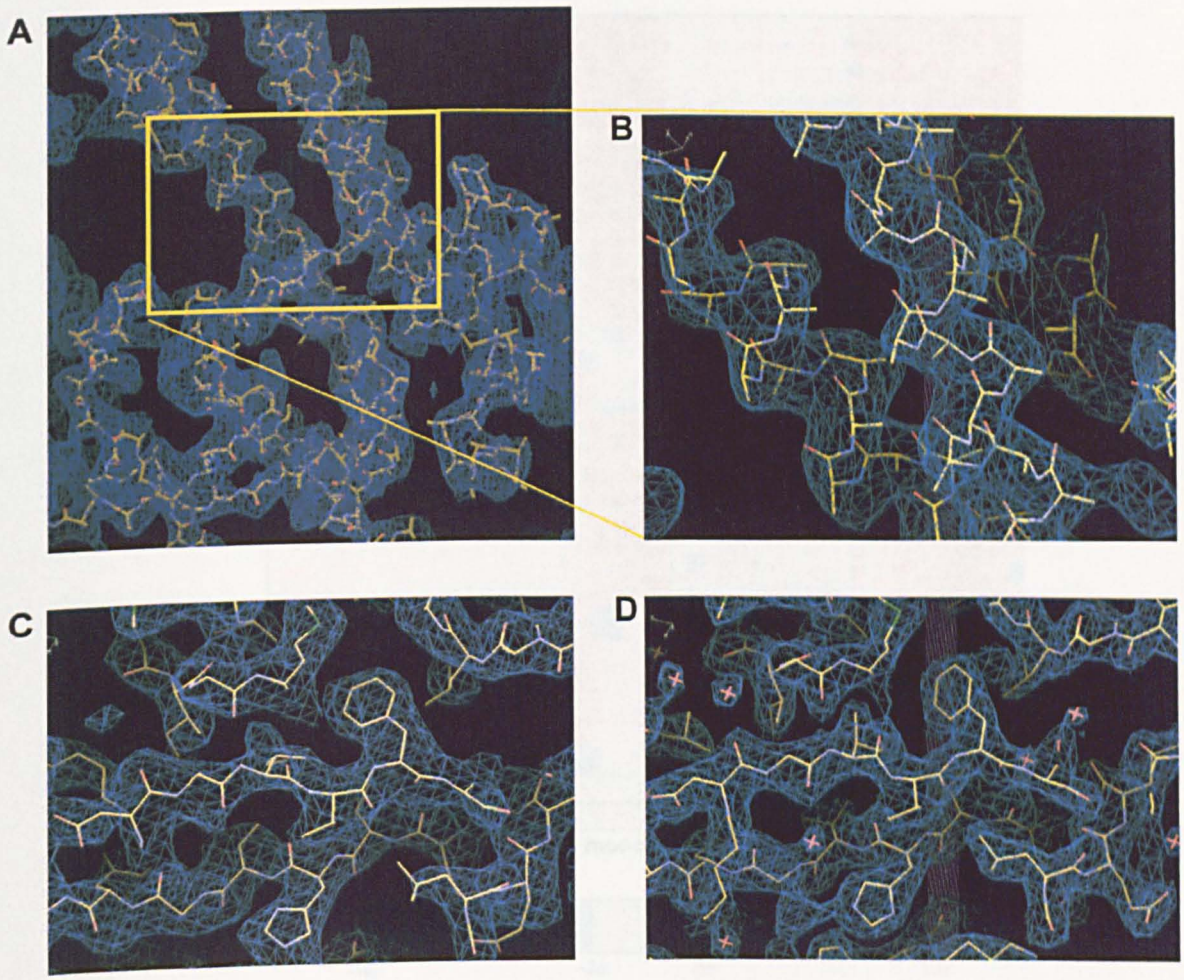


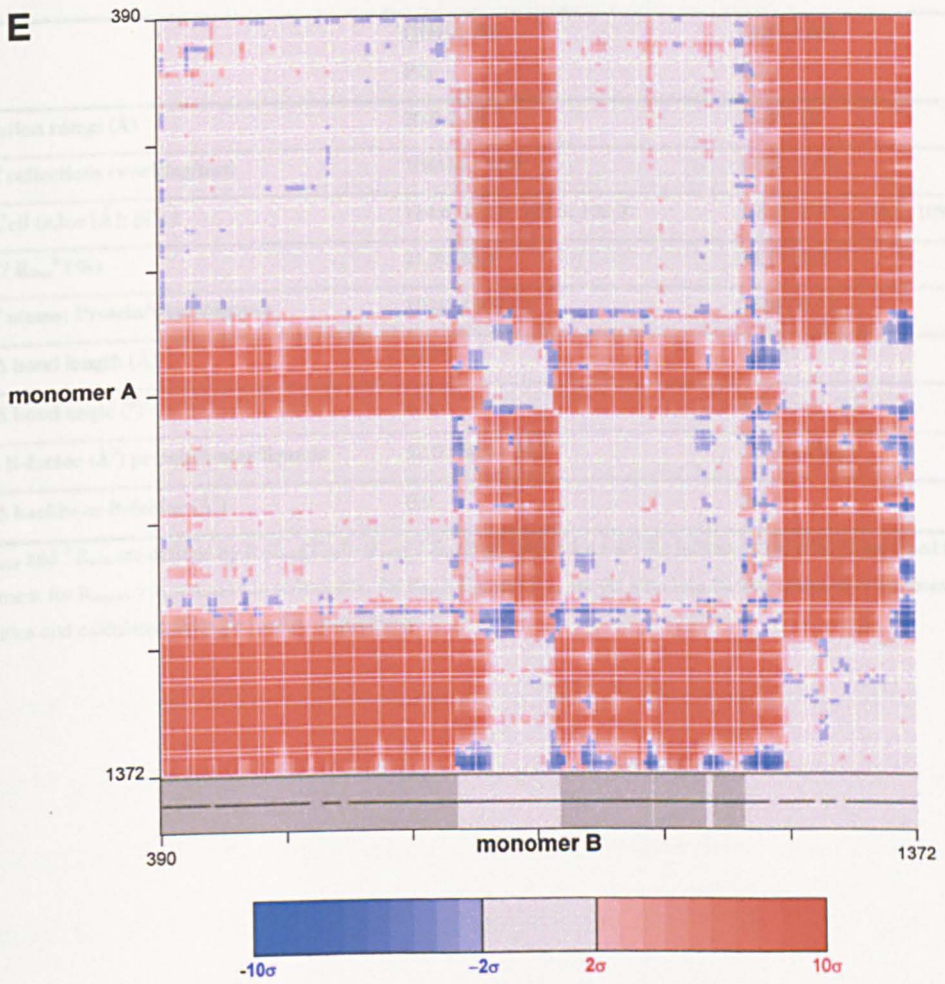
Figure 4.5. Model building steps

A. Initial polyalanine model (C – yellow, O – red, N – blue) and RESOLVE electron density map (blue chicken wire), contoured at 1.5σ.

B. Zoomed view of the initial polyalanine model in the region of Fig. 4.5.B indicated by a yellow box.

C. QDE-1 ΔN model refined with CNS, after several rounds of manual and automated model building. A section containing several well-ordered residues is shown (atoms represented as in A). CNS SIGMAA 2|Fo|-|Fc| map contoured at 1.5σ is represented as in A.

D. Final QDE-1 ΔN model refined. A section containing several well-ordered residues is shown (atoms represented as in A). SIGMAA 2|Fo|-|Fc| map contoured at 1.5σ is represented as in A.



E. ESCET result matrix. Vertical axis correspond to monomer A model, horizontal axis to monomer B. Grey areas represent residues whose positions in the two monomers varies between -2 and 2σ, red areas posititive variations above 2σ and blue regions negative variants of more than -2σ.

Table 4.6. **Refinement statistics of QDE-1 ΔN $P2_1$ and $C2$ crystal forms**

	QDE-1 ΔN $P2_1$	QDE-1 ΔN C2
Resolution range (Å)	20.0-2.30	20.0-3.50
No. of reflections (working/test)	108414 / 5748	14947 / 773
Unit Cell (a,b,c [Å]; β[°])	114.6, 124.0, 101.9; 108.9	101.0, 122.6, 114.7; 108.9
R_{factor}[a]/ R_{free}[b] (%)	21.7 / 26.4	35.3 / 36.0
No. of atoms: Protein/water/ligands	15018 / 922 / 7	7520 / 0 / 1
Rms Δ bond length (Å)	0.013	0.013
Rms Δ bond angle (°)	1.5	1.5
Mean B-factor (Å²) protein/water/ligands	52.0 / 50.4 / 54.3	119.3 / 150.0
Rms Δ backbone B-factor (Å²)	0.9	2.4

* [a] R_{factor} and [b] 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_{factor}; 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

4.6. Structural description and analysis

The QDE-1 ΔN refined model reveals a molecular dimer in the asymmetric unit, with two slightly misaligned non-crystallographic two-fold axes running through the dimer interface, one relating the top third of each monomeric subunit, whilst the second axis relates the lower two thirds. Throughout the following discussions, the standard orientation for the dimeric molecule places these axes in the plane of the page with subunit A represented to the left and subunit B to the right of the axes, as shown in the first panel of Fig. 4.6.A. The residues of each monomer form 41 α-helices and 25 β-strands, creating a four domain protein with a fold previously undescribed for RNA-dependent RNA polymerases (Fig. 4.6.A). The N-terminal residues (390-646) form a mixed α/β "slab" domain leading into a domain (residues 647-807; 914-1162) that includes the proposed three catalytic aspartic acid residues (D1007, D1009, D1011) (Makeyev and Bamford, 2002) and is hence referred to as "catalytic". The "neck" domain, that lies close to the molecular two-fold axis, comprises three long α-helices (residues 808-836; 887-913; 1162-1195) and connects the catalytic domain to the 'head' (residues 837-888; 1196-1372), which is mainly α-helical (Fig. 4.6.A).

The most important tertiary structure feature of the catalytic domain are two six-stranded β barrel motifs (residues 680-782 and 916-1018) (Fig. 4.7.A) that display the typical double-psi topology, with the parallel β strands forming two psi structures (Fig. 4.7.B), as described for different protein superfamilies (Castillo *et al.*, 1999). The three aspartates reside in a loop in double-psi β barrel 2 (DPBB2, residues 916-1018) facing the interface of the two β barrels (DPBB). A peak in the

SIGMAA |Fo|-|Fc| difference Fourier maps at sigma levels of 4.5 has been assigned as a Mg^{2+} ion, coordinated by these aspartates. DPBB1 contributes to the active site cleft with a set of positively charged residues, namely Q736, K743 and K767, establishing a network of hydrogen bonds with water molecules, linking the two DPBB sub-domains. The catalytic domain also contains a separate 'flap' sub-domain (residues 1025-1161), constituted mainly by α–helices, peripheral to the active site cleft.

Most of the inter-monomer contacts responsible for dimerization are established between the heads (3458 $Å^2$ out of the total contact area of 4461 $Å^2$). To exclude the possibility that dimerization is a crystallographic artefact, QDE-1 ΔN protein in solution was analysed by gel filtration and sedimentation assays (see Appendix II.4). The determined molecular mass in both essays was ~230kDa, indicating that it behaves as a dimer in solution (predicted molecular mass for monomeric QDE-1 ΔN is ~120kDa).

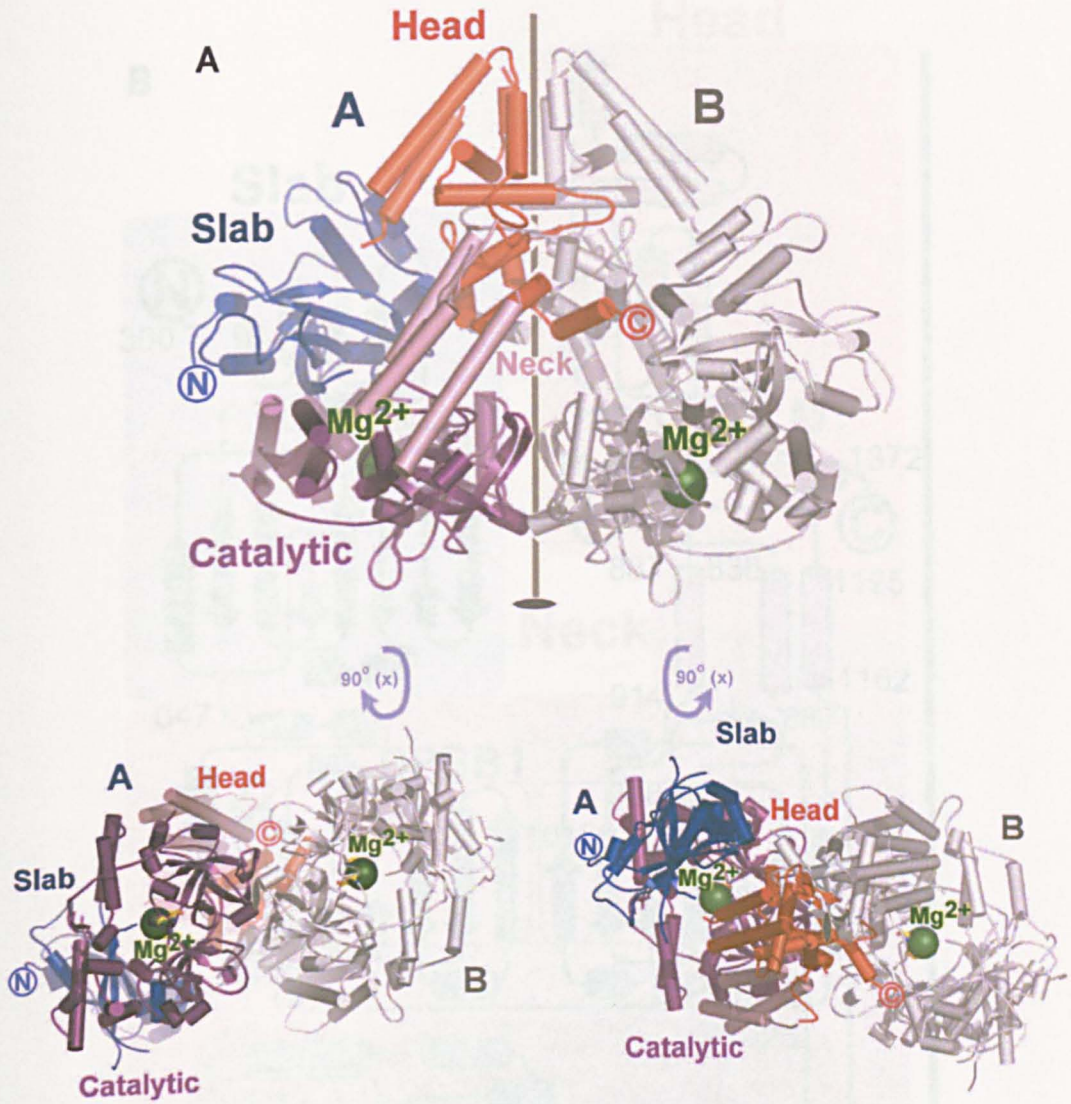
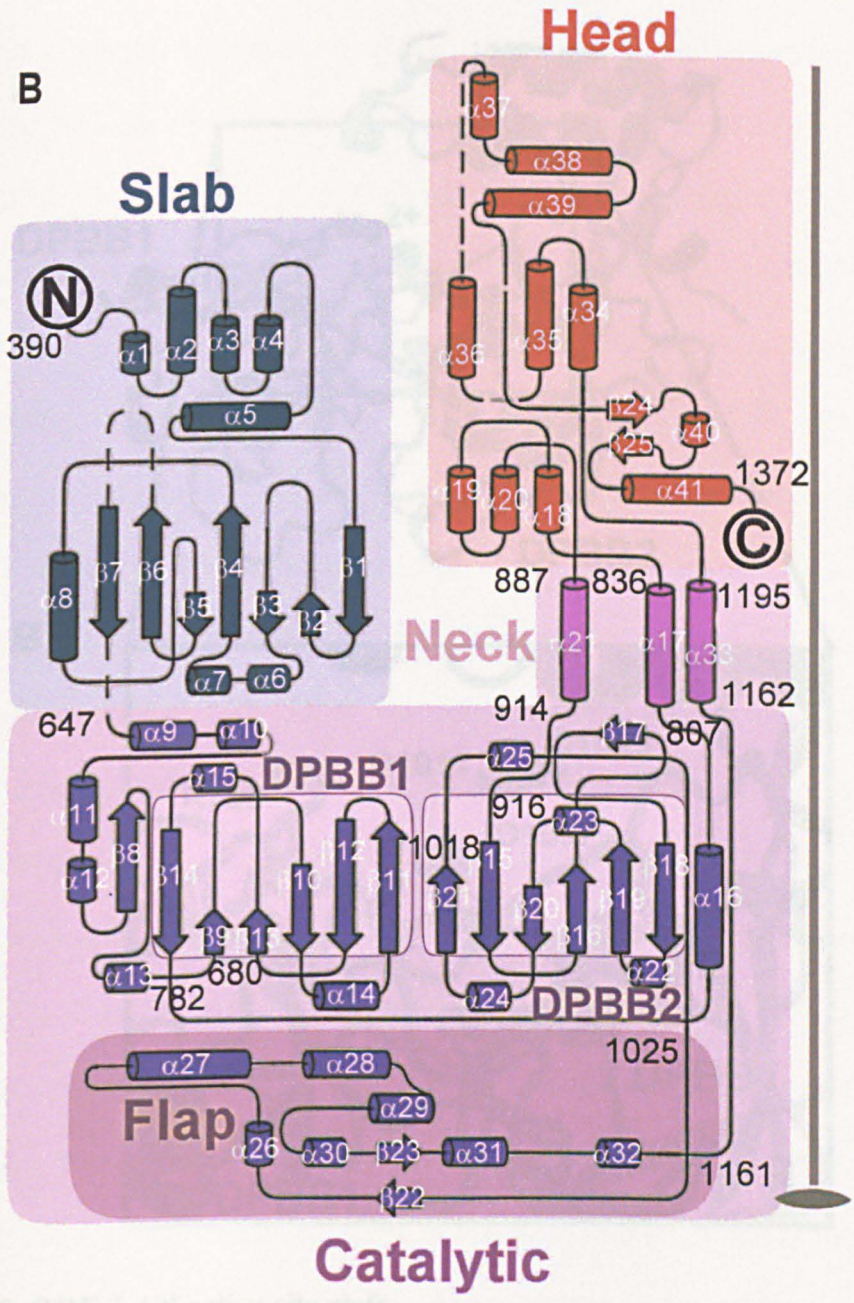


Figure 4.6. The structure of QDE-1 ΔN

A. Cartoon representation of the secondary structure elements in QDE-1 ΔN $P2_1$ homodimer - subunit A coloured according to domains: slab – dark blue (residues 390-646), catalytic - dark purple (residues 647-807 and 914-1161), neck – pink (residues 808-836; 817-913; 1162-1195), head - red (residues 837-888 and 1196-1372); subunit B coloured grey. The non-crystallographic two-fold is represented as a grey line. Two magnesium ions are shown as green spheres. Disordered regions corresponding to 10 residues at the N-terminus, 30 residues at the C-terminal, 45 residues in monomer A and 48 in monomer B belonging to 4 loops (A: residues 590-603, 628-640, 1241-1251, 1271-1281; B: 591-606, 627-640, 1241-1251, 1271-1281) could not be traced in the electron density maps. Top left panel: front view, with two-fold axis aligned on the plane of the page. Top right panel: top view (rotated 90° clockwise in *x* from front view.) Bottom left panel: bottom view (rotated 90° counter-clockwise in *x* from front view)



B. Diagram representation of the topology of QDE-1 ΔN subunit A, coloured as in 4.6.A. Shadowed areas outline each domain boundaries and are coloured in a similar scheme. The catalytic sub-domains DPBB1, DPBB2 and flap are outlined by shadowed boxes. The non-crystallographic 2-fold is represented as in 4.6.A. Disordered loops are represented by dashes lines. Secondary structural elements are numbered white and residues that delimit each sub-domain in black.

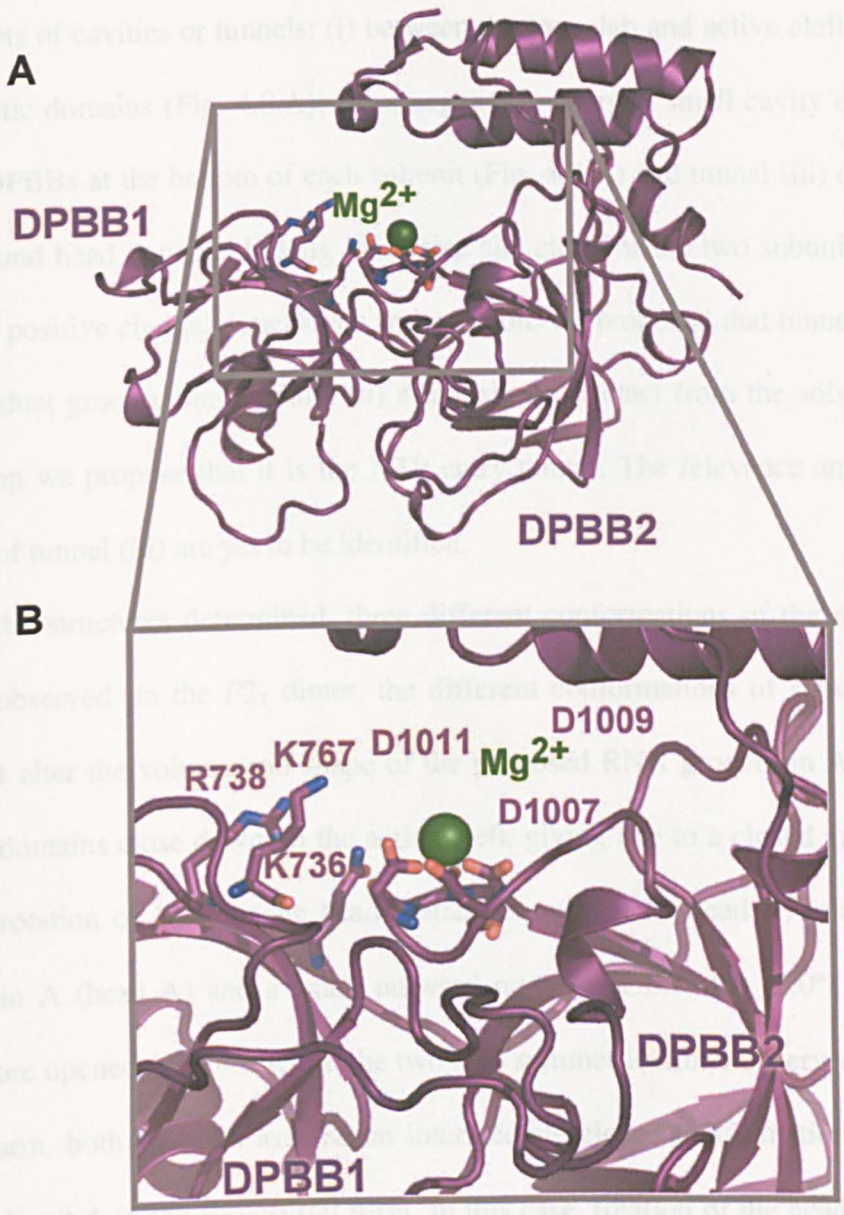


Figure 4.7. QDE-1 ΔN active site cleft

A. Cartoon representation of QDE-1 ΔN homodimer subunit A active site cleft. The two double-psi β-barrels that form the active cleft, DPBB1 (residues 680-782) and DPBB2 (residues 916-1018) are labelled. Catalytic aspartates in DPBB2 and positively charged residues from DPBB1 are shown in ball-and-stick representation (C – dark purple, O – red, N – blue); Mg^{2+} is shown as a green sphere.

B. Zoomed view of active site loop and surrounding residues, showing details of a chain of residues that are likely to be important for polymerisation. Catalytic aspartates and positively charged residues surrounding the active site and Mg^{2+} ions represented as in A.

Analysis of the molecular surface of the dimer allowed identification of three distinct sets of cavities or tunnels: (i) between the top, slab and active cleft region of the catalytic domains (Fig. 4.8.A); (ii) a negatively charged small cavity defined by the two DPBBs at the bottom of each subunit (Fig. 4.8.A) and tunnel (iii) defined by the neck and head domains linking the active site clefts of the two subunits. Due to its highly positive charge, dimensions and position, we proposed that tunnel (i) is the RNA product groove. Since tunnel (ii) establishes a contact from the solvent to the active loop we propose that it is the NTP entry tunnel. The relevance and possible function of tunnel (iii) are yet to be identified.

In the structures determined, three different conformations of the monomeric unit are observed. In the $P2_1$ dimer, the different conformations of subunit A and subunit B alter the volume and shape of the proposed RNA groove. In A, the head and slab domains close down on the active cleft, giving rise to a closed groove. The outward rotation of 11.2° of the head domain in subunit B (head B) relative to its position in A (head A) and a small outward rotation of the slab (2.0°) render the cavity more opened (Fig. 4.9.A). In the two-fold symmetric dimer observed in the *C*2 crystal form, both subunits assume an intermediate closed conformation, closer to that of subunit A in the $P2_1$ crystal form. In this case, rotation of the head domain is less sharp – only 3.8° from the position of head A, compared to 8.3° relative to head B. The slab domain, however, seems to undergo more significant changes in orientation in this crystal form: it is displaced in an upward, slightly outward direction by 3.6° relative to the position of slab A and 2.3° relative to slab B (Fig. 4.9.B).

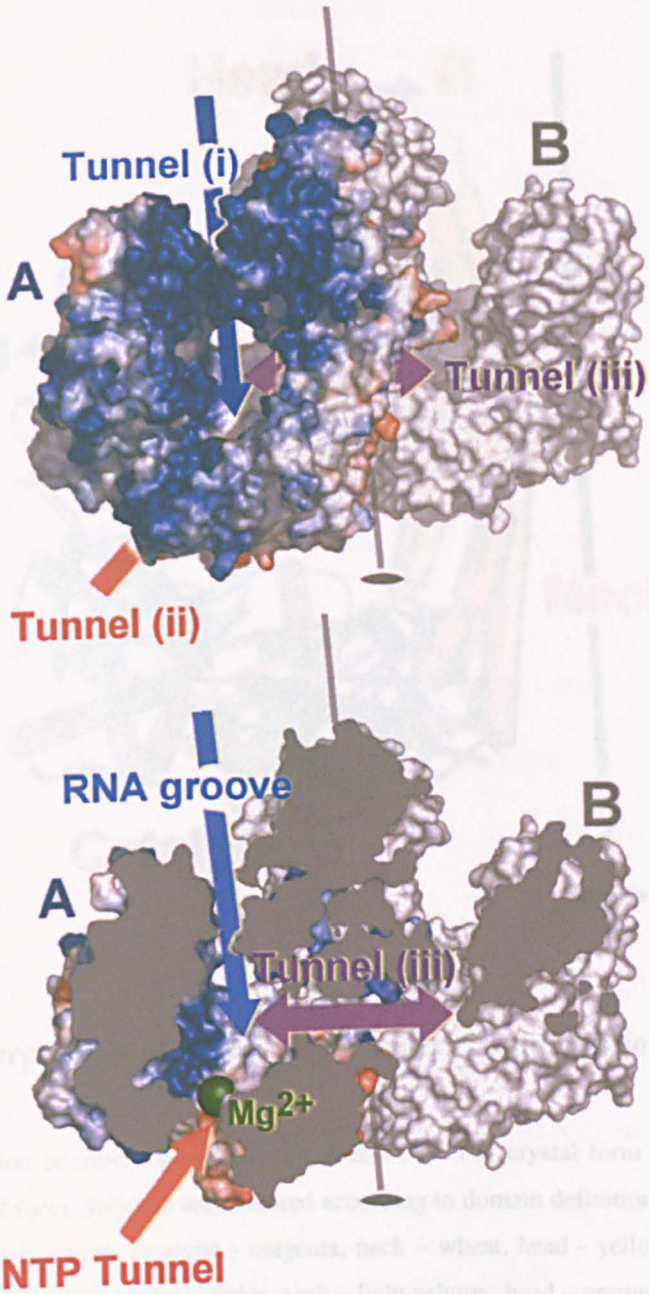


Figure 4.8. Surface charge representation of QDE-1 ΔN

Surface charge representation of QDE-1 ΔN subunit A (subunit B in grey). Blue arrow: tunnel (i), proposed RNA product groove; red arrow: tunnel (ii), proposed NTP tunnel; purple arrow: tunnel (iii) across active sites in dimer. The magnesium ions are shown as green spheres. Top panel: homodimer view as in Fig. 4.6.A (left top panel). Bottom panel: slice through top panel representation showing the three sets of cavities and tunnels.

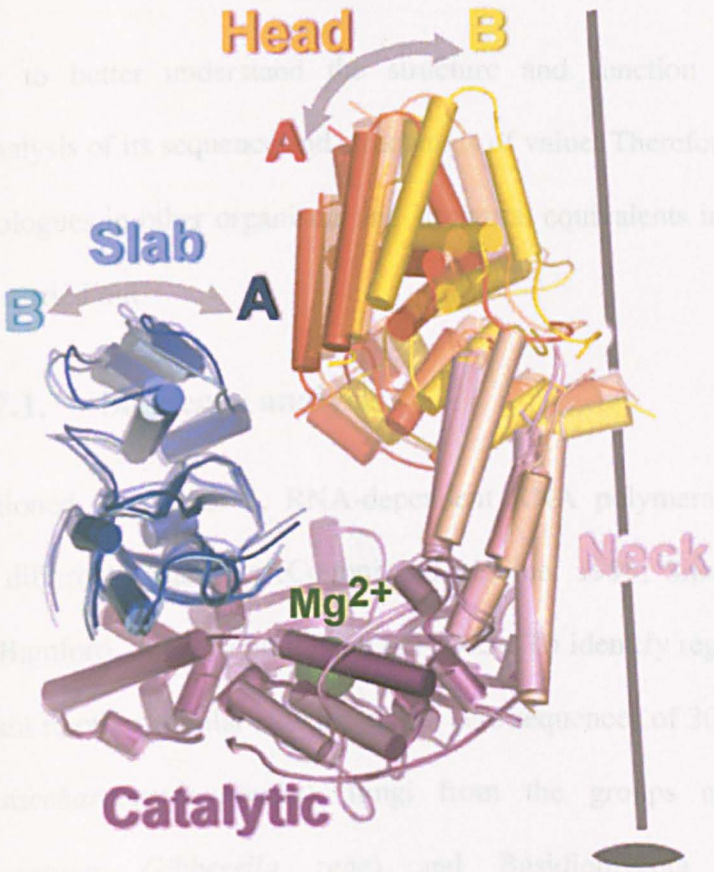


Figure 4.9. Superposition of observed monomeric conformations of QDE-1 ΔN structure

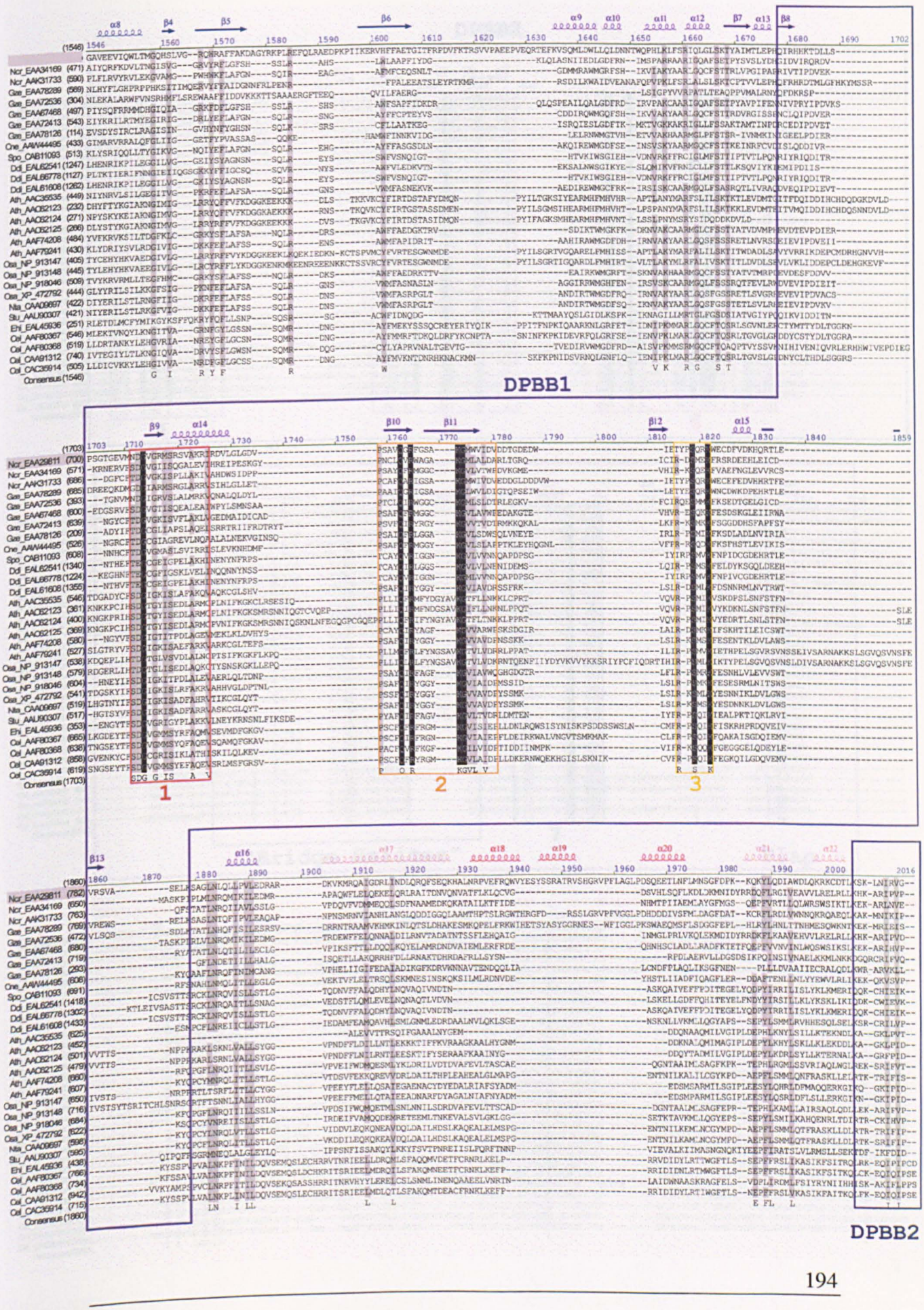
Cartoon representation of superimposed subunit A and B in $P2_1$ crystal form and each monomeric subunit in *C*2 crystal form. Subunits are coloured according to domain definition: subunit A as in Fig. 4.6.A.; subunit B: slab – cyan, catalytic - magenta, neck – wheat, head - yellow; *C*2 subunit (semi-transparent): slab – marine, catalytic – violet, neck – light salmon, head – orange. The direction of the movement of the slab and head domains is indicated by grey arrows and the non-crystallographic two-fold along the dimer interface is represented as in Fig. 4.6.A. Magnesium is represented as a green sphere.

4.7. Bioinformatics analysis

In order to better understand the structure and function of QDE-1, a comparative analysis of its sequence and structure is of value. Therefore, searches for sequence homologues in other organisms and structural equivalents in know protein structures was carried out.

4.7.1. Sequence analysis

As mentioned in chapter 1, RNA-dependent RNA polymerases have been recognised in different organisms (Cogoni and Macino, 1999; Sijen *et al.*, 2001; Makeyev and Bamford, 2002; Schramke *et al.*, 2005). To identify regions in QDE-1 that are invariant in other cellular RdRPs, amino acid sequences of 30 cRdRPs from: yeast (*Schizosaccharomyces pombe*), fungi from the groups of Ascomycota (*Neurospora crassa, Gibberella zeae*) and Basidiomycota (*Cryptococcus neoformans*), slime molds (*Dictyostelium discoideum*), dicot plants (*Arabidopsis thaliana, Solanum tuberosum, Nicotiana tabacum*), monocot plants (*Oryza sativa*), protozoa (*Entamoeba histolytica*) and nematodes (*Caenorhabditis elegans*) were aligned using standard settings of ClustalW algorithm (Thompson *et al.*, 1994). This alignment allowed us to identify seven sequence motifs (Fig. 4.10.A), clustered around the invariant proposed catalytic aspartates. All these sequence motifs map onto the DPBBs in QDE-1, apart from motif 7, which corresponds to α-helices 29 and 30 at the interface of DPBB2 (Fig. 4.10.B). Sequence motifs 1 to 3 belong to the structural motif DPBB1 whilst motifs 4 to 6 form most of DPBB2. These results indicate that, despite their overall low sequence homology, all cell-encoded RdRPs might share a structurally equivalent active site cleft.


DPBB1
DPBB2
1
2
3

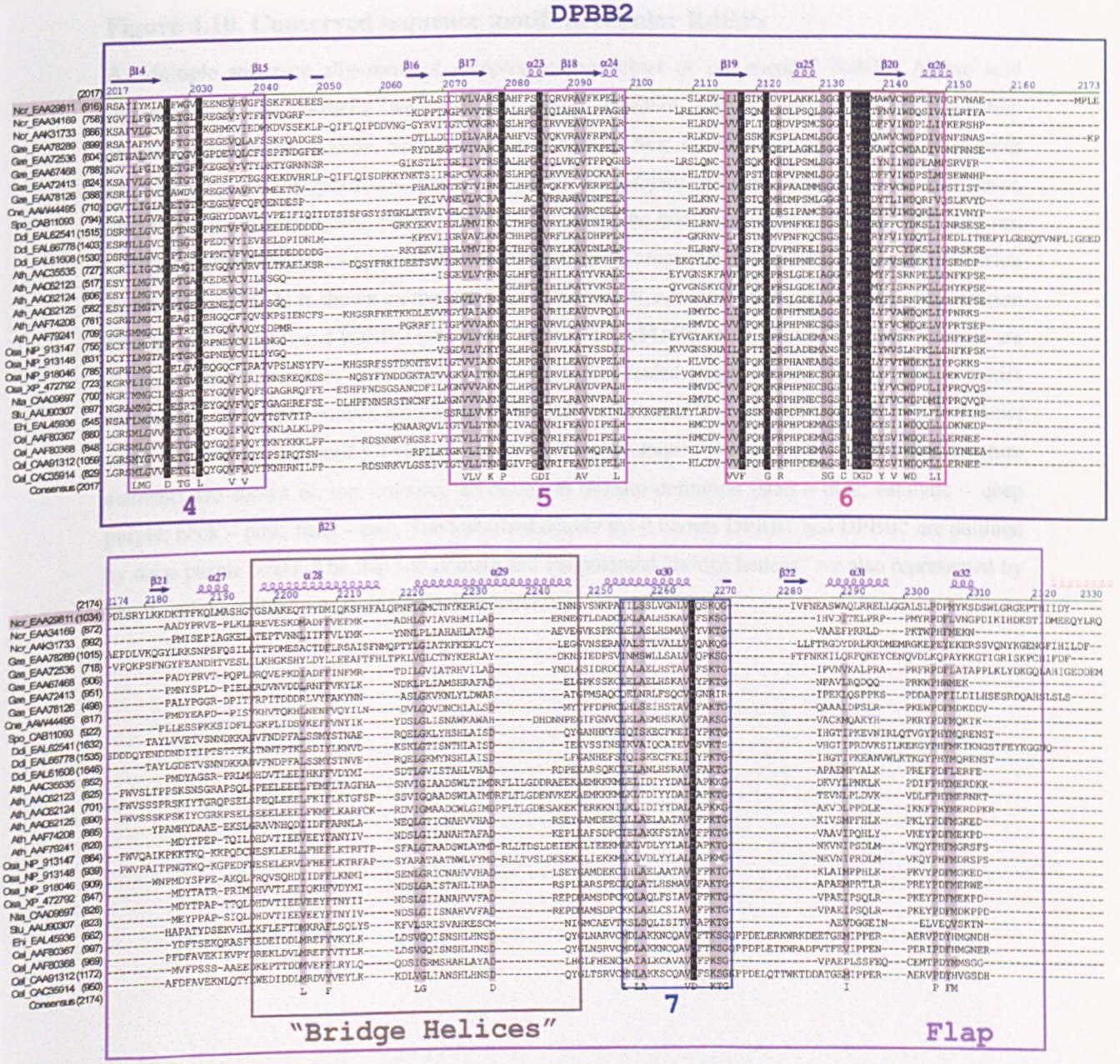
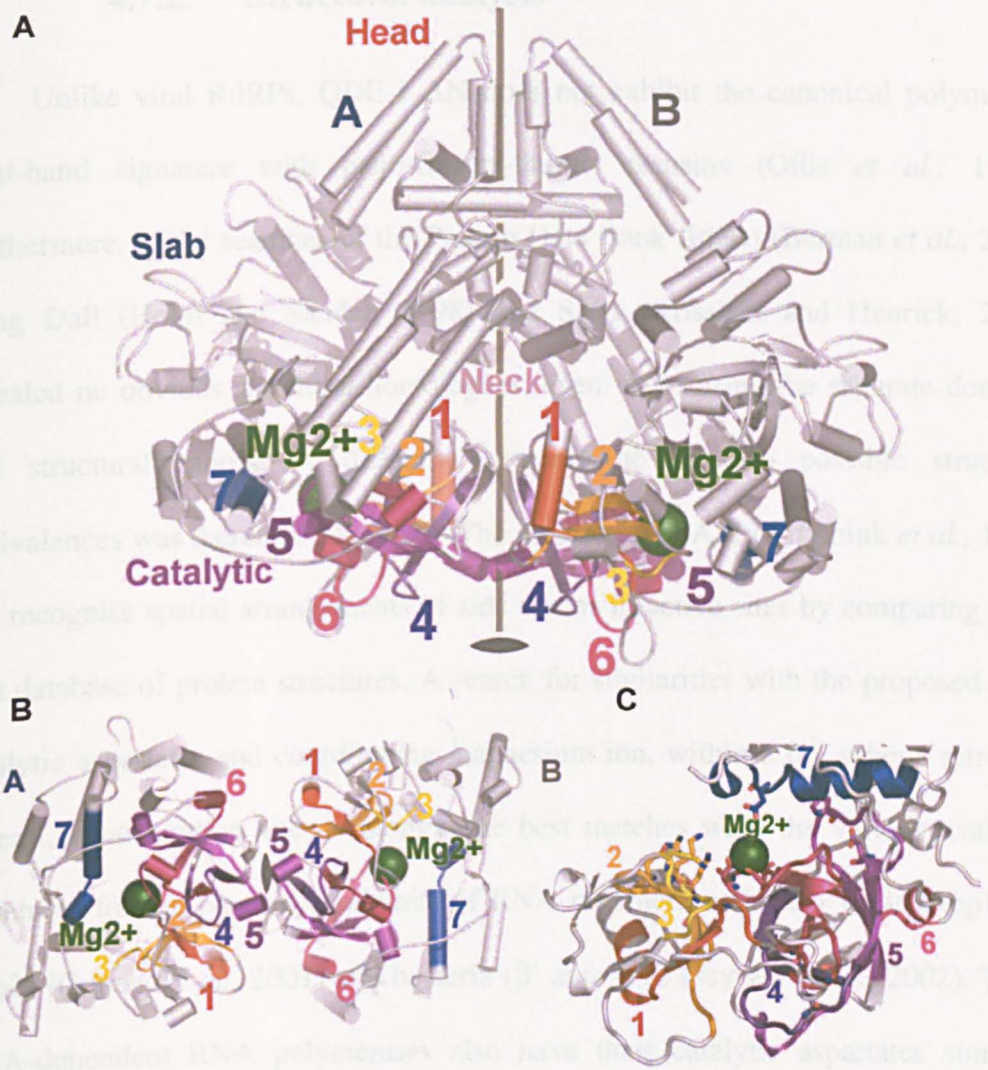

DPBB2
4
5
6
7
"Bridge Helices"
Flap

Figure 4.10. Conserved sequence motifs in cellular RdRPs

A. Multiple sequence alignment of a representative subset of cell-encoded RdRPs. Amino acid sequences of 30 cRdRPs were aligned using standard settings of ClustalW algorithm: *Schizosaccharomyces pombe*, Spo; *Neurospora crassa*, Ncr; *Gibberella zeae*, Gze; *Cryptococcus neoformans*, Cne; *Dictyostelium discoideum*, Ddi; *Arabidopsis thaliana*, Ath; *Solanum tuberosum*, Stu; *Nicotiana tabacum*, Ntu; *Oryza sativa*, Osa; *Entamoeba histolytica*, Ehi; *Caenorhabditis elegans*, Cel Local alignment was improved by manual editing. *N. crassa* QDE-1 protein sequence (accession number EAA29811) is shown on the top. Two additional *N. crassa* cRdRP genes - SAD-1 (accession number AAK31733) and RdRP-3 (accession number EAA34169) are included. Invariant residues are shaded in black; residues with ≥80% conservation are shaded in grey. Conserved sequence motifs comprising invariant residues are outlined: motif 1 – red; motif 2 – orange; motif 3 – dark yellow; motif 4 - dark purple; motif 5 – violet; motif 6 – light pink; motif 7 – blue. QDE-1 secondary structure elements are shown on top, coloured according to domain definition (slab – blue; catalytic – deep purple; neck – pink; head – red). The identified double psi-β barrels DPBB1 and DPBB2 are outlined by deep purple boxes. The flap sub-domain and the potential "bridge helices" are also represented by boxes, coloured light purple and grey, respectively.



B. Cartoon representation of the secondary structure elements in QDE-1 ΔN $P2_1$ homodimer with the sequence conserved motifs across other cellular RdRPs highlighted: motif 1 – red; motif 2 – orange; motif 3 – dark yellow; motif 4 – purple; motif 5 – dark pink; motif 6 – bright pink; motif 7 – blue. The non-crystallographic two-fold is represented as a grey line. Two magnesium ions are shown as green spheres. Top panel: front view, with two-fold axis aligned on the plane of the page. Bottom left panel: bottom view (rotated 90° counter-clockwise in *x* from front view). Bottom right panel: QDE-1 ΔN homodimer subunit A active site cleft. In this panel, the invariant residues are shown as ball-and-stick representation (C – according to motif colour; N – blue; O – red).

4.7.2. Structural analysis

Unlike viral RdRPs, QDE-1 ΔN does not exhibit the canonical polymerase right-hand signature with palm-fingers-thumb domains (Ollis *et al.*, 1985). Furthermore, initial searches of the Protein Data Bank (PDB) (Berman *et al.*, 2000) using Dali (Holm and Sander, 1998) and SSM (Krissinel and Henrick, 2004) revealed no obvious structural homologues, even when using the separate domains and structural motifs. A different approach to identify possible structural equivalences was therefore necessary. The program ASSAM (Artymiuk *et al.*, 1994) can recognise spatial arrangements of side chains in active sites by comparing them to a database of protein structures. A search for similarities with the proposed three catalytic aspartates and coordinating magnesium ion, within a 6Å sphere, retrieved several metal binding sites. Amongst the best matches were the various available structures for the two large subunits of RNA polymerase II from yeast (Rbp1 and Rbp2) (Cramer *et al.*, 2001) and bacteria (β' and β) (Vassylyev *et al.*, 2002). These DNA-dependent RNA polymerases also have their catalytic aspartates similarly spaced on alternating residues (DxDxD sequence motif). ASSAM also provided a rotation matrix and translation vector to superimpose both the yeast and bacterial DdRP subunits onto to our QDE-1 ΔN model. Unless otherwise stated, the coordinates of QDE-1 ΔN subunit A were used in these structural analysis.

A. Comparison with DNA-dependent RNA polymerases

Optimization of the superposition of these polymerases with QDE-1 ΔN using SHP (Stuart *et al.*, 1979) revealed that the double-psi β-barrel domains in QDE-1 ΔN are structurally related with the DPBBs in the bacterial RNApolII β' and β subunits and yeast equivalents Rbp1 and Rbp2, respectively (Fig. 4.11.A). Surprisingly, whilst the DdRPs have a double-psi β barrel domain in each subunit, with one contributing to the active site with the catalytic aspartates and the other with several positively charged residues (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Vassylyev *et al.*, 2002; Artsimovitch *et al.*, 2004; Westover *et al.*, 2004), QDE-1 ΔN has both domains on a single polypeptide chain. Optimized superposition based only on the DPBB domains reveals that 81 residues of DPBB2 in QDE-1 ΔN and the bacterial β' subunit DPBB are structurally equivalent, within an rms deviation in Cα positions of 2.2Å. Yeast Rbp1 DPBB and QDE-1 ΔN DPBB2 have 85 residues with Cα positions within an rms deviation of 2.1Å. The other β-barrel domain is not as strongly conserved, although 74 (bacterial) and 67 (yeast) residues are superimposable to DPBB1 of QDE-1 ΔN (rms deviation in Cα positions 3.0 & 3.1 Å, respectively). Furthermore, QDE-1 ΔN catalytic aspartates lie within 1.4 & 1.3 Å of the bacterial and yeast catalytic residues, respectively. A magnesium ion similarly coordinated by the three catalytic aspartates has also been identified in the DdRPs. A long helix (bridge helix) proposed to be relevant for nucleic acid – protein interactions during translocation of the RNA-DNA duplex in yeast Rbp2 (residues 810-846) (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Westover *et al.*, 2004) is structurally equivalent to α-helices 27 and 28 in QDE-1 ΔN. The fact that the bacterial equivalent to the yeast

bridge helix also consists of two smaller helices (residues 1067-1081 and 1083-1093) supports the hypothesis that they serve a common functional role. However, a structure-based sequence alignment (Fig. 4.12) reveals that the sequence homology is very low, even within the closely structurally related DPBBs. Extended searches using other domains and structural motifs of QDE-1 ΔN, even when the smaller subunits of the RNA polymerase II complex were included, retrieved no further sequence, structural or functional motifs, showing how divergent the overall of QDE-1 ΔN and DdRPs are (Fig.4.11.B).

In recent years, structures of yeast RNApolII elongation and initiation complexes have been solved (Armache *et al.*, 2003; Cramer, 2004; Kettenberger *et al.*, 2004) so a superposition of an RNApolII - RNA-DNA duplex oligonucleotide complex with QDE-1 ΔN was also carried out. The position of DPBBs in both models was used, with rms deviations of the Cα positions similar to those observed for the superposition of the yeast apo enzyme. The RNA-DNA duplex maps onto tunnel (i) confirming that it corresponds to the RNA product groove (Fig. 4.13.A). Since two possible conformations for the RNA tunnel are observed in the $P2_1$ dimer, we also superimposed the RNA-DNA duplex onto the coordinates of subunit B. Some of the clashes observed between the head domain in subunit A and the RNA-DNA duplex are not present in subunit B opened conformation (Fig. 4.13.B). These changes are indicative of the conformational rearrangements that would be necessary upon RNA binding and during catalysis.

Overall, the DdRPs are much more elaborate than QDE-1 ΔN and, despite the fact that no significant structural homology, beyond the vicinity of the active site, is detectable, there may be some functional relationship at the level of protein domains.

The flexibility of the head domain could allow it to close down on the RNA duplex product, in a mechanism similar to the closing down on the RNA-DNA duplex of the clamp region in RNApolII (Cramer *et al.*, 2001). By analogy to the function of the protrusion-lobe domains of RNApolII (Cramer *et al.*, 2001), we can also propose that accommodation of the duplex and its stabilization is aided by the rearrangement of the slab domain.

The presence of a number of structurally equivalent residues within the sequence conserved motifs in cRdRPs reinforces the view that the mechanism of both types of enzyme is similar. Some of the strictly conserved residues in cRdRPs have structurally equivalents in DdRPs, namely R962 and P964 which occupy positions equivalent to R446 and P448 in yeast Rbp1, known to interact with the daughter strand in the RNA-DNA duplex. The superposition of the RNA-DNA duplex into the QDE-1 ΔN model indicates that this interaction would also be important in stabilizing the RNA duplex. Furthermore, the conserved positively charged residues on DPBB1 (namely Q736, K743 and K767) have structural equivalencies with residues in Rbp2 DPBB (K979, K987 and R1020, respectively) that are important for interaction with incoming NTPs and in stabilization of the active cleft (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Westover *et al.*, 2004). Recent studies of the unique replication strategy of hepatitis Δ virus revealed, surprisingly, that a human DdRP, polymerase II (PolII), equivalent to yeast RNApolII, is likely to be involved in RNA replication (Lai, 2005), strengthen the hypothesis that these enzymes share a common reaction mechanism.

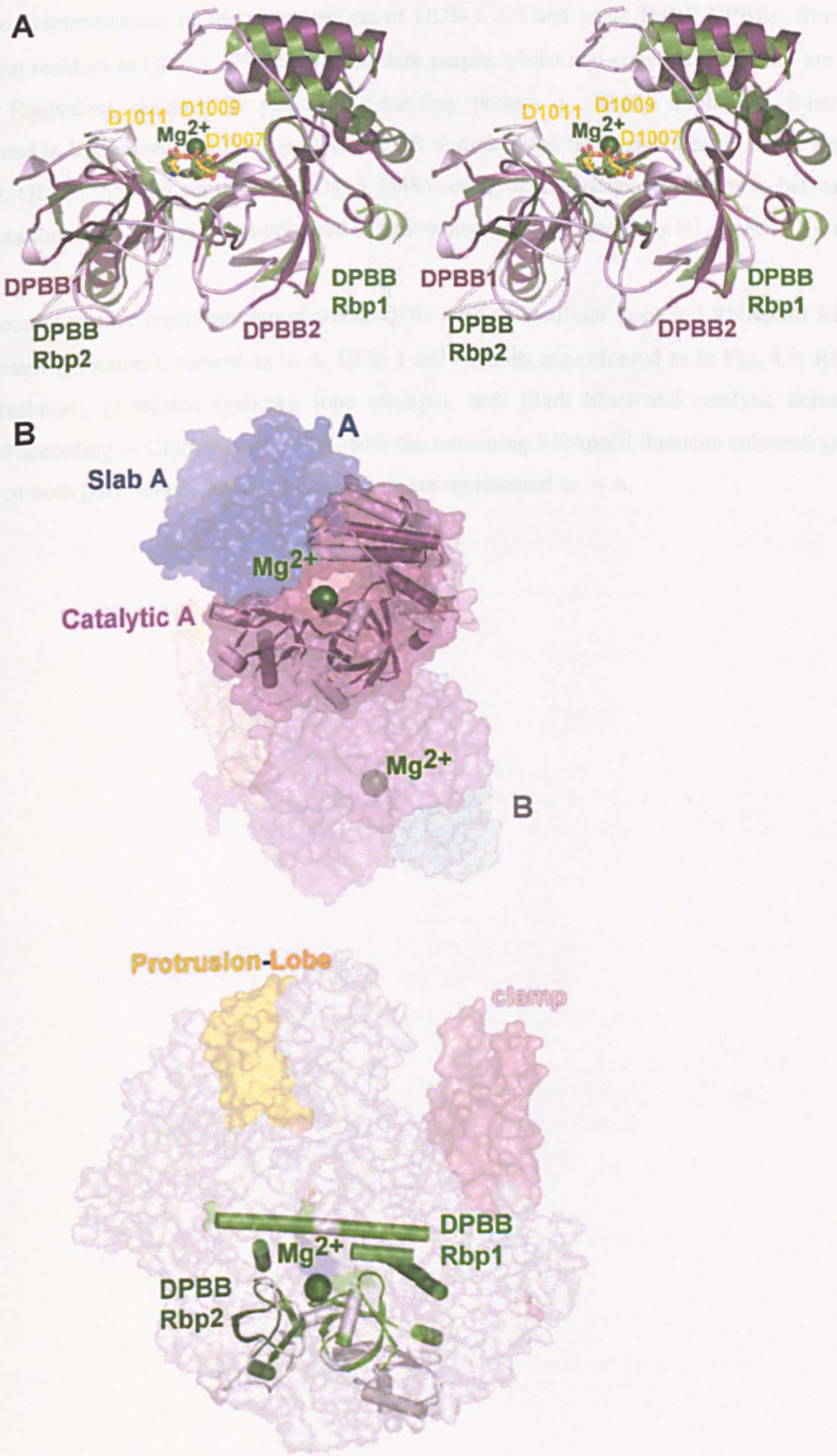

A
D1011
D1009
Mg2+
D1007
DPBB1
DPBB Rbp2
DPBB2
DPBB Rbp1
B
A
Slab A
Mg2+
Catalytic A
Mg2+
B
Protrusion-Lobe
clamp
DPBB Rbp1
Mg2+
DPBB Rbp2

Figure 4.11. Comparison of QDE-1 ΔN and multisubunit DdRPs

A. Stereo representation of the superposition of QDE-1 ΔN and yeast DdRP DPBBs. Structurally equivalent residues in QDE-1 ΔN are coloured dark purple, whilst non-equivalent residues are in light purple. Equivalent residues in yeast RNA-binding protein 1 (Rbp1) double-psi β-barrel are represented in light green and those in Rbp2 DPBB in dark green (non-equivalent residues are shown in grey). QDE-1 ΔN and yeast (D481, D483, D485) catalytic aspartates are shown in ball-and-stick representation , with carbon atoms coloured in yellow and green, respectively (C – green, O – red, N – blue).

B. Molecular surface representation of whole QDE-1 ΔN homodimer (top) and RNApolII Rbp1 and Rbp2 complex (bottom), viewed as in A. QDE-1 ΔN subunits are coloured as in Fig. 4.9; RNApolII clamp (salmon), protrusion (yellow), lobe (orange), wall (dark blue) and catalytic domains are coloured according to Cramer *et al.*, 2001, with the remaining RNApolII domains coloured grey. The DPBBs of both polymerases and magnesium ions are represented as in A.

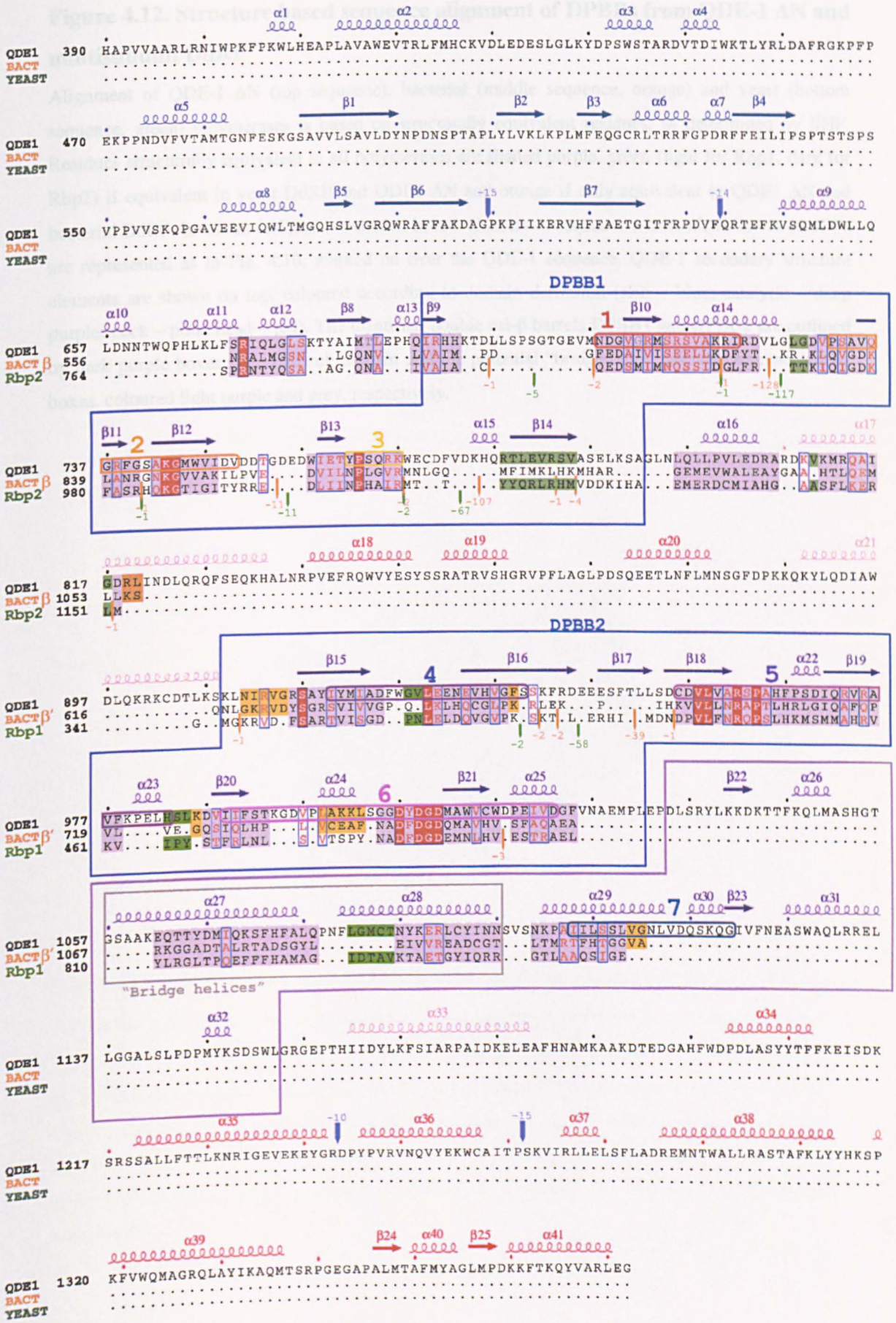

DPBB1
DPBB2
"Bridge helices"
QDE1
BACT
YEAST
Rbp2
Rbp1

Figure 4.12. Structure based sequence alignment of DPBBs from QDE-1 ΔN and multisubunit DdRP

Alignment of QDE-1 ΔN (top sequence), bacterial (middle sequence, orange) and yeast (bottom sequence, green) polymerases is based on structurally equivalent residues, as determined by SHP. Residues structurally equivalent in all polymerases are shaded purple, green (light for Rbp1, dark for Rbp2) if equivalent in yeast DdRP and QDE1 ΔN and orange if only equivalent in QDE1 ΔN and bacterial DdRP. Invariant residues are shaded in red. Conserved sequence motifs identified in cRdRPs are represented as in Fig. 4.10, marked on over the QDE-1 sequence. QDE-1 secondary structure elements are shown on top, coloured according to domain definition (slab – blue; catalytic – deep purple; neck – pink; head – red). The identified double psi-β barrels DPBB1 and DPBB2 are outlined by dark purple boxes. The flap sub-domain and the potential "bridge helix" are also represented by boxes, coloured light purple and grey, respectively.

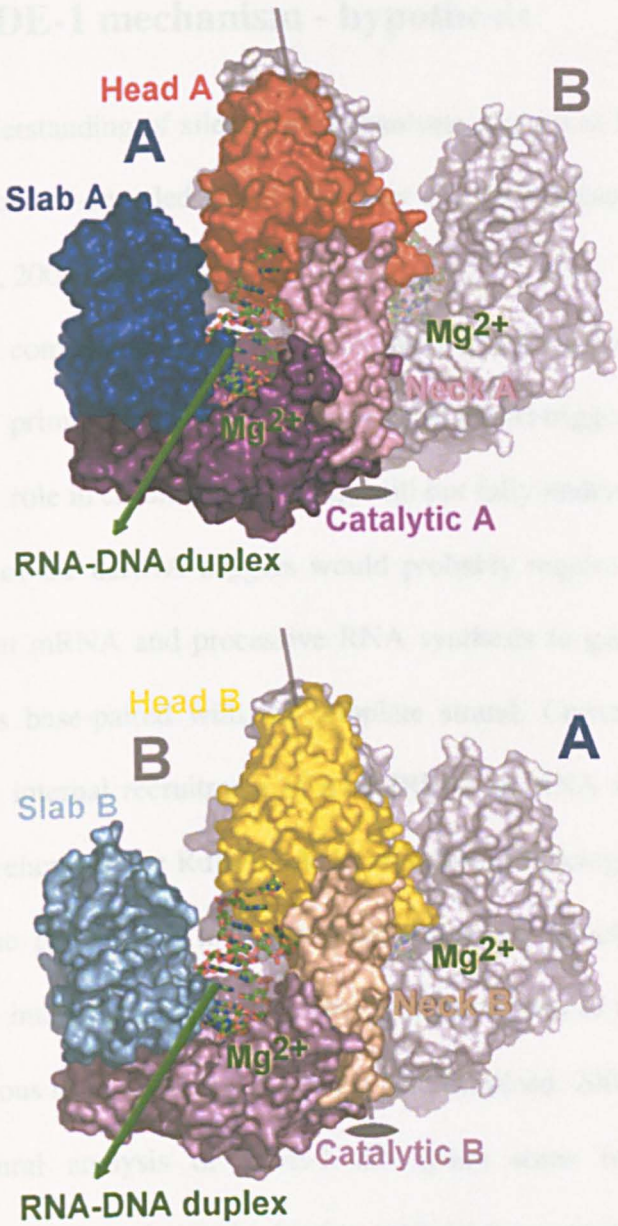


Figure 4.13. RNA-DNA duplex fitted onto open and closed conformations

A. Surface representation of the homodimer from monomer A (coloured by domain definition, as in Fig. 4.6.B) with superimposed RNA-DNA duplex from the yeast RNApolII elongation complex.

B. Surface representation of the homodimer from monomer B (orientation equivalent to A, rotated by 180°), coloured according to domain definition, as in Fig. 4.6.B, with superimposed RNA-DNA duplex.

RNA duplex is shown as ball-and-stick representation (C – green, O – red, N – blue, P – purple) in both panels.

4.8. QDE-1 mechanism - hypothesis

Current understanding of silencing mechanisms suggest at least three possible functional roles for cell-encoded RdRPs (Sijen *et al.*, 2001; Makeyev and Bamford, 2002; Baulcombe, 2004; Motamedi *et al.*, 2004):

(i) conversion of aberrant RNA into dsRNA triggers in PTGS

(ii) primer-dependent amplification of RNAi triggers

(iii) role in chromatin silencing, still not fully understood.

Production of the dsRNA triggers would probably require initiation at the 3' end of an aberrant mRNA and processive RNA synthesis to generate a full-length copy that persists base-paired with the template strand. Conversely, reaction (ii) would require an internal recruitment of a cRdRP to its RNA targets. Despite not being fully comprehended, the RdRPs role in chromatin silencing is likely to involve recruitment of the polymerase to the nascent transcript, which can also only be achieved through internal initiation. The two possible modes of initiation have been observed in previous *in vitro* studies (Makeyev and Bamford, 2002).

The structural analysis of QDE-1 ΔN gives some hints to a possible polymerisation mechanism. Internal initiation without the need of a free 3' end could potentially be achieved within a relatively open conformation as seen in subunit B. In this case, a particular "opened" subunit could be loaded to a nascent RNA transcript for reaction to occur. Initiation at the 3' end of an RNA template as thought to be involved in production of dsRNA triggers from aberrant RNA (aRNA) would possibly be favoured by the closed conformation observed in subunit A. Dimerization would, therefore, allow a control of the flexibility and stability of the

heads, providing a way to regulate the initiation mode of the reaction. This proposed mechanism implies that only one QDE-1 subunit is catalytically active at any particular time, therefore priming the other site for binding by holding it in an open conformation. Indeed, the molecular architecture might provide steric constraints that favour the molecule working as a "two-stroke motor", with facilitated active site switching. The intermediate conformations observed in the *C*2 crystal form may represent a transition state or an inactive, dormant form of the polymerase dimer, which might be activated by rearrangements of the head and slab domains upon template binding and/or interaction with other proteins.

In the fission yeast *Schizosaccharomyces pombe*, the cellular RNA-dependent RNA polymerase Rdp1 is part of a complex – RNA-dependent RNA polymerase complex RDRC – together with an helicase Hrr1 and a protein related to the poly(A) polymerases. This complex has been shown to be essential for heterochromatin silencing (Motamedi *et al.*, 2004). The RDRC complex co-localizes in the RNA-induced transcriptional silencing (RITS) complex at noncoding centromeric RNAs and the complexes associate together in a manner dependent of Dcr1, the *S. pombe* DICER protein, and with Clr4, a histone methyltransferase. A possible explanation for these observations is that RITS acts as a "priming complex" for the dsRNA synthesis activity of RDRC (Motamedi *et al.*, 2004). Other studies have shown that, in *S. pombe*, the RdRP activity is essential for heterochromatin silencing and recruitment of RITS components to the centromeric RNAs (Sugiyama *et al.*, 2005). In *N. crassa*, however, DNA methylation has been shown to be independent of the quelling components (Freitag *et al.*, 2004). Nevertheless, interactions of QDE-1 with components of the RNA silencing pathway, namely the DICER and Argonaute

proteins might be relevant for the mechanism of QDE-1. Namely, a "priming" function of the silencing complexes could provide the 3' end that one of the subunits of the QDE-1 dimer binds to in an opened conformation, and uses as a template to produce dsRNA products. The interactions of QDE-1 with other members of the silencing pathway might be mediated by the N-terminal domain of the protein. Since this domain is not necessary for activity *in vitro*, it is possible that its interactions with other RNA silencing proteins allow extra levels of regulation of the polymerase activity, and/or are involved in recruitment of the polymerase to particular cell compartments or signal molecules. In any case, the precise mechanism of regulation of the activity of QDE-1 and the pathways it is involved, together with full molecular and biochemical details of RNA silencing mechanisms are still only beginning to be understood and further experiments have to be carried out to test these hypotheses.

Chapter 5

Conclusions and future directions

5.1. Φ6pol and other viral RdRPs

The structural studies of Φ6pol described in this thesis provide a comprehensive view of the enzyme, giving a better understanding of template entry and specificity, divalent ion effects and the events leading to initiation of polymerization. Moreover, structural studies of complex Φ6pol-RNA-Mg^{2+}-GTP have demonstrated *in crystallo* ϕ6pol polymerase activity, hence validating the conclusions drawn from these studies. A more complete and detailed mechanism for initiation can, therefore, be proposed. In the description of this model, the template and daughter nucleotides are numbered as defined by Butcher *et al.* (2001) and described in Chapter 1.

Butcher *et al.* (2001) proposed that the plough-like feature at the entrance of the template tunnel would facilitate RNA strand separation, feeding one strand directly into the positively charged template tunnel (see Chapter 1, Fig. 1.10.C). During transcription, selective attachment of the minus strand in the template tunnel then ensures that the correct positive RNA strand leaves the polymerase complex into the cytoplasm where it acts as mRNA. As the RNA template binds to the empty enzyme (Fig. 5.1.A), it is stabilized with interactions of positively charged residues lining the tunnel. Slight conformational rearrangements of these residues accompany the movement of the RNA template within the tunnel (Fig. 5.1.B) until its 3' cytidine

base fits snugly in the specificity pocket site S (Fig. 5.1.C). This pocket lies well beyond the catalytic site C, within the C-terminal sub-domain and specifically recognizes the preferred cytidine T1. The presence of an extra hydroxyl group in the ribose sugar rings of RNA nucleotides allows extra hydrogen bonds to be established between the template and the protein, stabilizing the bound RNA oligonucleotides within the tunnel, when compared to DNA. These interactions account, at a molecular level, for the known preference of Φ6pol for RNA templates. Viral RNA-dependent RNA polymerases all exhibit a similarly formed template tunnel, more restricted than in other template dependent single subunit polymerases. Therefore, the observed interactions between Φ6pol and template that stabilize it within the tunnel are likely to be a general feature of vRdRPs. Picornaviruses polymerases, nevertheless, use a primer-dependent initiation mode and the need to accommodate a bulkier template might lead to a slightly different set of interactions being involved in template stabilisation.

As RNA migrates down the template tunnel, different NTPs bind within the substrate channel, presumably in rapid exchange. A set of arginine residues – R225, R268 and R270 – defines an "interrogation site I" and are proposed to prime for the correct NTP (Fig. 5.1.D). This is defined as the interrogation state, where interactions with the arginines in site I locks the triphosphate backbone and allows the NTP base to interrogate the template base positioned at the catalytic site C. As the NTP is directed towards site C, a range of interactions with the "interrogating arginines", together with one of the conserved catalytical aspartates – D453 – and Y630 of the "initiation stabilizing platform P" loop allows reorientation of the nucleoside within the tunnel (Fig. 5.1.E-I). Finally, the nucleoside is tethered within site C, stabilized

by interactions of the priming arginines with the phosphate backbone and the O2' group in the ribose sugar with the catalytic D453. These interactions lock the GTP in a position that allows interactions with the bases of nucleotides in positions T2 and T3 (Fig. 5.1.J). This is a post-interrogation state, where the incoming NTP is positioned ready to establish base-pairing interactions with the template but the RNA template is still tethered in site S. The fact that all vRdRPs possess a positively lined fingertips sub-domain indicates that they might have a similar role in priming and directing the incoming NTPs and stabilizing the post-interrogation state in other viral polymerases.

The incoming GTP molecule (denoted henceforth as D2) in the post-interrogation state then rearranges to establish Watson-Crick base-pair interactions with T2, stabilized by base-stacking interactions between the base of D2 and the ring of Y630 (Fig. 5.1.K). When correct base-pairing interactions are established, a conformational change is then presumably triggered in the molecule that ultimately results in a displacement of the template away from the S pocket (Fig. 5.1.L). This hypothesis is strengthened by the fact that some regions of the thumb, fingers and C-terminal sub-domains (residues 37-100, 503-551 & 567-617 – thumb; 204-268 - fingers and fingertips; 618-646 - C-terminal) show a considerable degree of disorder in the post-interrogation state (see Chapter 2, Fig. 2.9). Once the second GTP (termed D1 as it will base-pair with T1) is directed towards the P site by interactions with the priming residues, a pre-initiation complex is formed: GTP molecule D1 base-pairs with RNA nucleotide T1 whilst D2 base-pairs with T2 (Fig. 5.1.M). These interactions are further stabilized by base-stacking interactions with Y630 from the "initiation platform" loop. A magnesium ion is found coordinating the phosphate

backbone of D2, whilst the D1 backbone is displaced away by hydrogen bonds with Y630 and D453 (Fig. 5.1M). Presumably, a second metal ion is transiently present, accompanying the second GTP molecule, although we could only detect weak density for this second metal ion. Alternatively, the second magnesium could subsequently bind to the complex between the phosphate moieties of both GTP molecules. The coordination of the phosphate backbone of D1 by this second ion places it within the correct distance and orientation for catalysis to occur, as described for the general "two metal ion" mechanism (see Chapter 1, Fig. 1.3). This conformational arrangement of the enzyme, template and substrate molecules corresponds to the initiation competent state (Fig. 5.1.N).

Once catalysis occurs and the bond between the β and α-phosphates of D2 is broken, pyrophosphate is formed and is transiently stabilized by the arginine patch at site I. The RNA template-daughter duplex is no longer stabilized by interactions with the arginine residues and, presumably, can only be further stabilized if it ratchets down by one nucleotide towards the C-terminal sub-domain. As described previously (Butcher *et al.*, 2001), this sub-domain creates a physical barrier and has to be displaced in order to allow translocation of the dsRNA product during elongation (Fig. 5.1.O). Similarly, structural features equivalent to the C-terminal sub-domain in polymerases from Hepatitis C (HCV) and Bovine Viral Diarrhoea (BVDV) viruses are also proposed to be flexible enough to give way to the nascent dsRNA (Hong *et al.*, 2001).

The conformational changes involved in the switch from initiation to elongation in these vRdRPs are poorly understood. The switch to elongation can not be captured in the present Φ6pol crystallographic system by soaking experiments

since the C-terminal sub-domain is involved in crystallographic contacts which would be disrupted upon domain movement, hence distorting the crystal packing. Furthermore, structural studies of a Φ6pol mutant that preferentially initiates in a back-priming mode have provided evidence that those protruding structural features present in HCVpol, BVDVpol and Φ6pol are also relevant to primer-independent initiation. In Φ6pol, the primer-independent initiation form of the polymerase is stabilized by a discreet set of contacts of the C-terminal with other sub-domains, that also seem to control the switch from the initiation to elongation phase. Relatively modest changes in those contacts can have long range effects and result in a premature conformational switching that produces a structure with a preference for back-priming initiation mode. A similar mechanism is, therefore, likely to be relevant in HCVpol and BVDVpol. Moreover, these studies provide a structural explanation for the absence of such features in vRdRPs that utilise primer-dependent mechanisms.

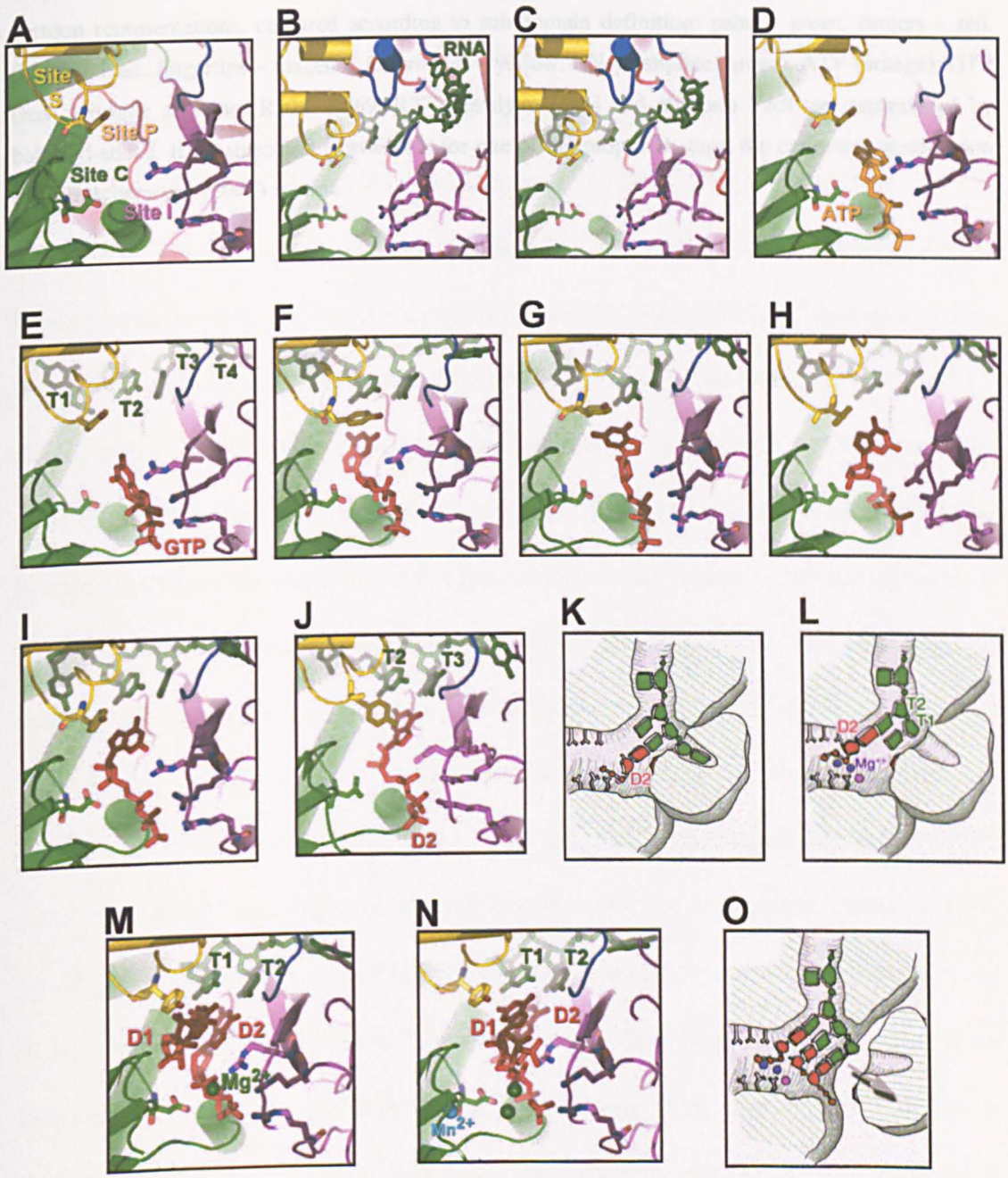


Figure 5.1. Detailed mechanism of Φ6pol initiation

A. Φ6pol active cleft with interrogation (I) and catalytic (C) sites, S pocket and platform (P) identified. **B.** Intermediate conformation of the RNA template. **C.** RNA template tethered in the S pocket. **D.** ATP bound at the I site (Butcher *et al.*, 2001). **E-I.** GTP migration though the substrate tunnel, "primed" and oriented by the arginine patch of site I, platform residue Y630 and catalytic D453. **J.** "post-interrogation" state. **K.** Proposed state with GTP (D2) base-pairing with T2. **L.** RNA template is displaced from the S pocket, ready for the binding of second GTP molecule. **M.** "pre-initiation" state. **N.** Initiation competent state (Butcher *et al.*, 2001). **O.** Switch to elongation with dsRNA product coupled to the displacement of the C-terminal sub-domain.

Steps in the initiation mechanism structurally characterised, viewed from the NTP entry, are shown as cartoon representations, coloured according to sub-domain definition: palm – green, fingers – red, thumb – blue, fingertips – magenta, C-terminal – yellow. RNA template (green), ATP (orange), GTP (red), priming arginines R225, R269, R270, catalytic D453 and platform Y630 are represented by ball-and-sticks. If no structural is available for one of the proposed steps, the cartoon representation from Butcher *et al.* (2001) is used.

The structural basis for the widely known stimulatory effect of manganese ions in polymerisation (Ferrari *et al.*, 1999; Alaoui-Lsmaili *et al.*, 2000; Kim *et al.*, 2000; Lampio *et al.*, 2000; Makeyev and Bamford, 2000a; Yang *et al.*, 2001; Ranjith-Kumar *et al.*, 2002b; Ranjith-Kumar *et al.*, 2003; Yang *et al.*, 2003; Laurila *et al.*, 2005) however, still remains unclear. The fact that "manganese-free" Φ6pol structures showed no structural changes compared to the wild type protein seems to exclude its effect on overall structural stability. However, the structural studies of a Φ6pol mutant with a changed coordination affinity for manganese and its complexes with oligonucleotides and/or NTPs revealed an intrinsically more flexible structure indicating that the coordination of the ion can affect the protein structural stability. If so, the role of manganese might be to lock the Φ6pol in a more rigid structure that ensures efficient initiation, hence stimulating polymerisation. These effects might not have been noticed in the "manganese-free" structural studies since no co-crystallization or soaking experiments with template oligonucleotides and/or NTPs were carried out. Alternatively, the stimulatory effect of manganese could possibly be related to the early stages of the mechanism, namely the rearrangements involved from interrogation to initiation competent states. The hypothesis that manganese stimulation lowers the GTP dissociation constant from the *de novo* initiation complex, therefore providing additional stabilization, would provide a possible general mechanism for the observed effect of the ion in polymerisation. However, direct evidence of this effect has proved elusive both in Φ6pol and other RdRPs. It is also unclear whether manganese is involved in the initiation competent state, in the formation of the first bond or in both stages of the reaction. Furthermore, in the case of Φ6pol, HCVpol and BVDVpol that have C-terminal protruding sub-domains that

block the exit of the dsRNA product, it is also possible that the ion could be involved in the switch from initiation to elongation. Therefore, it would be interesting to measure binding constants for GTP in the initiation complex with RNA template for E491Q mutants and the WT Φ6pol to check the involvement of the different manganese binding sites. Determining the efficiency of polymerization using an RNA template that contains only two cytidines at the 3' end (5'-...CC-3') would be an indirect way to measure the productive GTP binding and the first round of nucleotidyl transfer. An alternative would be to measure the binding constant directly using a radioactively labelled non-hydrolysable GTP analogue. Clearly, further biochemical and structural studies of Φ6pol and other vRdRPs are necessary to provide a definite mechanism for the manganese stimulation of RNA polymerisation and these are only illustrations of possible experiments.

5.2. QDE-1 and cellular RdRPs

The structure of the RNA-dependent RNA polymerase QDE-1 from *Neurospora crassa* is the first cellular RdRP solved to date, providing important insights into the architecture of these enzymes, its evolutionary relationship with other polymerases and its role in the RNA silencing mechanism. Surprisingly, the protein was found to be a dimer both in the crystals and in solution, in the conditions that allowed its polymerisation activity to be studied (Makeyev and Bamford, 2002). The observation that the subunits in the dimer have an "opened" and "closed" conformation gives some indication of the possible mechanism of polymerisation. We propose that QDE-1 might function as a "two-stroke motor", with only one QDE-1 subunit catalytically active at any particular time. Template binding in one subunit would cause it to close down on the template/product, therefore priming the other site for binding by holding it in an open conformation. The dimeric molecular architecture would hence provide the steric constraints that allow the switching between catalytic active sites.

Sequence analysis of other cellular RdRPs proposed to be involved in RNA silencing mechanisms revealed seven conserved motifs that map onto to the active cleft in QDE-1 (see Chapter 4, section 4.7.1). This indicates that, despite their weak sequence homology, all cellular RdRPs would share a similar active cleft architecture formed by two double-psi β-barrels (DPBB). It is reasonable to suppose that they would also have flexible domains that can clamp down on the template/product during polymerisation, a mechanism described for multisubunit DdRPs whose active clefts are also formed by DPBBs. This would imply the possibility that dimerization

can occur in other, if not all, cellular RdRPs and that the "two-stroke motor" could be a general mechanism for the activity of cRdRPs.

A possible way to test these hypotheses in QDE-1 is to determine the structure of complexes with NTPs and/or RNA/DNA oligonucleotides. Several attempts to obtain structures of such complexes where carried out, unfortunately with little success. Although in some cases we were able to detect electron density features in the proposed RNA tunnel, these were too weak for an RNA model to be fitted. One way to improve the signal-to-noise ratio is to use RNA oligos with nucleotides labelled with heavy atoms, for example iodine or bromide. If data were collected at a wavelength that maximised the anomalous signal of the heavy atoms, not only could the binding of RNA be more clearly identified by stronger electron density features, but also its orientation within the tunnel could be determined if the RNA oligos were designed so to that the pattern of heavy atoms clearly defined the 3' end. Soaking experiments with iodinated RNA oligonucleotides were attempted but, as had happened with previous soaking experiments, the quality of the diffraction was so poor the data could not be processed. This, however, does not imply that this experiment could not be carried out successfully if more extensive studies are carried out to determine the optimal conditions of soaking, which clearly seem to vary with crystal size and age. Cross-linking the RNA template oligonucleotides with the protein would be an alternative way to determine the structure of a complex, providing the resulting protein was crystallisable. Furthermore, as more knowledge about the specific requirements for QDE-1 is gained by our collaborators at the University of Helsinki, more informed oligonucleotides can be designed in order to obtain a complex with more tightly bound RNA. Another strategy to provide a

clearer view of the mechanism of QDE-1 would be to design mutated versions of the protein where residues involved in dimer contacts are altered in a way predicted to disrupt dimerization. Biochemical and structural studies of these mutated proteins would provide a powerful way to address the "two-stroke" motor hypothesis and the biological relevance of dimerization. An alternative way to test the "two-stroke" mechanism hypothesis would be to produce a form of the protein where one of the catalytic aspartates is mutated. A mixture of mutant and WT QDE-1 protein (1:1 ratio) would presumably be an heterogeneous population with one third of dimers composed of two WT monomers, one third of mutant:mutant dimers and one third of mutant:WT dimers. If the two active sites in the dimer are independent, the presence of one catalytic active site would still allow the protein to polymerise RNA. Catalytic activity of the mixture would therefore be approximately half of those observed for an homogeneous WT:WT dimers population. Conversely, if, as we propose, the protein functions as a "two-stroke" motor, only the WT:WT dimers would exhibit significant catalytic activity and the observed catalysis levels would then be a quarter of that observed for a homogeneous population of WT dimers. Finally, structural studies of the full-length protein would expand the understanding of this cellular RdRP and possibly provide clues to the interactions that QDE-1 is involved in the RNA silencing pathways. Early attempts to crystallise the full-length protein have been carried out, with very small crystals found in some conditions of different crystallisation screens. These crystals were tested and did not diffract, probably due to their very small size, but they are good starting points for further structural studies.

Structural and biochemical studies of other cellular RdRPs are necessary to further elucidate the role of these enzymes in the RNA silencing mechanisms.

Moreover, recent studies have indicated that RdRPs and other silencing components might be involved in different pathways in different organisms – as for example the need for RdRP in *S. pombe* (Motamedi *et al.*, 2004; Sugiyama *et al.*, 2005) chromatin silencing but apparently not in *N. crassa* (Freitag *et al.*, 2004). Further studies of the RNA silencing pathways in different organisms and the different interactions of the proteins and their complexes involved are necessary to test some of these hypotheses and allow a detailed understanding of this complex mechanism.

5.3. Evolutionary implications

Classification of cellular organisms and their evolutionary relationship has its roots in a Linnean morphology-based approach. The advent of genomic efforts and the wealth of genetic information generated, has moved the focus towards gene-based evolutionary relationships between organisms. This approach has proved to be effective in studies of systems that are reasonably closely related, particularly when applied to virus classification. However, when considering long evolutionary time spans, whilst structure and function can be preserved, sequence changes to a point where no recognisable similarity might be identified. Recently, Bamford, Burnett and Stuart (2002) proposed an alternative way to identify viral lineages based on the atomic structures of viral proteins that belong to the so-called virus "self". The evolution of protein structure is slowed by functional constraints. Hence, comparison of atomic structures of particular proteins involved in vital cellular and/or viral functions provides a clearer view of the phylogeny of these proteins and can elucidate the intricate evolutionary relationships of cellular organisms as well as virus lineage. Indeed, the accumulating structural data suggest that any given protein fold can be achieved by very different amino acid sequences, as long as secondary structure elements and critical intramolecular contacts are preserved. From considerations of the template dependent polymerases, it is clear that similar tertiary structures are often formed by very dissimilar amino acid sequences. One possibility is that this is a result of convergent evolution, when evolutionarily unrelated proteins independently "invent" analogous, functionally competent folds. The three different folds described for template dependent nucleotidyl transferases – right-hand, Polβ and multisubunit DdRPs – have been considered examples of this hypothesis (Steitz,

1998). Alternatively, more substantial structural similarity without sequence homology signifies the likely divergence from a common ancestor [as an example, see (Benson *et al.*, 1999)]. Since vRdRP and cRdRP have dissimilar overall folds, despite the conservation of carboxylate moieties at the active site, at least two distinct pathways have arrived at a functional RdRP enzyme. The demonstration of multiple solutions to the functional requirements of RNA-dependent RNA synthesis makes it less likely that structural similarity arises from convergent evolution.

The unexpected evolutionary link between cellular RNA-dependent RNA polymerases and multisubunit DNA-dependent DNA polymerases revealed by the structure of QDE-1 prompted us to analyse the evolution of all different template dependent nucleotidyl transferases solved to date based on their structures, with a particular emphasis on double-psi β-barrel (DPBB) containing enzymes and viral RdRPs. A modified version of the program SHP (Stuart *et al.*, 1979) was used to superimpose the coordinates of each vRdRP solved to date, QDE-1 DPBB1 and DPBB2 and the yeast and bacterial equivalents, Polβ, HIV-1 reverse transcriptase, T7pol and the Klenow fragment from DNA polymerase I. Each polymerase was systematically superimposed with all other enzymes and the respective evolutionary distances were calculated based on the sum of all probabilities of equivalent residues as calculated by SHP (Stuart *et al.*, 1979) and the average number of protein residues (see legend to Fig. 5.2). An evolutionary tree was then built that reflects the different relationships of all the polymerases compared in this analysis using the PHYLIP package (Felsenstein, 1989) (Fig. 5.2).

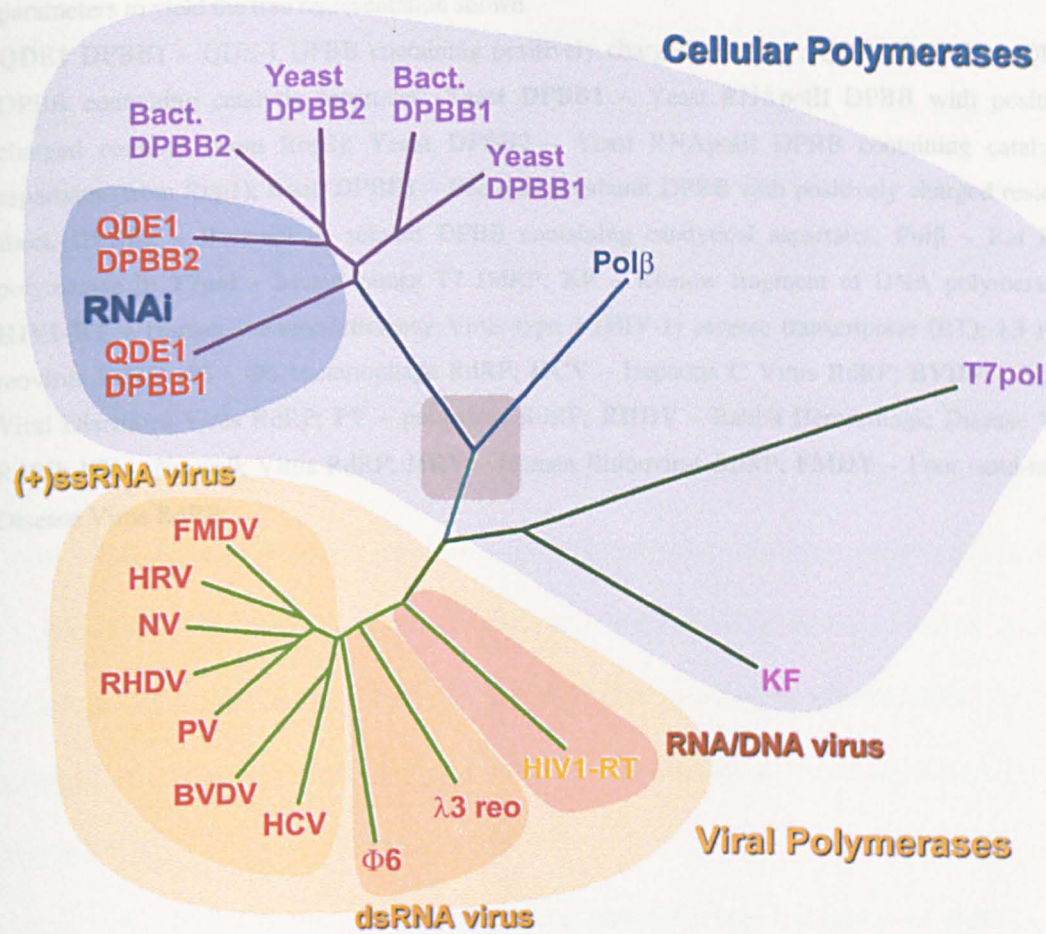


Figure 5.2. Structure based phylogenetic tree of template dependent polymerases

Cellular (light blue shade) and viral (pale orange shade) polymerases phylogenetic relationships based on structural comparisons. Viral polymerases are shaded also according to virus type: (+)ssRNA virus - light orange; dsRNA virus – orange; retrovirus – pale red. RdRPs are labelled red, DdRPs lilac, DdDPs dark pink and RdDPs orange. Branches are coloured according to structural fold: green – right hand (dark – cellular; light – viral), Polβ – blue; DPBB-containing fold – dark magenta. RNA silencing cellular RdRPs subset is shaded blue. We do not imply that all polymerases originated form a common ancestor. To indicate this, the root of the tree is shaded grey. The tree is calculated by performing a gap-penalty-weighted superposition using a modified version of SHP (Stuart *et al.*, 1979). This procedure maximises the total sum of probabilities of equivalence between pairs of residues for the proteins being analysed. The summed probability is then converted into an estimate of evolutionary distance using the empirical expression D=-log[(P-2)/(<N>-2)], where D is the evolutionary distance, P is the sum of probabilities and <N> the mean number of residues in the two molecules (Stuart *et al.*, unpublished). A full matrix of evolutionary distances was then calculated, using as input the coordinates of all analysed proteins or particular sub-domains when appropriate.

This matrix was then analysed using the PHYLIP package (Felsenstein, 1989) using the default parameters to yield the tree representation shown.

QDE1 DPBB1 – QDE-1 DPBB containing positively charged residues ; **QDE1 DPBB2** – QDE-1 DPBB containing catalytic aspartates; **Yeast DPBB1** – Yeast RNApolII DPBB with positively charged residues (from Rbp2); **Yeast DPBB2** – Yeast RNApolII DPBB containing catalytical aspartates (from Rbp1); **Bact. DPBB1** – Bacterial β subunit DPBB with positively charged residues; **Bact. DPBB2** – Bacterial β' subunit DPBB containing catalytical aspartates; **Polβ** – Rat DNA polymerase β; **T7pol** – bacteriophage T7 DdRP; **KF** – Klenow fragment of DNA polymerase I; **HIV1-RT** – Human Immunodeficiency Virus type 1 (HIV-1) reverse transcriptase (RT); **λ3 reo** – reovirus RdRP; Φ6 - **Φ6** bacteriophage RdRP; **HCV** – Hepatitis C Virus RdRP; **BVDV** – Bovine Viral Diarrhoea Virus RdRP; **PV** – poliovirus RdRP; **RHDV** – Rabbit Hemorrhagic Disease Virus RdRP; **NV** – Norwalk Virus RdRP; **HRV** – Human Rhinovirus RdRP; **FMDV** – Foot –and-mouth Disease Virus RdRP;

As expected, it is clear from our analysis that the right hand, Polβ and multisubunit DdRP folds are essentially unrelated. Surprisingly, cellular polymerases T7pol and the Klenow fragment of DNA polymerase I seem to be more distant relatives than previously proposed, although clearly evolved from the same ancestor. Furthermore, the close evolutionary link between dsRNA cystoviridal Φ6pol and RdRPs from (+)ssRNA flavivirus HCV and BVDV strengthens the hypothesis that these viral families have an evolutionary relationship. The fact that RdRPs from calici- and picornaviruses seem to be more closely related with each other than with other vRdRPs is somewhat surprising, as their polymerases were previously thought to be closely related to *Cystoviridae* and *Flaviviridae* RdRPs. More detailed structures of the poliovirus and recent determination of other structures of *Picorna-* and *Caliciviridae* polymerases is probably the reason why these relationships emerge only now.

A number of models have been proposed for the evolution of dsRNA and (+)ssRNA viruses, based on sequence and structural analysis of their proteins, particularly RdRPs (Bruenn, 1991; Koonin, 1992; Makeyev and Grimes, 2004). Sequence alignment of the C-terminal 200 residues of RdRPs from several 43 (+)ssRNA viruses and seven dsRNA viruses allowed identification of different clusters of related virus families. This led to the proposed hypothesis that ssRNA viruses, apart from picornavirus, might have originated from the dsRNA group. Furthermore, it was suggested that the separation between viruses infecting prokaryotes and those infecting eukaryotes arose prior to the emergence of dsRNA viruses (Bruenn, 1991). An alternative view (Koonin, 1992), also based on sequence comparisons, suggested occasional emergence of dsRNA families from different

(+)ssRNA. This controversy suggests that viral classification and lineage identification based solely on sequence comparisons, particularly for distantly related family virus, does not provide enough unambiguous information to derive robust phylogenetic relationships. More recently, an alternative approach has been proposed based on structural comparisons (Makeyev and Grimes, 2004). In this model, both (+)ssRNA and dsRNA viruses are related to an "RNA protovirus" that is envisioned as a vesicle containing RdRP and a dsRNA genome. Most dsRNA viruses would have emerged from the "protovirus", preserving the dsRNA configuration of the genome and substituting the lipid vesicle by the proteinaceous core that serves as a compartment for RNA synthesis. Conversely, most (+)ssRNA viruses inherited the membrane-associated RNA synthesis, which may proceed via dsRNA intermediates (Makeyev and Grimes, 2004).

Our structure-based analysis indicates that all viral polymerases whose structures are known have emerged from a common ancestor (Fig. 5.2). Furthermore, dsRNA *Cystoviridae* and (+)ssRNA *Flaviviridae* are closely related, as indicated by the structural similarity of their RdRPs (Fig. 5.2). The RdRPs from (+)ssRNA *Picorna-* and *Caliciviridae* are closely related and have emerged from a somewhat separate path in relation to (+)ssRNA *Flaviviridae*. Strikingly, the dsRNA reovirus and Φ6 polymerases are not particularly related, indicating that, based on the polymerase structure, Φ6 is no more closer to reovirus than it is to the *Flaviviridae* viruses. These results seem to be in line with the "RNA protovirus" model, from which both dsRNA and (+)ssRNA viruses would have evolved. Divergent pathways would then have originated the different viral families, with the dsRNA *Cystoviridae*

and (+)ssRNA *Flaviviridae* being closer to that ancestral viral form, whilst (+)ssRNA *Picorna-* and *Caliciviridae* viral families would have emerged later.

As structures of vRdRPs from other viral families become available, together with attempts to obtain complexes of these enzymes at several stages in reaction, the mechanism of polymerisation and the specific regulation and control mechanisms different family virus adopted are expected to be further elucidated. In particular, great interest surrounds the determination of the structure of a viral RNA-dependent RNA polymerase from a (-)ssRNA virus, such as influenza virus, that have so far remained elusive. It will most certainly shed new light into the evolution of vRdRPs and viral families, allowing a broader understanding of polymerisation and viral phylogeny.

Multisubunit cellular DdRPs and QDE-1 are apparently unrelated to the other folds and they must have arisen separately far back in evolutionary history. Previous sequence analysis studies (Iyer *et al.*, 2003) which suggested an evolutionary link between DdRPs and cellular RdRPs, recognised the DPBB containing the DxDxD conserved motif in cellular RdRPs. QDE-1 structural studies, identifying a complete active site cleft formed by two β-barrels equivalent to those described in DdRPs on a single polypeptide chain (DPBB1 and DPBB2), indicate a much stronger link, as shown in our structure based evolutionary analysis (Fig. 5.2). Moreover, the two DPBBs in QDE-1 are structurally related and would have originated from a common ancestral, catalytically active DPBB (Fig. 5.2). It is likely that gene duplication would have generated the two DPBBs, with one (DPBB2) retaining the catalytic aspartates and the other (DPBB1) acquiring several positively charged residues and more elaborate inserts. Furthermore, the metal ion coordination

pattern by three neighbouring residues is also conserved between the multisubunit DdRPs and QDE-1. In these enzymes, all three catalytic aspartates are present in the same motif – DxDxD - and coordinate the magnesium ion both in the apo form of the enzyme as well as during the reaction (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Armache *et al.*, 2003; Artsimovitch *et al.*, 2004; Westover *et al.*, 2004; Armache *et al.*, 2005). The other RNA polymerase folds are characterised by having their two catalytic aspartates within different conserved motifs - A and C - and a third non-catalytic aspartate in motif C that stabilizes the initiation complex (Delarue *et al.*, 1990; Steitz, 1998). This suggests that the DxDxD motif and subsequent mode of metal coordination are a signature of the multisubunit DdRP fold, possibly related to the need for tight coordination to guarantee correct stabilization of the template and complexes formed during polymerization. These observations strongly suggest that cellular RdRPs and multisubunit DdRPs share a common ancestor and that QDE-1 and related cRdRPs are closer to that ancestor than RNApolII-like polymerases. An evolutionary path for RNA polymerases evolution can therefore be proposed (Fig. 5.3).

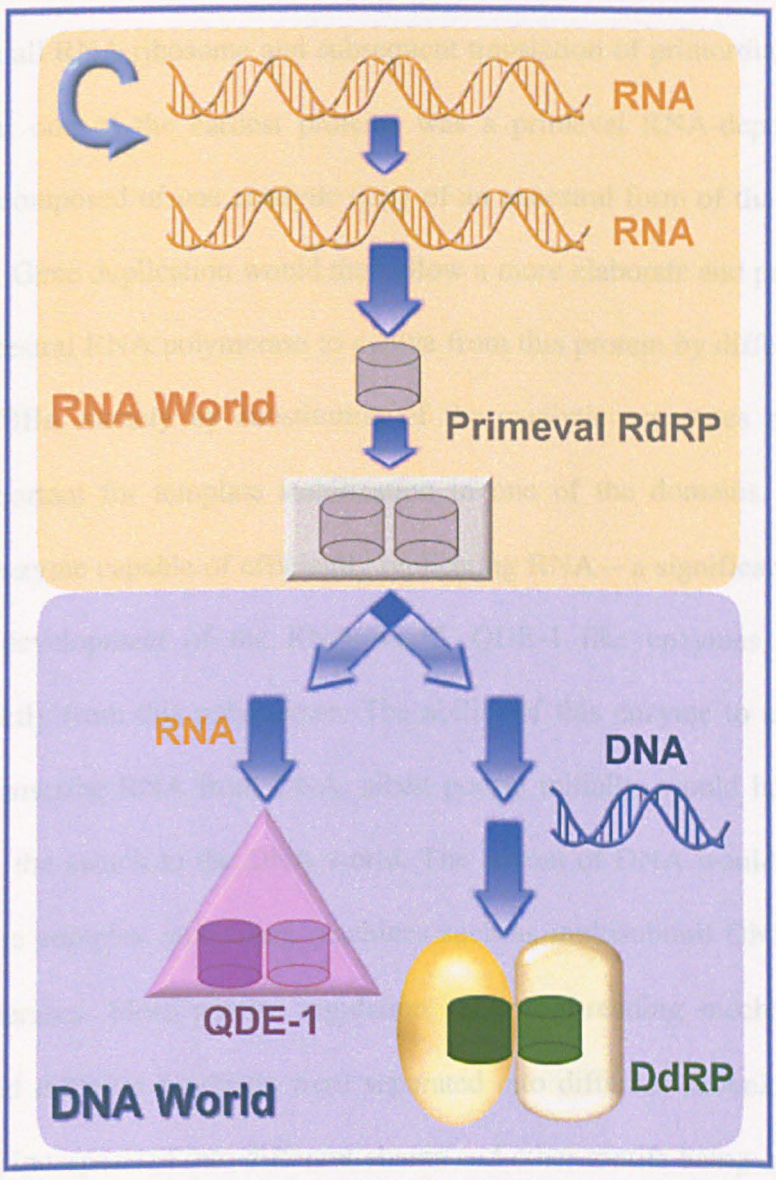


Figure 5.3. **Model for the evolution of cellular RdRPs and multisubunit DdRPs.**

In the original all RNA world, RNA would have to replicate itself, until the advent of an all RNA ribosome and subsequent translation of primordial proteins. It is likely that one of the earliest proteins was a primeval RNA-dependent RNA polymerase composed of one catalytic copy of an ancestral form of the DPBB on a single chain. Gene duplication would then allow a more elaborate and possibly more effective ancestral RNA polymerase to evolve from this protein by differentiation of the two DPBBs, namely by substitution of the catalytic aspartates by the basic residues important for template stabilization in one of the domains. This would provide an enzyme capable of efficiently replicating RNA – a significant step in the complexity development of the RNA world. QDE-1 like enzymes would have evolved directly from this polymerase. The ability of this enzyme to also replicate DNA and transcribe RNA from DNA, albeit poorly initially, would have been the key to allow the switch to the DNA world. The advent of DNA would presumably result in more complex replicating machines such as multisubunit DNA-dependent RNA polymerases. More precise regulation and proof-reading mechanisms were necessary and different functions were separated into different subunits, with each DPBB being incorporated into different chains and other motifs being acquired. The need for further specificity as the complexity of polymerisation paths increased resulted in the addition of extra motifs and other regulating subunits to the machinery, giving rise to the elaborate multisubunit complexes seen nowadays.

Separately, the "RNA protovirus" would have evolved a different fold to catalyse RNA-dependent RNA polymerisation. It can be speculated that the cellular "right hand" fold polymerases observed nowadays could have been a result of viruses infecting early primitive biological systems that already possessed a

polymerase containing the DPBB architecture and recruited the protovirus polymerase to other functions in the cell.

The evolutionary link between DNA-dependent RNA polymerases and QDE-1, an RdRP involved in RNA silencing mechanisms with homologues in several organisms, from plants to *C. elegans*, also suggests that there are evolutionary links between the primordial RNA mechanism described above and current RNA silencing pathways. Once DNA replaced RNA as the molecule of genetic information storage, it is likely that RNA polymerisation mechanisms became less relevant to the core replication of the cell. However, some of its components were recruited for different pathways. Thus, a plausible hypothesis is that at least part of the RNA silencing machinery in some organisms is reminiscent of a complex set of reactions involving RNA polymerization and translation, together with several control mechanisms that assured the correct transmission of RNA genomic information in an all RNA world. As further studies of the RNA silencing mechanisms become available and understanding of its intricate pathways and the proteins involved increases, it should become clearer how RNAi might have evolved and what primordial pathways have been recruited for new functions in a DNA world.

References

Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. and Miyano, M. (1999) Crystal Structure of the RNA-Dependent RNA Polymerase of Hepatitis C Virus. *Structure Fold Des*, **7**, 1417-1426.

Alaoui-Lsmaili, M.H., Hamel, M., L'Heureux, L., Nicolas, O., Bilimoria, D., Labonte, P., Mounir, S. and Rando, R.F. (2000) The Hepatitis C Virus NS5B RNA-Dependent RNA Polymerase Activity and Susceptibility to Inhibitors Is Modulated by Metal Cations. *J Hum Virol*, **3**, 306-316.

Armache, K.J., Kettenberger, H. and Cramer, P. (2003) Architecture of Initiation-Competent 12-Subunit RNA Polymerase II. *Proc Natl Acad Sci U S A*, **100**, 6964-6968.

Armache, K.J., Mitterweger, S., Meinhart, A. and Cramer, P. (2005) Structures of Complete RNA Polymerase II and its subcomplex, Rpb4/7. *J Biol Chem*, **280**, 7131-7134.

Arnold, J.J., Ghosh, S.K. and Cameron, C.E. (1999) Poliovirus RNA-Dependent RNA Polymerase (3d(Pol)). Divalent Cation Modulation of Primer, Template, and Nucleotide Selection. *J Biol Chem*, **274**, 37060-37069.

Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M.N., Hosaka, T., Ochi, K., Yokoyama, S. and Vassylyev, D.G. (2004) Structural Basis for Transcription Regulation by Alarmone PPGPP. *Cell*, **117**, 299-310.

Artymiuk, P.J., Poirrette, A.R., Grindley, H.M., Rice, D.W. and Willett, P. (1994) A Graph-Theoretic Approach to the Identification of Three-Dimensional Patterns of Amino Acid Side-Chains in Protein Structures. *J Mol Biol*, **243**, 327-344.

Bamford, D.H., Palva, E.T. and Lounatmaa, K. (1976) Ultrastructure and Life Cycle of the Lipid-Containing Bacteriophage Φ 6. *J Gen Virol*, **32**, 249-259.

Bamford, D.H., Ojala, P.M., Frilander, M., Walin, L. and Bamford, J.K.H. (1995) Isolation, Purification, and Function of Assembly Intermediates and Subviral Particles of Bacteriophages Prd1 and Φ6. In Adolph, K.W. (ed.), *Methods in Molecular Genetics*. Academic Press, San Diego, Vol. 6, pp. 455-474.

Bamford, D.H., Burnett, R.M. and Stuart, D.I. (2002) Evolution of Viral Structure. *Theor Popul Biol*, **61**, 461-470.

Bamford, D.H. (2003) Do Viruses Form Lineages across Different Domains of Life? *Res Microbiol*, **154**, 231-236.

Baulcombe, D. (2002) RNA Silencing. *Curr Biol*, **12**, 82-84.

Baulcombe, D. (2004) RNA Silencing in Plants. *Nature*, **431**, 356-363.

Behrens, S.E., Tomei, L. and De Francesco, R. (1996) Identification and Properties of the RNA-Dependent RNA Polymerase of Hepatitis C Virus. *EMBO J*, **15**, 12-22.

Bell, S.D. and Jackson, S.P. (1998) Transcription in Archaea. *Cold Spring Harb Symp Quant Biol*, **63**, 41-51.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucl. Acids Res.*, **28**, 235-242.

Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001b) Role for a Bidentate Ribonuclease in the Initiation Step of RNA Interference. *Nature*, **409**, 363-366.

Biswal, B.K., Cherney, M.M., Wang, M., Chan, L., Yannopoulos, C.G., Bilimoria, D., Nicolas, O., Bedard, J. and James, M.N. (2005) Crystal Structures of the RNA-Dependent RNA Polymerase Genotype 2a of Hepatitis C Virus Reveal Two Conformations and Suggest Mechanisms of Inhibition by Non-Nucleoside Inhibitors. *J Biol Chem*, **280**, 18202-18210.

Blumenthal, T. and Carmichael, G.G. (1979) RNA Replication: Function and Structure of Qβ-Replicase. *Annu Rev Biochem*, **48**, 525-548.

Blumenthal, T. (1980) Q Beta Replicase Template Specificity: Different Templates Require Different GTP Concentrations for Initiation. *Proc Natl Acad Sci U S A*, **77**, 2601-2605.

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R. and Rey, F.A. (1999) Crystal Structure of the RNA-Dependent RNA Polymerase of Hepatitis C Virus. *Proc Natl Acad Sci U S A*, **96**, 13034-13039.

Bressanelli, S., Tomei, L., Rey, F.A. and De Francesco, R. (2002) Structural Analysis of the Hepatitis C Virus RNA Polymerase in Complex with Ribonucleotides. *J Virol*, **76**, 3482-3492.

Brown, J., *et al.* (2003) A Procedure for Setting up High-Throughput Nanolitre Crystallization Experiments. II. Crystallization Results. *JouRNAl of Applied Crystallography*, **36**, 315-318.

Bruenn, J.A. (1991) Relationships among the Positive Strand and Double-Strand RNA Viruses as Viewed through Their RNA-Dependent RNA Polymerases. *Nucleic Acids Res*, **19**, 217-226.

Bruker. (2001) Xprep. Bruker AXS Inc., Madison, Wisconsin, USA.

Brunger, A.T., *et al.* (1998) Crystallography & Nmr System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr D Biol Crystallogr*, **54 (Pt 5)**, 905-921.

Bushnell, D.A., Cramer, P. and Kornberg, R.D. (2001) Selenomethionine Incorporation in Saccharomyces Cerevisiae RNA Polymerase II. *Structure (Camb)*, **9**, 11-14.

Butcher, S.J., Dokland, T., Ojala, P.M., Bamford, D.H. and Fuller, S.D. (1997) Intermediates in the Assembly Pathway of the Double-Stranded RNA Virus Φ6. *EMBO J*, **16**, 4477-4487.

Butcher, S.J., Makeyev, E.V., Grimes, J.M., Stuart, D.I. and Bamford, D.H. (2000) Crystallization and Preliminary X-Ray Crystallograϕc Studies on the Bacteriophage Φ6 RNA-Dependent RNA Polymerase. *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 11)**, 1473-1475.

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H. and Stuart, D.I. (2001) A Mechanism for Initiating RNA-Dependent RNA Polymerization. *Nature*, **410**, 235-240.

Castillo, R.M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T.L. and Murzin, A.G. (1999) A Six-Stranded Double-Psi [Beta] Barrel Is Shared by Several Protein Superfamilies. *Structure*, **7**, 227-236.

Catalanotto, C., Azzalin, G., Macino, G. and Cogoni, C. (2000) Gene Silencing in Worms and Fungi. *Nature*, **404**, 245.

Catalanotto, C., Pallotta, M., ReFalo, P., Sachs, M.S., Vayssie, L., Macino, G. and Cogoni, C. (2004) Redundancy of the Two Dicer Genes in Transgene-Induced Posttranscriptional Gene Silencing in Neurospora Crassa. *Mol Cell Biol*, **24**, 2536-2545.

Chicas, A., Forrest, E.C., Sepich, S., Cogoni, C. and Macino, G. (2005) Small Interfering RNAs That Trigger Posttranscriptional Gene Silencing Are Not Required for the Histone H3 Lys9 Methylation Necessary for Transgenic Tandem Repeat Stabilization in Neurospora Crassa. *Mol Cell Biol*, **25**, 3793-3801.

Cho, M.W., Richards, O.C., Dmitrieva, T.M., Agol, V. and Ehrenfeld, E. (1993) RNA Duplex Unwinding Activity of Poliovirus RNA-Dependent RNA Polymerase 3dpol. *J Virol*, **67**, 3010-3018.

Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevear, D.C. and Rossmann, M.G. (2004) The Structure of the RNA-Dependent RNA Polymerase from Bovine Viral Diarrhea Virus Establishes the Role of GTP in De novo Initiation. *Proc Natl Acad Sci U S A*, **101**, 4425-4430. Epub 2004 Mar 4419.

Cogoni, C. and Macino, G. (1999a) Gene Silencing in Neurospora Crassa Requires a Protein Homologous to RNA-Dependent RNA Polymerase. *Nature*, **399**, 166-169.

Cogoni, C. and Macino, G. (1999b) Posttranscriptional Gene Silencing in Neurospora by a Recq DNA Helicase. *Science*, **286**, 2342-2344.

Collaborative Computational Project, N. (1994) *Acta Cryst. D*, **50**, 760-763.

Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001) Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution. *Science*, **292**, 1863-1876.

Cramer, P. (2004) RNA Polymerase II Structure: From Core to Functional Complexes. *Curr Opin Genet Dev*, **14**, 218-226.

Dauter, Z. (2002) New Approaches to High-Throughput Phasing. *Curr Opin Struct Biol*, **12**, 674-678.

Davies, J.F., 2nd, Almassy, R.J., Hostomska, Z., Ferre, R.A. and Hostomsky, Z. (1994) 2.3Å Crystal Structure of the Catalytic Domain of DNA Polymerase Beta. *Cell*, **76**, 1123-1133.

Day, L.A. and Mindich, L. (1980) The Molecular Weight of Bacteriophage Φ 6 and Its Nucleocapsid. *Virology*, **103**, 376-385.

de Graaff, M., Houwing, C.J., Lukacs, N. and Jaspars, E.M. (1995) RNA Duplex Unwinding Activity of Alfalfa Mosaic Virus RNA-Dependent RNA Polymerase. *FEBS Lett*, **371**, 219-222.

de Haas, F., Paatero, A.O., Mindich, L., Bamford, D.H. and Fuller, S.D. (1999) A Symmetry Mismatch at the Site of RNA Packaging in the Polymerase Complex of dsRNA Bacteriophage Φ6. *J Mol Biol*, **294**, 357-372.

de la Fortelle, E. and Bricogne, B. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology*, **276**, 472-494.

DeLano, W.L. (2004) The Pymol Molecular Graphics System. DeLano Scientific LLC, San Carlos.

Delarue, M., Poch, O., Tordo, N., Moras, D. and Argos, P. (1990) An Attempt to Unify the Structure of Polymerases. *Protein Eng*, **3**, 461-467.

Diprose, J.M., *et al.* (2001) Translocation Portals for the Substrates and Products of a Viral Transcription Complex: The Bluetongue Virus Core. *EMBO J*, **20**, 7229-7239.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal Structure of a Bacteriophage T7 DNA Replication Complex at 2.2Å Resolution. *Nature*, **391**, 251-258.

Drenth, J. (1999) Principles of Protein X-Ray Crystallography. Springer, New York.

Ebright, R.H. (2000) RNA Polymerase: Structural Similarities between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *J Mol Biol*, **304**, 687-698.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001a) Duplexes of 21-Nucleotide RNAs Mediate RNA Interference in Cultured Mammalian Cells. *Nature*, **411**, 494-498.

Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001b) RNA Interference Is Mediated by 21- and 22-Nucleotide RNAs. *Genes Dev*, **15**, 188-200.

Emori, Y., Iba, H. and Okada, Y. (1983) Transcriptional Regulation of Three Double-Stranded RNA Segments of Bacteriophage Φ 6 in Vitro. *J Virol*, **46**, 196-203.

Emsley, P. and Cowtan, K. (2004) Coot: Model-Building Tools for Molecular Graφcs. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.

Esnouf, R.M. (1997) Polyalanine Reconstruction from CALPHA Positions Using the Program CALPHA Can Aid Initial Phasing by Molecular Replcament Procedures. *Acta Crystallogr D*, **53**, 665-672.

Esnouf, R.M. (1999) Further Additions to Molscript Version 1.4, Including Reading and Contouring of Electron-Density Maps. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 4)**, 938-940.

Evans, G. and Pettifer, R.F. (2001) *J. Appl. Cryst*, **34**, 82-86.

Ewen, M.E. and Revel, H.R. (1990) RNA-Protein Complexes Responsible for Replication and Transcription of the Double-Stranded RNA Bacteriophage Φ 6. *Virology*, **178**, 509-519.

Felsenstein, J. (1989) Phylip - Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164-166.

Ferrari, E., Wright-Minogue, J., Fang, J.W., Baroudy, B.M., Lau, J.Y. and Hong, Z. (1999) Characterization of Soluble Hepatitis C Virus RNA-Dependent RNA Polymerase Expressed in Escherichia Coli. *J Virol*, **73**, 1649-1654.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E. and Verdaguer, N. (2004) Structure of Foot-and-Mouth Disease Virus RNA-Dependent RNA Polymerase and Its Complex with a Template-Primer RNA. *J Biol Chem*, **279**, 47212-47221.

Forrest, E.C., Cogoni, C. and Macino, G. (2004) The RNA-Dependent RNA Polymerase, QDE-1, Is a Rate-Limiting Factor in Post-Transcriptional Gene Silencing in Neurospora Crassa. *Nucleic Acids Res*, **32**, 2123-2128.

Freitag, M., Lee, D.W., Kothe, G.O., Pratt, R.J., Aramayo, R. and Selker, E.U. (2004) DNA Methylation Is Independent of RNA Interference in Neurospora. *Science*, **304**, 1939.

Frilander, M. and Bamford, D.H. (1995) In Vitro Packaging of the Single-Stranded RNA Genomic Precursors of the Segmented Double-Stranded RNA Bacteriophage Φ 6: The Three Segments Modulate Each Other's Packaging Efficiency. *J Mol Biol*, **246**, 418-428.

Glusker, J.P., Katz, A.K., Bock, C.W. (1999) Metal Ions in Biological Systems. *The Rigaku JouRNAl*, **16**, 8-16.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. and Kornberg, R.D. (2001) Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3Å Resolution. *Science*, **292**, 1876-1882.

Gottlieb, P., Metzger, S., Romantschuk, M., Carton, J., Strassman, J., Bamford, D.H., Kalkkinen, N. and Mindich, L. (1988) Nucleotide Sequence of the Middle dsRNA Segment of Bacteriophage Φ 6: Placement of the Genes of Membrane-Associated Proteins. *Virology*, **163**, 183-190.

Gottlieb, P., Strassman, J., Qiao, X., Frilander, M., Frucht, A. and Mindich, L. (1992) In Vitro Packaging and Replication of Individual Genomic Segments of Bacteriophage Φ 6 RNA. *J Virol*, **66**, 2611-2616.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Zientara, S., Mertens, P.P. and Stuart, D.I. (1998) The Atomic Structure of the Bluetongue Virus Core. *Nature*, **395**, 470-478.

Hansen, J.L., Long, A.M. and Schultz, S.C. (1997) Structure of the RNA-Dependent RNA Polymerase of Poliovirus. *Structure*, **5**, 1109-1122.

Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (Mad): A Vehicle for Direct Determination of Three-Dimensional Structure. *EMBO J*, **9**, 1665-1672.

Hendrickson, W.A. (1991) Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science*, **254**, 51-58.

Hendrickson, W.A. and Ogata, C.M. (1997) Phase Determination from Multiwavelength Anomalous Diffraction Measurements. *Methods in Enzymology*, **276**, 494-523.

Holm, L. and Sander, C. (1998) Touring Protein Fold Space with Dali/FSSP. *Nucleic Acids Res*, **26**, 316-319.

Hong, Z., Cameron, C.E., Walker, M.P., Castro, C., Yao, N., Lau, J.Y. and Zhong, W. (2001) A Novel Mechanism to Ensure Terminal Initiation by Hepatitis C Virus NS5B Polymerase. *Virology*, **285**, 6-11.

Hoogstraten, D., Qiao, X., Sun, Y., Hu, A., Onodera, S. and Mindich, L. (2000) Characterization of Φ8, a Bacteriophage Containing Three Double-Stranded RNA Genomic Segments and Distantly Related to Φ6. *Virology*, **272**, 218-224.

Huang, H., Chopra, R., Verdine, G.L. and Harrison, S.C. (1998) Structure of a Covalently Trapped Catalytic Complex of HIV-1 Reverse Transcriptase: Implications for Drug Resistance. *Science*, **282**, 1669-1675.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. and Zamore, P.D. (2001) A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the Let-7 Small Temporal RNA. *Science*, **293**, 834-838.

Iyer, L.M., Koonin, E.V. and Aravind, L. (2003) Evolutionary Connection between the Catalytic Subunits of DNA-Dependent RNA Polymerases and Eukaryotic RNA-Dependent RNA Polymerases and the Origin of RNA Polymerases. *BMC Struct Biol*, **3**, 1.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard. (1991) Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr A*, **47 (Pt 2)**, 110-119.

Joyce, C.M. and Steitz, T.A. (1995) Polymerase Structures and Function: Variations on a Theme? *J Bacteriol*, **177**, 6321-6329.

Juuti, J.T. and Bamford, D.H. (1995) RNA Binding, Packaging and Polymerase Activities of the Different Incomplete Polymerase Complex Particles of dsRNA Bacteriophage Φ 6. *J Mol Biol*, **249**, 545-554.

Kao, C.C. and Sun, J.H. (1996) Initiation of Minus-Strand RNA Synthesis by the Brome Mosaicvirus RNA-Dependent RNA Polymerase: Use of Oligoribonucleotide Primers. *J Virol*, **70**, 6826-6830.

Kao, C.C., Del Vecchio, A.M. and Zhong, W. (1999) De novo Initiation of RNA Synthesis by a Recombinant Flaviviridae RNA-Dependent RNA Polymerase. *Virology*, **253**, 1-7.

Kao, C.C., Singh, P. and Ecker, D.J. (2001) De novo Initiation of Viral RNA-Dependent RNA Synthesis. *Virology*, **287**, 251-260.

Kato, A., Akamatsu, Y., Sakuraba, Y. and Inoue, H. (2004) The Neurospora Crassa Mus-19 Gene Is Identical to the QDE-3 Gene, Which Encodes a Recq Homologue and Is Involved in Recombination Repair and Postreplication Repair. *Curr Genet*, **45**, 37-44.

Kettenberger, H., Armache, K.J. and Cramer, P. (2004) Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with Ntp and TfIIs. *Mol Cell*, **16**, 955-965.

Kim, M.J., Zhong, W., Hong, Z. and Kao, C.C. (2000) Template Nucleotide Moieties Required for De novo Initiation of RNA Synthesis by a Recombinant Viral RNA-Dependent RNA Polymerase. *J Virol*, **74**, 10312-10322.

Kissinger, C.R., Gehlhaar, D.K. and Fogel, D.B. (1999) Rapid Automated Molecular Replacement by Evolutionary Search. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 2)**, 484-491.

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A. (1992) Crystal Structure at 3.5Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor. *Science*, **256**, 1783-1790.

Koonin, E.V. (1992) Evolution of Double-Stranded RNA Viruses: A Case for Polyphyletic Origin from Different Groups of Positive-Stranded RNA Viruses. *Semin Virol*, **3**.

Krissinel, E. and Henrick, K. (2004) Secondary-Structure Matching (Ssm), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallographica Section D*, **60**, 2256-2268.

Lai, M.M. (1998) Cellular Factors in the Transcription and Replication of Viral RNA Genomes: A Parallel to DNA-Dependent RNA Transcription. *Virology*, **244**, 1-12.

Lai, M.M. (2005) RNA Replication without RNA-Dependent RNA Polymerase: Surprises from Hepatitis Delta Virus. *J Virol*, **79**, 7951-7958.

Lampio, A., Kilpelainen, I., Pesonen, S., Karhi, K., Auvinen, P., Somerharju, P. and Kaariainen, L. (2000) Membrane Binding Mechanism of an RNA Virus-Capping Enzyme. *J Biol Chem*, **275**, 37853-37859.

Lamzin, V.S. and Wilson, K.S. (1993) Automated Refinement of Protein Models. *Acta Crystallogr D Biol Crystallogr*, **49**, 129-147.

Larsson, A.M., Stahlberg, J. and Jones, T.A. (2002) Preparation and Crystallization of Selenomethionyl Dextranase from Penicillium Minioluteum Expressed in Pichia Pastoris. *Acta Crystallogr D Biol Crystallogr*, **58**, 346-348.

Larsson, A.M., Andersson, R., Stahlberg, J., Kenne, L. and Jones, T.A. (2003) Dextranase from Penicillium Minioluteum: Reaction Course, Crystal Structure, and Product Complex. *Structure (Camb)*, **11**, 1111-1121.

Laurila, M.R., Makeyev, E.V. and Bamford, D.H. (2002) Bacteriophage Φ6 RNA-Dependent RNA Polymerase: Molecular Details of Initiating Nucleic Acid Synthesis without Primer. *J Biol Chem*, **in press**.

Laurila, M.R., Salgado, P.S., Makeyev, E.V., Nettelship, J., Stuart, D.I., Grimes, J.M. and Bamford, D.H. (2005a) Gene Silencing Pathway RNA-Dependent RNA Polymerase of Neurospora Crassa: Yeast Expression and Crystallization of Selenomethionated QDE-1 Protein. *J Struct Biol*, **149**, 111-115.

Laurila, M.R., Salgado, P.S., Stuart, D.I., Grimes, J.M. and Bamford, D.H. (2005B) Back-Priming Mode of Φ6 RNA-Dependent RNA Polymerase. *J Gen Virol*, **86**, 521-526.

Lee, T.I. and Young, R.A. (2000) Transcription of Eukaryotic Protein-Coding Genes. *Annu Rev Genet*, **34**, 77-137.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J. and Carthew, R.W. (2004) Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the SiRNA/MiRNA Silencing Pathways. *Cell*, **117**, 69-81.

Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F. and Weber, P.C. (1999) Crystal Structure of the RNA-Dependent RNA Polymerase from

Hepatitis C Virus Reveals a Fully Encircled Active Site. *Nat Struct Biol*, **6**, 937-943.

Leveque, V.J., Johnson, R.B., Parsons, S., Ren, J., Xie, C., Zhang, F. and Wang, Q.M. (2003) Identification of a C-Terminal Regulatory Motif in Hepatitis C Virus RNA-Dependent RNA Polymerase: Structural and Biochemical Analysis. *J Virol*, **77**, 9020-9028.

Li, Y., Kong, Y., Korolev, S. and Waksman, G. (1998) Crystal Structures of the Klenow Fragment of Thermus Aquaticus DNA Polymerase I Complexed with Deoxyribonucleoside Triphosphates. *Protein Sci*, **7**, 1116-1123.

Lingel, A., Simon, B., Izaurralde, E. and Sattler, M. (2004) Nucleic Acid 3'-End Recognition by the Argonaute2 Paz Domain. *Nat Struct Mol Biol*, **11**, 576-577.

Lippman, Z. and Martienssen, R. (2004) The Role of RNA Interference in Heterochromatic Silencing. *Nature*, **431**, 364-370.

Lopez Vazquez, A.L., Martin Alonso, J.M. and Parra, F. (2001) Characterisation of the RNA-Dependent RNA Polymerase from Rabbit Hemorrhagic Disease Virus Produced in Escherichia Coli. *Arch Virol*, **146**, 59-69.

Losick, R. and Chamberlin, M. (1976) *RNA Polymerase*, Cold Spring Harbor Laboratory, NY.

Love, R.A., Maegley, K.A., Yu, X., Ferre, R.A., Lingardo, L.K., Diehl, W., Parge, H.E., Dragovich, P.S. and Fuhrman, S.A. (2004) The Crystal Structure of the RNA-Dependent RNA Polymerase from Human Rhinovirus: A Dual Function Target for Common Cold Antiviral Therapy. *Structure (Camb)*, **12**, 1533-1544.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T. and Patel, D.J. (2005) Structural Basis for 5'-End-Specific Recognition of Guide RNA by the A. Fulgidus Piwi Protein. *Nature*, **434**, 666-670.

Makeyev, E.V. and Bamford, D.H. (2000a) Replicase Activity of Purified Recombinant Protein P2 of Double-Stranded RNA Bacteriophage Φ6. *EMBO J*, **19**, 124-133.

Makeyev, E.V. and Bamford, D.H. (2000b) The Polymerase Subunit of a dsRNA Virus Plays a Central Role in the Regulation of Viral RNA Metabolism. *EMBO J*, **19**, 6275-6284.

Makeyev, E.V. (2001) Phd Thesis: RNA-Dependent RNA Polymerase of Bacteriophage Φ6.

Makeyev, E.V. and Bamford, D.H. (2001) Primer-Independent RNA Sequencing with Bacteriophage Φ6 RNA Polymerase and Chain Terminators. *RNA*, **7**, 774-781.

Makeyev, E.V. and Bamford, D.H. (2002) Cellular RNA-Dependent RNA Polymerase Involved in Posttranscriptional Gene Silencing Has Two Distinct Activity Modes. *Mol Cell*, **10**, 1417-1427.

Makeyev, E.V. and Grimes, J.M. (2004) RNA-Dependent RNA Polymerases of dsRNA Bacteriophages. *Virus Res*, **101**, 45-55.

Mancini, E.J., Kainov, D.E., Grimes, J.M., Tuma, R., Bamford, D.H. and Stuart, D.I. (2004) Atomic Snapshots of an RNA Packaging Motor Reveal Conformational Changes Linking ATP Hydrolysis to RNA Translocation. *Cell*, **118**, 743-755.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. (2002) Single-Stranded Antisense SiRNAs Guide Target RNA Cleavage in RNAi. *Cell*, **110**, 563-574.

Martinez, J. and Tuschl, T. (2004) RISC Is a 5' Phosphomonoester-Producing RNA Endonuclease. *Genes Dev*, **18**, 975-980.

Mathews, B.W. (1968) Solvent Content of Protein Crystals. *J Mol Biol*, **33**, 491-497.

McGraw, T., Mindich, L. and Frangione, B. (1986) Nucleotide Sequence of the Small Double-Stranded RNA Segment of Bacteriophage Φ 6: Novel Mechanism of Natural Translational Control. *J Virol*, **58**, 142-151.

McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*. John Wiley and Sons, Inc., New York.

Mello, C.C. and Conte, D., Jr. (2004) Revealing the World of RNA Interference. *Nature*, **431**, 338-342.

Merritt, E.A.a.B., D.J. (1997) Raster3d: Photorealistic Molecular Graφcs. *Macromolecular Crystallography*, **277**, 505-524.

Mindich, L. (1988) Bacteriophage Φ 6: A Unique Virus Having a Lipid-Containing Membrane and a Genome Composed of Three dsRNA Segments. *Adv Virus Res*, **35**, 137-176.

Mindich, L. and Bamford, D.H. (1988) Lipid-Containing Bacteriophages. In Calendar, R. (ed.), *The Bacteriophages*. Plenum Press, New York, Vol. 2, pp. 475-520.

Mindich, L., Nemhauser, I., Gottlieb, P., Romantschuk, M., Carton, J., Frucht, S., Strassman, J., Bamford, D.H. and Kalkkinen, N. (1988) Nucleotide Sequence of the Large Double-Stranded RNA Segment of Bacteriophage Φ 6: Genes Specifying the Viral Replicase and Transcriptase. *J Virol*, **62**, 1180-1185.

Mindich, L., Qiao, X., Qiao, J., Onodera, S., Romantschuk, M. and Hoogstraten, D. (1999) Isolation of Additional Bacteriophages with Genomes of Segmented Double-Stranded RNA. *J Bacteriol*, **181**, 4505-4508.

Motamedi, M.R., Verdel, A., Colmenares, *S.U., Gerber, S.A., Gygi, S.P. and* Moazed, D. (2004) Two RNAi Complexes, RITS and Rdrc, Physically Interact and Localize to Noncoding Centromeric RNAs. *Cell*, **119**, 789-802.

Mourrain, P., *et al.* (2000) Arabidopsis Sgs2 and Sgs3 Genes Are Required for Posttranscriptional Gene Silencing and Natural Virus Resistance. *Cell*, **101**, 533-542.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

Nason, E.L., Rothagel, R., Mukherjee, S.K., Kar, A.K., Forzan, M., Prasad, B.V. and Roy, P. (2004) Interactions between the Inner and Outer Capsids of Bluetongue Virus. *J Virol*, **78**, 8059-8067.

Navaza, J. (1994) Amore: An Automated Package for Molecular Replacement. *Acta Crystallografca Section A*, **50**, 157-163.

Ng, K.K., Cherney, M.M., Vazquez, A.L., Machin, A., Alonso, J.M., Parra, F. and James, M.N. (2002) Crystal Structures of Active and Inactive Conformations of a Caliciviral RNA-Dependent RNA Polymerase. *J Biol Chem*, **277**, 1381-1387.

Ng, K.K., Pendas-Franco, N., Rojo, J., Boga, J.A., Machin, A., Alonso, J.M. and Parra, F. (2004) Crystal Structure of Norwalk Virus Polymerase Reveals the Carboxyl Terminus in the Active Site Cleft. *J Biol Chem*, **279**, 16638-16645.

O'Farrell, D., Trowbridge, R., Rowlands, D. and Jäger, J. (2003) Substrate Complexes of Hepatitis C Virus RNA Polymerase (Hc-J4): Structural Evidence for Nucleotide Import and De-Novo Initiation. *JouRNAl of Molecular Biology*, **326**, 1025-1035.

Okai, Y. (1982) Calcium Effects on Free and Chromatin-Bound RNA Polymerase II Reactions. *FEBS Lett*, **140**, 139-141.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Structure of Large Fragment of Escherichia Coli DNA Polymerase I Complexed with DTmp. *Nature*, **313**, 762-766.

Onodera, S., Qiao, X., Qiao, J. and Mindich, L. (1998) Isolation of a Mutant That Changes Genomic Packaging Specificity in Φ6. *Virology*, **252**, 438-442.

O'Reilly, E.K. and Kao, C.C. (1998) Analysis of RNA-Dependent RNA Polymerase Structure and Function as Guided by Known Polymerase Structures and Computer Predictions of Secondary Structure. *Virology*, **252**, 287-303.

Otwinowski, Z. and Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Macromolecular Crystallography*, **276**, 307-326.

Paul, A.V., van Boom, J.H., Filippov, D. and Wimmer, E. (1998) Protein-Primed RNA Synthesis by Purified Poliovirus RNA Polymerase. *Nature*, **393**, 280-284.

Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. and Kraut, J. (1994) Structures of TeRNAry Complexes of Rat DNA Polymerase Beta, a DNA Template-Primer, and DdCTP. *Science*, **264**, 1891-1903.

Perrakis, A., Harkiolaki, M., Wilson, K.S. and Lamzin, V.S. (2001) Arp/Warp and Molecular Replacement. *Acta Crystallogr D Biol Crystallogr*, **57**, 1445-1450.

Pickford, A., Braccini, L., Macino, G. and Cogoni, C. (2003) The QDE-3 Homologue Recq-2 Co-Operates with QDE-3 in DNA Repair in Neurospora Crassa. *Curr Genet*, **42**, 220-227.

Pickford, A.S., Catalanotto, C., Cogoni, C. and Macino, G. (2002) Quelling in Neurospora Crassa. *Adv Genet*, **46**, 277-303.

Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989) Identification of Four Conserved Motifs among the RNA-Dependent Polymerase Encoding Elements. *EMBO J*, **8**, 3867-3874.

Poranen, M. (2000) Bacteriophage Φ6 Nucleocapsid Assembly and Membrane Voltage-Dependent Interaction with Host Plasma Membrane. *Academic dissertation*, **University of Helsinki.**

Poranen, M.M., Paatero, A.O., Tuma, R. and Bamford, D.H. (2001) Self-Assembly of a Viral Molecular Machine from Purified Protein and RNA Constituents. *Mol Cell*, **7**, 845-854.

Prasad, B.V., Rothnagel, R., Zeng, C.Q., Jakana, J., Lawton, J.A., Chiu, W. and Estes, M.K. (1996) Visualization of Ordered Genomic RNA and Localization of Transcriptional Complexes in Rotavirus. *Nature*, **382**, 471-473.

Qiao, X., Qiao, J. and Mindich, L. (1997) Stoichiometric Packaging of the Three Genomic Segments of Double-Stranded RNA Bacteriophage Φ6. *Proc Natl Acad Sci U S A*, **94**, 4074-4079.

Qiao, X., Qiao, J., Onodera, S. and Mindich, L. (2000) Characterization of Φ 13, a Bacteriophage Related to Φ 6 and Containing Three dsRNA Genomic Segments. *Virology*, **275**, 218-224.

Ramagopal, U.A., Dauter, M. and Dauter, Z. (2003) Sad Manganese in Two Crystal Forms of Glucose Isomerase. *Acta Crystallogr D Biol Crystallogr*, **59**, 868-875.

Ranjith-Kumar, C.T., Kim, Y.C., Gutshall, L., Silverman, C., Khandekar, S., Sarisky, R.T. and Kao, C.C. (2002) Mechanism of De novo Initiation by the Hepatitis C Virus RNA-Dependent RNA Polymerase: Role of Divalent Metals. *J Virol*, **76**, 12513-12525.

Ranjith-Kumar, C.T., Gutshall, L., Kim, M.J., Sarisky, R.T. and Kao, C.C. (2002b) Requirements for De novo Initiation of RNA Synthesis by Recombinant Flaviviral RNA-Dependent RNA Polymerases. *J Virol*, **76**, 12526-12536.

Ranjith-Kumar, C.T., Gutshall, L., Sarisky, R.T. and Kao, C.C. (2003a) Multiple Interactions within the Hepatitis C Virus RNA Polymerase Repress Primer-Dependent RNA Synthesis. *J Mol Biol*, **330**, 675-685.

Ranjith-Kumar, C.T., *et al.* (2003b) Enzymatic Activities of the Gb Virus-B RNA-Dependent RNA Polymerase. *Virology*, **312**, 270-280.

Salgado, P.S., Makeyev, E.V., Butcher, S.J., Bamford, D.H., Stuart, D.I. and Grimes, J.M. (2004) The Structural Basis for RNA Specificity and Ca2+ Inhibition of an RNA-Dependent RNA Polymerase. *Structure (Camb)*, **12**, 307-316.

Salgado, P.S., Walsh, M.A., Laurila, M.R., Stuart, D.I. and Grimes, J.M. (2005) Going Soft and Sad with Manganese. *Acta Crystallogr D Biol Crystallogr*, **61**, 108-111.

Sargent, M.D. and Borsa, J. (1984) Effects of Ca2+ and Mg2+ on the Switch-on of Transcriptase Function in Reovirus in Vitro. *Can J Biochem Cell Biol*, **62**, 162-169.

Saris, N.E., Mervaala, E., Karppanen, H., Khawaja, J.A. and Lewenstam, A. (2000) Magnesium. An Update on Physiological, Clinical and Analytical Aspects. *Clin Chim Acta*, **294**, 1-26.

Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. and Kraut, J. (1994) Crystal Structure of Rat DNA Polymerase Beta: Evidence for a Common Polymerase Mechanism. *Science*, **264**, 1930-1935.

Schiebel, W., Haas, B., Marinkovic, S., Klanner, A. and Sanger, H.L. (1993a) RNA-Directed RNA Polymerase from Tomato Leaves. II. Catalytic in Vitro Properties. *J Biol Chem*, **268**, 11858-11867.

Schiebel, W., Haas, B., Marinkovic, S., Klanner, A. and Sanger, H.L. (1993b) RNA-Directed RNA Polymerase from Tomato Leaves. I. Purification and Physical Properties. *J Biol Chem*, **268**, 11851-11857.

Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sanger, H.L. and Wassenegger, M. (1998) Isolation of an RNA-Directed RNA Polymerase-Specific CDNA Clone from Tomato. *Plant Cell*, **10**, 2087-2101.

Schneider, T.R. (2004) Domain Identification by Iterative Analysis of Error-Scaled Difference Distance Matrices. *Acta Crystallogr D Biol Crystallogr*, **60**, 2269-2275.

Schramke, V., Sheedy, D.M., Denli, A.M., Bonila, C., Ekwall, K., Hannon, G.J. and Allshire, R.C. (2005) RNA-Interference-Directed Chromatin Modification Coupled to RNA Polymerase II Transcription. *Nature*, **19**, 19.

Sharp, P.A. (2001) RNA Interference--2001. *Genes Dev*, **15**, 485-490.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H. and Fire, A. (2001) On the Role of RNA Amplification in dsRNA-Triggered Gene Silencing. *Cell*, **107**, 465-476.

Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science*, **305**, 1434-1437.

Sousa, R., Chung, Y.J., Rose, J.P. and Wang, B.C. (1993) Crystal Structure of Bacteriophage T7 RNA Polymerase at 3.3Å Resolution. *Nature*, **364**, 593-599.

Steitz, T.A. (1993) DNA and RNA-Dependent DNA Polymerases. *Curr Opin Struct Biol*, **3**, 31-38.

Steitz, T.A. (1998) A Mechanism for All Polymerases. *Nature*, **391**, 231-232.

Stuart, D.I., Levine, M., Muirhead, H. and Stammers, D.K. (1979) Crystal Structure of Cat Muscle Pyruvate Kinase at a Resolution of 2.6 A. *J Mol Biol*, **134**, 109-142.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D. and Grewal, S.I. (2005) RNA-Dependent RNA Polymerase Is an Essential Component of a Self-Enforcing Loop Coupling Heterochromatin Assembly to SiRNA Production. *Proc Natl Acad Sci U S A*, **102**, 152-157.

Sun, J.H., Adkins, S., Faurote, G. and Kao, C.C. (1996) Initiation of (-)-Strand RNA Synthesis Catalyzed by the BMV RNA-Dependent RNA Polymerase: Synthesis of Oligonucleotides. *Virology*, **226**, 1-12.

Sun, X.L., Johnson, R.B., Hockman, M.A. and Wang, Q.M. (2000) De novo RNA Synthesis Catalyzed by Hcv RNA-Dependent RNA Polymerase. *Biochem Biophys Res Commun*, **268**, 798-803.

Sun, Y., Qiao, X., Qiao, J., Onodera, S. and Mindich, L. (2003) Unique Properties of the Inner Core of Bacteriophage Φ8, a Virus with a Segmented dsRNA Genome. *Virology*, **308**, 354-361.

Tao, Y., Farsetta, D.L., Nibert, M.L. and Harrison, S.C. (2002) RNA Synthesis in a Cage--Structural Studies of Reovirus Polymerase Lambda3. *Cell*, **111**, 733-745.

Terwilliger, T.C. and Berendzen, J. (1999) Automated Mad and Mir Structure Solution. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 4)**, 849-861.

Terwilliger, T.C. (2000) Maximum-Likelihood Density Modification. *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 8)**, 965-972.

Thompson, A.A. and Peersen, O.B. (2004) Structural Basis for Proteolysis-Dependent Activation of the Poliovirus RNA-Dependent RNA Polymerase. *EMBO J*, **23**, 3462-3471.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Clustal W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res*, **22**, 4673-4680.

van Dijk, A.A., Frilander, M. and Bamford, D.H. (1995) Differentitation between Minus- and Plus-Strand Synthesis: Polymerase Activity of dsRNA Bacteriophage Ø6 in an *in Vitro* Packaging and Replication System. *Virology*, **211**, 320-323.

van Dijk, A.A., Makeyev, E.V. and Bamford, D.H. (2004) Initiation of Viral RNA-Dependent RNA Polymerization. *J Gen Virol*, **85**, 1077-1093.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. and Yokoyama, S. (2002) Crystal Structure of a Bacterial RNA Polymerase Holoenzyme at 2.6Å Resolution. *Nature*, **417**, 712-719.

Wallace, A.C., Laskowski, R.A. and Thornton, J.M. (1995) Ligplot: A Program to Generate Schematic Diagrams of Protein-Ligand Interactions. *Protein Eng.*, **8**, 127-134.

Walter, T.S., Diprose, J., Brown, J., Pickford, M., Owens, R.J., Stuart, D.I. and Harlos, K. (2003) A Procedure for Setting up High-Throughput Nanolitre Crystallization Experiments. I. Protocol Design and Validation. *JouRNAl of Applied Crystallography*, **36**, 308-314.

Walter, T.S., *et al.* (2005) A Procedure for Setting up High-Throughput Nanolitre Crystallization Experiments. Crystallization Workflow for Initial Screening, Automated Storage, Imaging and Optimization. *Acta Crystallogr D Biol Crystallogr*, **61**, 651-657.

Weeks, C.M. and Miller, R. (1999) The Design and Implementation of Snb 2.0. *J. Appli. Cryst.*, **32**, 120-124.

Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004) Structural Basis of Transcription: Separation of RNA from DNA by RNA Polymerase II. *Science*, **303**, 1014-1016.

Winn, M.D., Isupov, M.N. and Murshudov, G.N. (2001) Use of Tls Parameters to Model Anisotropic Displacements in Macromolecular Refinement. *Acta Crystallogr D Biol Crystallogr*, **57**, 122-133.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants. *PLoS Biol*, **2**, E104.

Yang, H., Makeyev, E.V. and Bamford, D.H. (2001) Comparison of Polymerase Subunits from Double-Stranded RNA Bacteriophages. *J Virol*, **75**, 11088-11095.

Yang, H., Gottlieb, P., Wei, H., Bamford, D.H. and Makeyev, E.V. (2003a) Temperature Requirements for Initiation of RNA-Dependent RNA Polymerization. *Virology*, **314**, 706-715.

Yang, H., Makeyev, E.V., Butcher, S.J., Gaidelyte, A. and Bamford, D.H. (2003b) Two Distinct Mechanisms Ensure Transcriptional Polarity in Double-Stranded RNA Bacteriophages. *J Virol*, **77**, 1195-1203.

Yi, G.H., Zhang, C.Y., Cao, S., Wu, H.X. and Wang, Y. (2003) De novo RNA Synthesis by a Recombinant Classical Swine Fever Virus RNA-Dependent RNA Polymerase. *Eur J Biochem*, **270**, 4952-4961.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. and Darst, S.A. (1999) Crystal Structure of Thermus Aquaticus Core RNA Polymerase at 3.3Å Resolution. *Cell*, **98**, 811-824.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E. and Filipowicz, W. (2002) Human Dicer Preferentially Cleaves dsRNAs at Their Termini without a Requirement for ATP. *EMBO J*, **21**, 5875-5885.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004) Single Processing Center Models for Human Dicer and Bacterial RNAse III . *Cell*, **118**, 57-68.

Zhong, W., Gutshall, L.L. and Del Vecchio, A.M. (1998) Identification and Characterization of an RNA-Dependent RNA Polymerase Activity within the Nonstructural Protein 5B Region of Bovine Viral Diarrhea Virus. *J Virol*, **72**, 9365-9369.

Zhong, W., Uss, A.S., Ferrari, E., Lau, J.Y. and Hong, Z. (2000) De novo Initiation of RNA Synthesis by Hepatitis C Virus Nonstructural Protein 5B Polymerase. *J Virol*, **74**, 2017-2022.

Abbreviations

(-)ssRNA – Negative single-stranded RNA
(+)ssRNA – Positive single-stranded RNA
(d)NTP – Deoxy-nucleoside triphosphate
[$^{\alpha 32}$P] – Phosphorous isotope 32
°C – Celsius
A – Adenosine
Å – Ångstrom
a.u. – Asymmetric unit
A_{280} – Absorbance at 280nm
aa – Amino acid
Ago – Argonaute protein
aRNA – Aberrant RNA
BMV – Brmo mosaic virus
bp – Base pair
BTV – Blue-tongue Virus
BVDV – Bovine Viral Diarrhoea Virus
BVDVpol – Bovine Viral Diarrhoea Virus polymerase
C – Cytosine
C site – Catalytic site in Φ6pol
CCP4 – Collaborative Computing Project 4
Ci/mmol – Cintillations per millimol
cRdRP – Cell-encoded RNA-dependent RNA polymerase
C-terminal – Carboxyl terminal
CTP – Cytosine triphosphate
DCL-1 – Dicer-like protein 1 from *Neurospora crassa*
DCL-2 – Dicer-like protein 2 from *Neurospora crassa*
Dcr1 – Dicer protein from *Saccharomyces pombe*
ddCTP – Dideoxy-cytosine triphosphate
DdDP – DNA-dependent DNA polymerase
DdRP- DNA-dependent RNA polymerase
D*n* – Daughter strand nucleotide *n*, where *n* corresponds to the position of the template nucleotide that is base pairs with.
DNA – Deoxyribonucleic acid
DPBB – Double-psi β-barrel
dsRNA – Double-stranded RNA
DTT – Dithiothreitol
E491Q – Φ6pol protein with E491 residue mutated to Q
EDTA – Ethylene diamine tetraacetic acid
ESRF – European Synchrotron Radiation Facility, Grenoble, France
f' – Real component of the anomalous scattering factor
f'' – Complex component of the anomalous scattering factor
FMDV – Foot-and-Mouth Disease Virus
FMDVpol – Foot-and-Mouth Disease Virus polymerase
g – Unit of gravitational force
G – Guanosine

GTP – Guanidyl triphosphate
HCV – Hepatitis C Virus
HCVpol – Hepatitis C Virus polymerase
His-tag – Histidine tag
HIV-1 – Human Immunodeficiency virus type 1
HRV – Human Rhinovirus
HRVpol – Human Rhinovirus polymerase
I site – NTP interrogation site in Φ6pol
IPTG – Iso-propyl β-D-thio galactopyranoside
K – Kelvin
kDa – kiloDalton
KeV – Kilo electron Volts
KF – Klenow Fragment of DNA polymerase I
L – Large segment of genomic dsRNA of bacteriophage Φ6
l^+ – Positive sense strand of large segment of genomic dsRNA of bacteriophage Φ6
LB – Luria-Bertani medium
M – Medium segment of genomic dsRNA of bacteriophage Φ6
m^+ – Positive sense strand of medium segment of genomic dsRNA of bacteriophage Φ6
MAD – Multiple-wavelength Anomalous Dispersion
MDa – megaDalton
mg – milligrams
MIR – Molecular Isomorphous Replacement
miRNA – microRNAs
ml – millilitre
mM – milimolar
MPa - milipascal
mRNA – Messenger RNA
NC – Nucleocapsid
NCS – Non-crystallographic symmetry
nl – Nanoliter
nm – Nanometers
N*ru*I, N*si*I, N*de*I, P*st*I, H*ind*III, B*st*II, P*me*I – Restriction enzymes
NS2 – Non-structural protein from BTV
NSP2 – Non-structural protein from rotavirus
nt – Nucleotide
N-terminal – Amino terminal
NTP – Nucleoside triphosphate
NV – Norwalk Virus
NVpol – Norwalk Virus polymerase
OD_x – Optical density at x nm
OM – Outer membrane
OPPF – Oxford Protein Production Facility
P site – "Initiation stabilising platform" in Φ6pol
P1 – Major capsid protein from bacteriophage Φ6
P2 – Bacteriophage Φ6 polymerase
P4 – ATP-dependent helicase from bacteriophage Φ6
PAZ – PIWI/Argonaute/Zwile domain

PC – Polymerase Complex
PCR – Polymerase Chain Reaction
PEG – Polyethylene glycol
pEM55, pEM41, pYES2/CT, pEM69, INVSc1, pLM659 – Plasmids used in the production of QDE-1 ΔN
pEMG2, pRT2, pEM33, pNL18, pSVe4, pNL9 – Plasmids used in the production of Φ6pol
PG – Peptidoglycan
PM – Plasmatic membrane
Polβ – DNA polymerase β
PPi – Inorganic pyrophosphate
PPP-G-P-G – guanylyl(3'-5')-guanosine-5'-triphosphate
PTGS – Post-transcriptional gene silencing
PV – Poliovirus
PVpol – Poliovirus polymerase
qde-1 – Quelling defective gene 1 from *Neurospora crassa*
QDE-1 – Quelling defective protein 1 – RDRP from *Neurospora crassa*
QDE-1 ΔN – QDE-1 truncated protein (1-376 aa deleted)
qde-2 – Quelling defective gene 2 from *Neurospora crassa*
qde-3 – Quelling defective gene 3 from *Neurospora crassa*
Qβ – Coliphage Qβ
RBD – RNA binding domain
RdDP – RNA-dependent DNA polymerase
RDRC – RNA-dependent RNA polymerase complex
RdRP – RNA-dependent RNA polymerase
RecQ – ATP-dependent DNA helicase
RHDV – Rabbit Hemorrhagic Disease Virus
RHDVpol – Rabbit Hemorrhagic Disease Virus polymerase
RISC – RNA-induced silencing complex
RITS – RNA-induced transcriptional silencing
rmsd – Root mean square deviation
RNA – Ribonucleic acid
RNAi – RNA interference
RNApolII – RNA polymerase II
RNase – Ribonuclease
rNTP – Ribonucleoside triphosphate
rpm – Rotations per minute
RT – Reverse Transcriptase
S – Small segment of genomic dsRNA of bacteriophage Φ6
S pocket – Specificity site pocket in Φ6pol
s^+ – Positive sense strand of small segment of genomic dsRNA of bacteriophage Φ6
SAD – Single-wavelength Anomalous Dispersion
SB – Sodium boric acid electrophoresis buffer
SDS-PAGE – Sodium dodecylsulphate polyacrylamide gel electrophoresis
SeMet – Selenomethionine
SG – Φ6pol protein with 629QYWK632 residues mutated to SG
siRNA – small interfering RNAs
SRS – Synchrotron Radiation Source, Daresbury, UK

ssRNA – Single-stranded RNA
T – Thymidine
T7pol – Bacteriophage T7 polymerase
TGS – Transcriptional gene silencing
T*n* – Template strand nucleotide *n*, where *n* corresponds to the position of the nucleotide, starting from the 3' end
U – Uracil
UTP – Uracil triphosphate
vRdRP – viral RNA-dependent RNA polymerase
WT – Wild type
ε_{280} – Extinction coefficient at 280nm
λ3pol – Reovirus polymerase
µl – microlitre
σNS – Non-structural protein from orthoreovirus
Φ6 – Bacteriophage Φ6
Φ6pol – Bacteriophage Φ6 polymerase
Φ8 – Bacteriophage Φ8
Φ12 - Bacteriophage Φ12
Φ13 – Bacteriophage Φ13