DYNAMICS ON MULTIPLE TIMESCALES IN THE CYSTOVIRAL RNA-DIRECTED RNA POLYMERASE

by

ZHEN REN

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York
2012

APPROVAL PAGE

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. **Ranajeet Ghose**

Date Chair of Examining Committee

Dr. **Edward J. Kennelly**

Date Executive Officer

Dr**. Ranajeet Ghose**, City College of New York

Dr. **Zimei Bu**, City College of New York

Dr. **Ruth E. Stark,** City College of New York

Dr. **David Eliezer,** Weill Medical College/CUMC

Dr. **Stewart Shuman,** Sloan-Kettering Institute

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

DYNAMICS ON MULTIPLE TIMESCALES IN THE CYSTOVIRAL RNA-DIRECTED RNA POLYMERASE

by

Zhen Ren

Advisor: Professor Ranajeet Ghose

The RNA-dependent RNA polymerase P2 from cystovirus φ6 directs the replication and transcription of the double-stranded RNA genomes. In spite of the availability of multiple crystal structures at various points along its catalytic pathway, the dynamics role involving in the catalytic cycle and fidelity control remain unclear. Isoleucine residues are distributed over the three-dimensional fold of P2. By using the δ1 positions of 25 Ile residues as probes, we measured the dynamic modes and their alterations along P2 catalytic cycle with CPMG-based multiple quantum relaxation dispersion experiments. The results indicate that P2 is dynamic on the fast (ps-ns) and slow (μs-ms) timescale. The characteristics of several motional modes are altered in the presence of substrate analogs and single-stranded RNA templates. The timescale of the lower frequency motional modes that involve several conserved functional motifs coincides with the catalytic timescale (1-2 ms), which was determined from kinetic analyses of representative RdRPs. We further investigated the influence of the extreme 3'-end sequence of the single-stranded RNA templates and the nature of the substrate nucleotide triphosphates on the slow motional modes using multiple-quantum relaxation dispersion. We found that P2, in the presence of templates bearing the proper genomic 3'-ends or the preferred initiation nucleotide

(GTP), displays unique dynamic signatures that are different from those in the presence of non-physiological templates or substrates. This further suggests that dynamics may play a role in the fidelity of recognition of the correct substrates and template sequences to initiate RNA polymerization.

ACKNOWLEDGMENT

I would like to thank all people who have helped, inspired and support me during my doctoral study. This thesis wouldn't be possible to complete without them. I owe my deepest gratitude to my advisor, Professor Ranajeet Ghose. His enthusiasm and passion for research have inspired everyone around him. He is the person always reachable and supportive especially when you need. Research is fun in Ghose's lab. I am also very thankful to my committee members Dr. Zimei Bu, Dr. David Eliezer, Dr. Stewart Shuman and Dr. Ruth E. Stark for their thorough reviews, illuminating ideas and comments from proposal development to thesis completion.

I would like to show special gratitude to Dr. Hsin Wang. He patiently walked me through all the steps of NMR experiments. Without his help, my research wouldn't go smoothly. I am grateful to Dr. Matthew Franklin in NYSBC. Thanks for his help in solving φ12 P2 structure and leading me to the fascinating crystallography world.

It is a pleasure to thank for all the members of Ghose's lab who I have been worked with--Dr. Andrea Piserchio, who is always coming up with brilliant ideas to inspire people. Thanks him for proof reading my papers and thesis. Dr. Sebastien Alphonse, who is one of the most diligent scientists I have ever known. Thanks him for being always reachable in the lab and providing valuable ideas for my thesis. Dr. Deniz Temel, Dr. Ertan Eryilmaz and Dr. Aswin Natarajan, who are the former members of Prof. Ghose's lab. They became my dearest friends and make my Ph.D life colorful. Thanks Mr. Benjamin Reed, Dr. Padmanava Pradhan and other members in the lab, it is an enjoyable experience working with them.

Finally, but most importantly, I would like to thank for my family. Thanks for their unconditional love and support. They give me the courage to take challenges in life. Thanks my dearest friends in New York and China. They make my life more meaningful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Bis-Tris	2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol
BMV	Broom mosaic virus
BTV	Bluetongue virus
BVDV	Bovine diarrhea virus
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
DdDP	DNA-dependent DNA polymerase
DdRP	DNA-dependent RNA polymerase
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
HIV	Human immunodeficiency virus
MQ	Multiple-quantum
NC	Nucleocapsid
PC	Procapsid
PX	Polymerase complex
PV	Poliovirus
RDV	Rice dwarf virus
RdDP	RNA-dependent DNA polymerase
RdRP	RNA-dependent RNA polymerase
RF	Radiofrequency

RNA	Ribonucleic acid
RT	Reverse transcriptase
SDS	Sodium dodecyl sulfate
SQ	Single-quantum
ssRNA	single-stranded RNA
Tris	Tris-hydroxymethyl aminomethane
VC	Viral core
5-nt	5 nucleotides

Chapter 1

INTRODUCTION

Double-stranded RNA (dsRNA) viruses represent a diverse group of viruses that infect bacteria, fungi, plants, animals and humans. Some examples of dsRNA viruses include rotaviruses, bluetongue virus (BTV), rice dwarf virus (RDV) and cystoviruses (1). Rotaviruses are the most common cause of severe diarrhea among infants and young children, especially in developing countries. Rotavirus infection causes an estimated 450,000 to 700,000 deaths each year in young children below five globally (2). BTV causes severe illness in livestock (3), while RDV infects rice plants, vertebrates and even mammalian hosts causing significant economic losses (4). Due to the medical importance and economic significance, dsRNA viruses have been extensively studied. Among the best structurally characterized dsRNA viruses, are the class of bacteriophages known as the cystoviruses (5). Cystoviruses, most notably φ6 and φ12, have been proven to be of great utility as model systems in obtaining structural and mechanistic insights into dsRNA virus assembly, genome packaging, transcription and replication. Enzymes called RNA polymerases that are encoded in the viral genome control the latter two processes. In this study we used a variety of techniques especially solution-state NMR spectroscopy to obtain structural, functional and dynamic insights into the catalytic activity of the RNA polymerase from cystovirus. These studies provide mechanistic insight not only into cystoviral RNA polymerases, but also into other RNA polymerases encoded by a larger family of RNA viruses including those discussed above, the highly pathogenic poliovirus (PV) (6) and hepatitis C virus (HCV) (7).

1.1. Cystoviruses – Overall Structure and Organization

Until now, nine members of cystovirus family, all affecting various species of *Pseudomonas* have been isolated (bacteriophage φ6-14). Among these nine cystoviruses, bacteriophage φ6, 7, 9, 10 and 11 are genetically similar, whereas φ8, 12, and 13 contain a more distantly related genome (8,9). The cystoviral genome contains three dsRNA segments that are classified based on size: Small (S segment) – 2.9 kb, Medium (M segment) – 4.0 kb, and Large (L segment) – 6.4 kb.

The outermost layer of the cystoviral virion (**Figure 1.1**) is a bilayer that consists of lipids, which are acquired from the host during the lysis phase of the life-cycle (see **Section 1.2** below). The outer coat contains 4 integral membrane proteins: P6 that anchors the spike protein, P3, to the membrane; P3 binds the host cell-surface receptor; P9, a major envelope protein; P13, a non-essential protein; and P10, likely a holin (5). All of these proteins are encoded by the genomic M segment except P9, which is encoded by the S segment. Within this lipid bilayer is the nucleocapsid (NC) containing two sets of protein coats. The outer coat consists principally of the coat protein P8, arranged in an icosahedral T=13 arrangement (10,11) and the lytic enzyme P5, both of which are encoded by the S segment. Inside the first protein layer is the viral core (VC) containing the 4-protein (P1, P2, P4 and P7 encoded by the genomic L segment) polymerase complex (PX, see **Section 1.3**) and the viral genome.

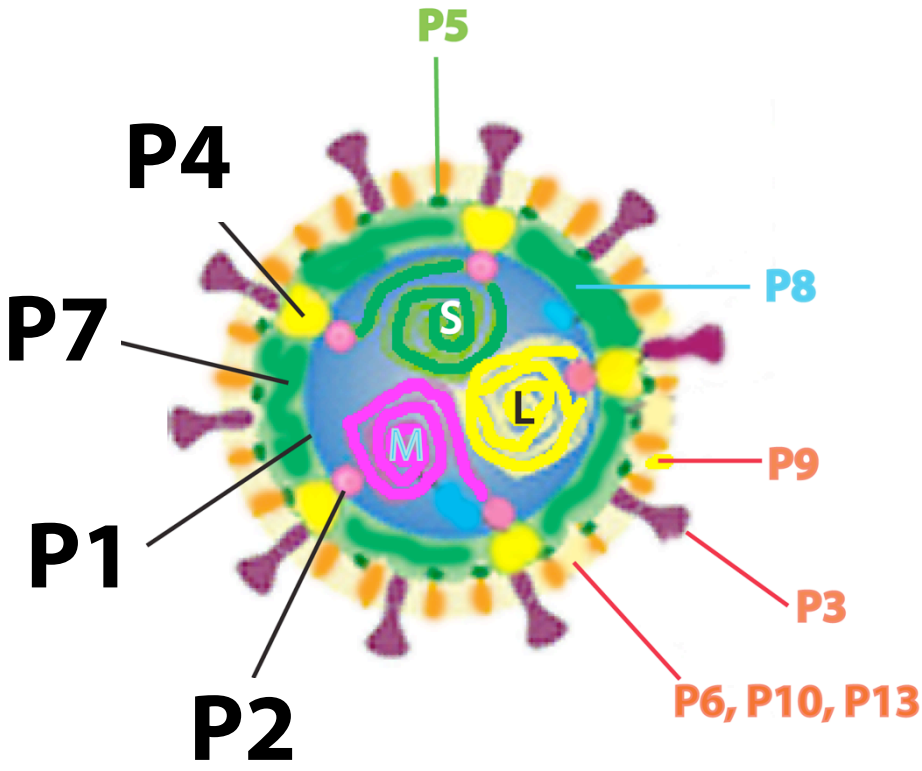


Figure 1.1. Overall Organization of Cystoviruses.
The cystoviruses contain an outer lipid coat and a double protein capsid. Embedded in the lipid core is the invasion and fusion apparatus (orange) consisting of several integral membrane proteins. The outer protein core is composed of the capsid protein P8 (cyan). The viral core (VC) comprises the 4-protein polymerase complex (shown in large black lettering). The three-segmented double-stranded RNA genome is contained within the VC. Figure modified from **Poranen et. al**. (5)

1.2. The Cystoviral Life cycle

During infection, P3 recognizes the host cell-surface receptor followed by membrane fusion mediated by protein P6, leading to the release of the NC into the host periplasm. Next, the protein P5 lyses the peptidoglycan layer, and releases the VC into the host cytoplasm. The VC particle then utilizes its constituent RNA-directed RNA polymerase (P2, see **Section 1.3**) to catalyze semi-conservative transcription by utilizing (-) sense single-stranded RNA (ssRNA) segments as template to synthesize (+) sense ssRNA segments. The nascent (+) sense ssRNA segments are extruded into cytoplasm, directing the translation of protein P1, P2, P4, and P7,

which self-assemble to form the dodecahedral procapsid (PC). The newly formed PC is capable of recognizing and packaging one copy of each (+) sense ssRNA segments into the protein shell. After packaging, the RNA polymerase, P2, utilizes (+) sense ssRNA segments as a template to replicate i.e. direct (-) sense RNA synthesis generating the dsRNA genome. Interestingly, only after the complete synthesis of (-) sense ssRNA segments, does the transcription step take place (12).

Elegant *in vitro* studies have revealed the fascinating mechanism of RNA packaging in cystoviruses (13). About 200 nucleotides from the 5'-end of each transcript segment are sequences called *pac* sequences that are necessary for packaging (14). The *pac* sequence is different for each of the three genomic segments. However, most of these *pac* sequences contain a number of stem-loop secondary structures. In cystoviruses, close relatives share similar stem-loop structure except for differences in a few base (14). The currently accepted model of the genome packaging process is as follows (15): The (+) sense ssRNA transcripts are packaged into the PC in a sequential manner, first s^+, then m^+, and last l^+ ssRNA (14). Studies in which specific RNA segments were crosslinked to the PC (15) demonstrated that the 5'-end of the genomic segments specifically bind onto P1 in the presence of ATP. The location of this binding site is still unknown and whether all the genomic segments share the same binding site is also not known. After binding, the ssRNA segments are drawn into the PC by P4 (see **Section 1.3.2**), a hexameric ATPase, through the hydrolysis of ATP (16). Notably, the packaging process is volume dependent (17). For instance, if the transcripts are smaller than genomic RNA segments, PX will pack multiple RNA segments till reach the normal genomic content. Furthermore, if all three genomic segments, s^+, m^+, and l^+, are combined into a single ssRNA containing the 5' *pac* sequence from s^+, the entire RNA chain will be successfully packaged into the PC (18).

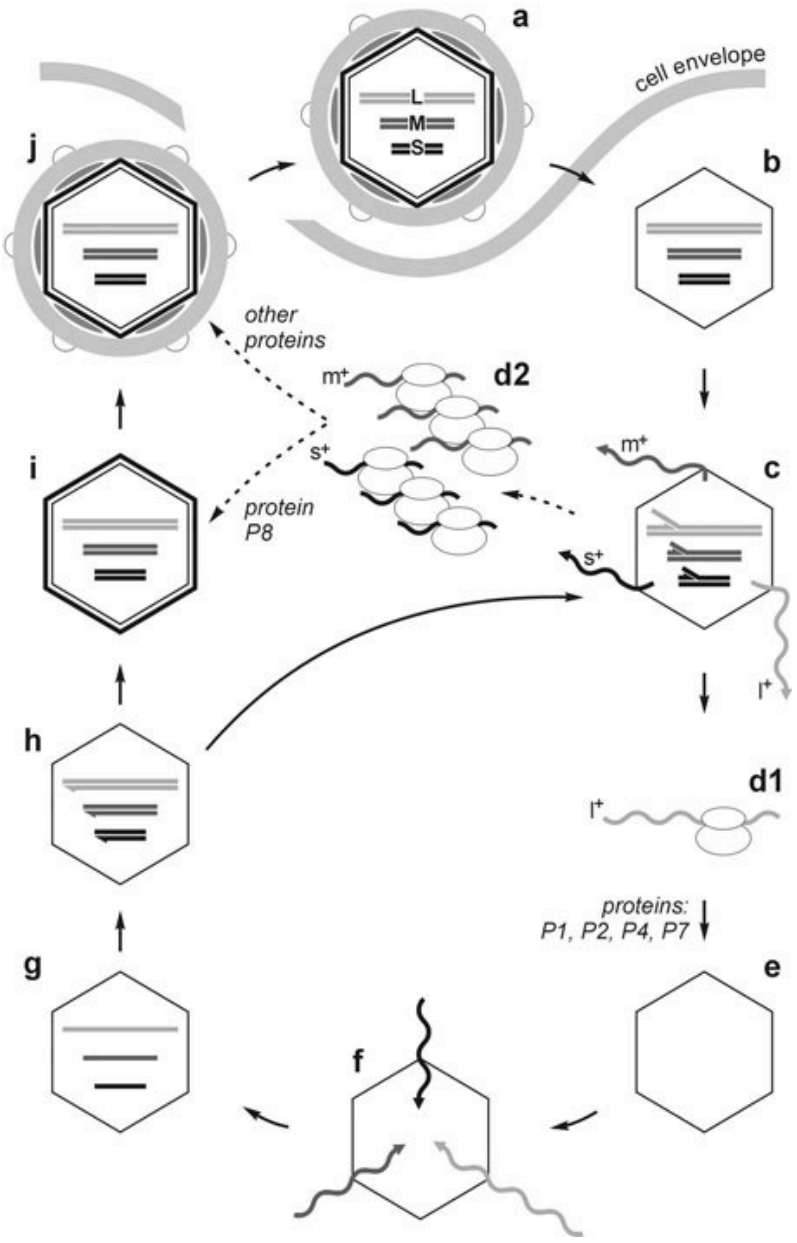


Figure 1.2. Schematic Representation of the Cystoviral Life-Cycle.
(a) The viral dsRNA genomic segments are brought into the host cell inside VC particle. (b) The VC particle catalyzes semi-conservative transcription. (c) (+) ssRNA transcripts l, m and s segments are extruded into the cytoplasm. The cellular protein synthesis apparatus translates l^+ RNA (d1), giving rise to proteins P1, P2, P4 and P7. (e) The newly produced proteins form empty PCs. (f) The PC particles package one copy of each of the (+) ssRNA segments (l^+, m^+ and s^+) per particle. (g) The PX then catalyzes (-) strand synthesis to reconstitute genomic dsRNA segments. (h) The particle at this stage can enter into an additional round of transcription (to step c) or alternatively mature into infectious virions. (i) The latter pathway uses proteins produced by the translation of m^+ and s^+ ssRNA segments (d2) and includes addition of the protein P8 shell to form the nuclocapsid. (j) This is followed by acquisition of the rest of the viral structural proteins together with the lipid membrane. The mature virus particles are released by lysis of the host cell. Figure from **Makeyev and Bamford; 2000** (16).

Once all three ssRNAs are packaged and the genomic dsRNA segments are formed, the mature particles are ready to enter in an additional round of transcription or alternatively transform into infectious NC by acquiring an outer P8 shell protein. These self-assembled NC then acquire a lipid envelop from the host cell plasma membrane and are released as infectious particles through cell lysis (19). (**Figure 1.2**)

1.3. Proteins of the Cystoviral Polymerase Complex (PX)

The cystoviral PX is composed of four proteins – P1, P2, P4 and P7, all encoded by the genomic L segment. 120 copies of P1 form the overall framework of the PX in a dodecahedral particle (20). Besides its structural role, P1 also recognizes the specific *pac* sequences near the 5'-ends of the (+) genomic strands to initiate sequential packaging. P4 is a hexameric ATPase that is located on the 5-fold axes (causing a symmetry mismatch) **(Figure 1.3)**, provides energy for genome packaging through ATP hydrolysis (21). It also hypothesized to act as a passive portal for the translocation of the newly synthesized (+) ssRNA segments during semi-conservative transcription (21). P7 appears to stabilize RNA packaging and could facilitate procapsid expansion through motional coupling with P1 segments (10,22). The location of P7 within the PX is unknown. P2 is an RNA dependent RNA polymerase (RdRP), which directs RNA synthesis and is essential to the viral life cycle. Recently, it was reported using cryo-electron microscopy that P2 is located on or near the 5-fold axes of the PX near the RNA entry and exit channels (23). However, the exact location of P2 remains unclear. We describe the various proteins of the PX in greater detail below.

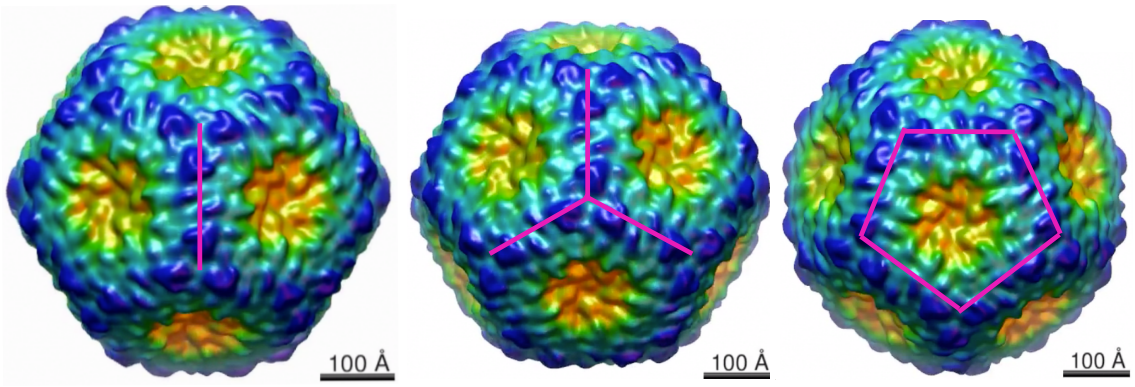
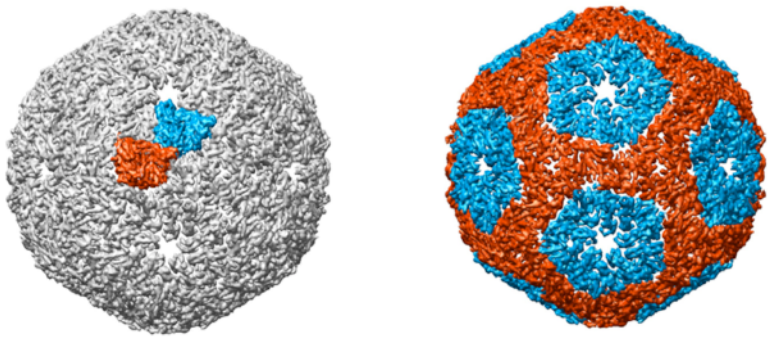


Figure 1.3. Symmetry Axes in a Dodecahedral Bacteriophage φ6 PX.
A dodecahedron contains 15 2-fold axes (left), 10 3-fold axes (center) and 6 5-fold axes (24). Images modified from **EM DATA BANK (EMDB) / 5355**

1. 3.1. Major Capsid Protein (P1)

The 85 kDa φ6 P1 protein forms the overall skeleton of the PX with 60 asymmetric P1 dimers arranged on a T=1 lattice, interacting with all other constituent proteins. Although P1 is the central component of the PX, it is the only one for which an atomic resolution structure is as of yet unavailable. The overall poor solubility of P1 constructs and the tendency to form heterogeneous aggregates has prevented detailed structural studies. The most detailed structure of P1 is available from high-resolution cryo-EM studies on φ6 (10). These studies suggest that P1 is a mixed α/β protein, and P1 molecules are located at the five-fold and at the three-fold vertices (11). It was proposed in addition to forming the framework of the PX, that the major role of P1 is to recognize the 5' *pac* sequence of the RNA segments. It has been speculated that the *pac* sequences bind a P1 dimer resulting in both local conformational changes for each P1 monomer as well as global conformation changes at the dimer interface, which are allowing the PX shell to expand and accommodate the newly packaged RNA segments (10). Additionally, the interaction of the flexible C-terminal tail of P7 (see **Section 1.3.3**) has been proposed to facilitate

this expansion (25).



A B

Figure 1.4. φ6 P1, Cryo EM Image Reconstruction.
(A) Two copies of P1 arrange as an asymmetric P1 dimer on a T=1 lattice. Two monomers are color coded as blue and orange. (B) 60 dimers comprise the Dodecahedral protein capsid. Figure is modified from **Butcher, S.J., Structure 2006** (10)**.**

1.3.2. Hexameric ATPase (P4)

The hexameric ATPase P4 (35 kDa monomer) is located at the 5-fold axes of the PX and functions as a packaging motor translocating (+) ssRNAs into the PX facilitated by conformational changes in P1 shell (described above). The structure of the φ12 P4 monomer illustrates its similarity with equivalent units of hexameric helicases such as RecA (26). A series of structures in addition to the apo state of φ12 P4, including a substrate ATP-bound state and an RNA-bound state, provide a vivid illustration of the conformational changes that occur along the P4 catalytic cycle (26). These structures indicate that ATP hydrolysis is the driving force for RNA translocation, and a conformational change as small as 1 Å at the ATP binding site can be amplified into a nanometer scale motion along the nucleic acid chain (27). In spite of extensive structural studies, biochemical analyses and recent hydrogen/deuterium exchange by mass spectrometry (HX-MS) studies (28), some crucial mechanism still need to be elucidated. The most central of these questions include insight into how hexameric P4 and P1 work in unison

to couple ATP hydrolysis and RNA translocation with capsid expansion and genome packaging.

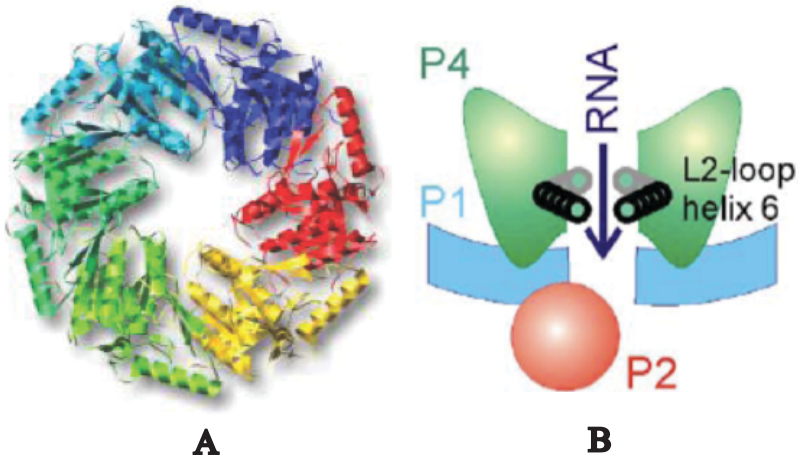


Figure 1.5. Hexameric ATPase, P4.
(A) Hexameric ATPase, P4. Each monomer was color-coded with yellow, red, purple, cyan, green and lemon green (PDB code: 1W48). (B) Schematic packaging illustration of hexameric P4. By consuming the energy form the ATP hydrolysis, the RNA strand was dragged downwards through L2-Loop and helix 6 by up and down sequential coordinated motions. After the genomic RNA strands packaging complete, P2 becomes active and starts semi-conservative transcription. Figure modified from **Kainov, D.E. et al., *NAR* 2004** (29).

1.3.3. Essential PX Protein (P7)

The 17 kDa P7 protein is a minor component of the PX and exists as a symmetric homodimer in solution. The crystal structure of φ12 P7 shows that it forms a novel α/β-fold with a disordered C-terminal tail (30). Structural and dynamic studies indicated that although highly disordered, the C-terminal tail of P7 possesses residual helical properties. P7 is capable of weakly interacting with RNA and this interaction leads to a significant reduction of the degree of disorder (22). While its functional role remains unclear, P7 was shown to greatly enhanced the assembly rate of PX particles (19). In addition, incomplete PX particles composed of the P1, P2, and P4 proteins (P124 particles) displayed minimal packaging of the (+) sense genomic strands in φ6. These results suggested that P7 likely functions both as an assembly cofactor and a packaging co-factor acting to stabilize the PX, most likely through interactions with P1 (25).

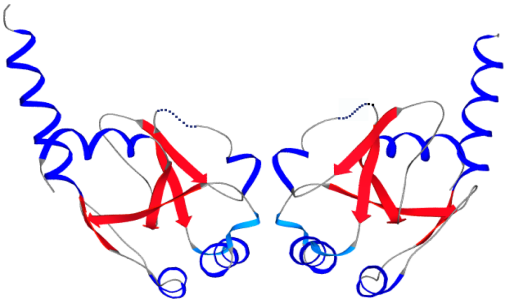


Figure 1.6. Structure of the P7 Core (P7ΔC)
P7 forms as a symmetric dimer in solution. P7ΔC crystal structure was solved in $P3_221$ space group forming a novel BRCT-like fold. BRCT domains found in DNA-repair proteins. (22)

1.3.4. RNA dependent RNA Polymerase (P2)

Viral polymerases play a crucial role in genome transcription and replication. Viral nucleic acid polymerases can be broadly classified into four distinct groups (31): (i) RNA-dependent RNA polymerase (RdRP) which transcribes RNA using a RNA template. These polymerases can perform RNA synthesis with the use of a primer (e.g Poliovirus RdRP, 3D; foot-and-mouth disease virus polymerase etc.) or *de novo* without using a primer (e.g. Hepatitis C virus polymerase, NS5B; cystovirus polymerase, P2) (32). (ii) RNA-dependent DNA polymerase (RdDP) that transcribes DNA using an RNA template, also known as a Reverse Transcriptase (RT) e.g. human immunodeficiency virus (HIV) RT. (iii) DNA-dependent RNA polymerase (DdRP) that utilizes DNA templates for RNA synthesis, e.g. in *vaccinia* or *variola viruses* (33) . (iv) DNA-dependent DNA polymerase (DdDP) that transcribes DNA from a DNA template, eg in adenovirus, herpes virus and small pox virus (33).

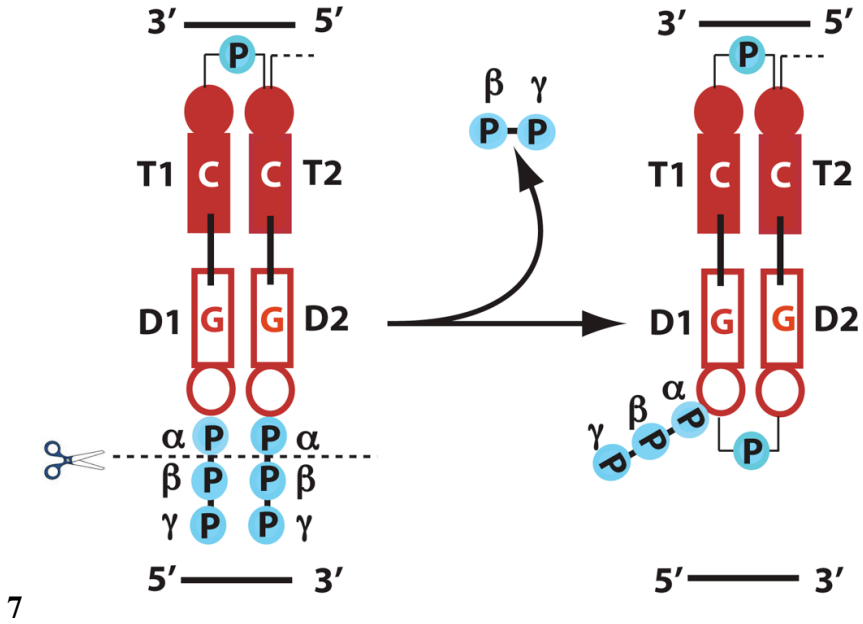
P2, the ~75 kDa cystoviral RdRP is a key component for the PX and essential for the life cycle of the virus. P2 acts as a replicase to synthesize the complementary (-) strand RNA from

(+) strand RNA templates, resulting in a dsRNA genome. In addition, P2 acts as a transcriptase by using the (-) strand of the dsRNA as template for (+) strand synthesis (16). Both reactions occur *de novo* without the use of a primer and from the 3' terminal nucleotide of the templates, replicating and transcribing the genome efficiently without loss minimal of information (34).

1.3.4.1. *De Novo* Initiation of RNA Synthesis by P2

Initiation of viral RNA synthesis is an important regulatory step in the replication and transcription of RNA viruses. Viral RNA initiation needs to take place at or near the 3' end of the viral template. There are two principal mechanisms - *de novo* initiation (primer-independent initiation) and primer-dependent initiation (35,36).

Cystoviruses utilize the *de novo* initiation mechanism requiring interactions of P2 with an RNA template, virus-specific initiation nucleotides and the initiation nucleoside triphosphate (37). The initiation nucleotide (D1), that would become the first base from the 5'-end of the daughter strand, enters the active site of P2 and base-pairs with a template initiation site (2), providing the 3'-hydroxyl for nucleotidyl transfer to a second substrate nucleoside triphosphate (D2), that base-pairs with the penultimate template nucleotide from the 3'-end (T2) (see **Figure 1.7**). The first phosphodiester bond is thus formed between D1 and D2 with the cleavage of the latter triphosphate moeity at the α-position and release of pyrophosphate. Since there is no need to generate primer during replication, *de novo* initiation is efficient and immediately followed by elongation. The elongation process continues with each incoming substrate nucleotide (CTP, UTP, ATP or GTP) complementary to the corresponding template nucleobase (**Figure 1.7**) (38).



7

Figure 1.7. Schematic Representation of *de novo* Initiation.
Two NTP bases (D1, D2) base pair with the two 3'-terminal nucleotides (T1, T2) of the template strand to form the initiation complex. Complex formation is followed by a nucleophilic attack by the 3'-OH of D1 on the α-phosphate of D2, forming the first phosphodiester bond of the daughter strand and releasing pyrophosphate.

1.3.4.2. Structural Studies on *de novo* Initiation by P2

The first insights into the molecular details of the initiation mechanism were provided by structural studies of φ6 P2 in complex with a 5-nt DNA template and NTP substrates (39,40). These studies provide a model in atomic detail for the primer-independent *de novo* initiation of RNA polymerization. Similar to the structures of other nuclei acid polymerases (41), φ6 P2 contains the classic polymerase fold structure with fingers, palm and thumb domains (**Figure 1.8**). In addition, P2 also contains a number of unique features to prove efficient and accurate initiation of RNA-dependent RNA synthesis. Unlike DNA polymerases that form an "open-hand" shape, allowing the fingers and thumb domains significant relative conformational flexibility, φ6 P2, reminiscent of HCV polymerase (38), contain loops known as fingertips that connect the fingers and thumb domain and create a "closed-hand" shape with an encircled

catalytic cavity (39,42). This closed structure creates a well-defined template channel and regulates the recognition of initiation site (39). The central cavity of P2 is not large enough to accommodate dsRNA and P2 is capable of unwinding dsRNA without the help of a helicase during transcription (16). The mechanism of this process of unraveling duplex RNA is still a mystery. P2 is also an unspecific RdRP and can use DNA templates to start initiation (39). P2 also contains the sequence motifs (A-E in its palm domain and F in the fingers domain) that are conserved in RdRPs (43). These motifs are essential for metal-ion co-ordination, NTP binding and catalysis. Notably, motif C contains the classic GDD sequence that is central for catalysis (also see **Figure 1.8** below).

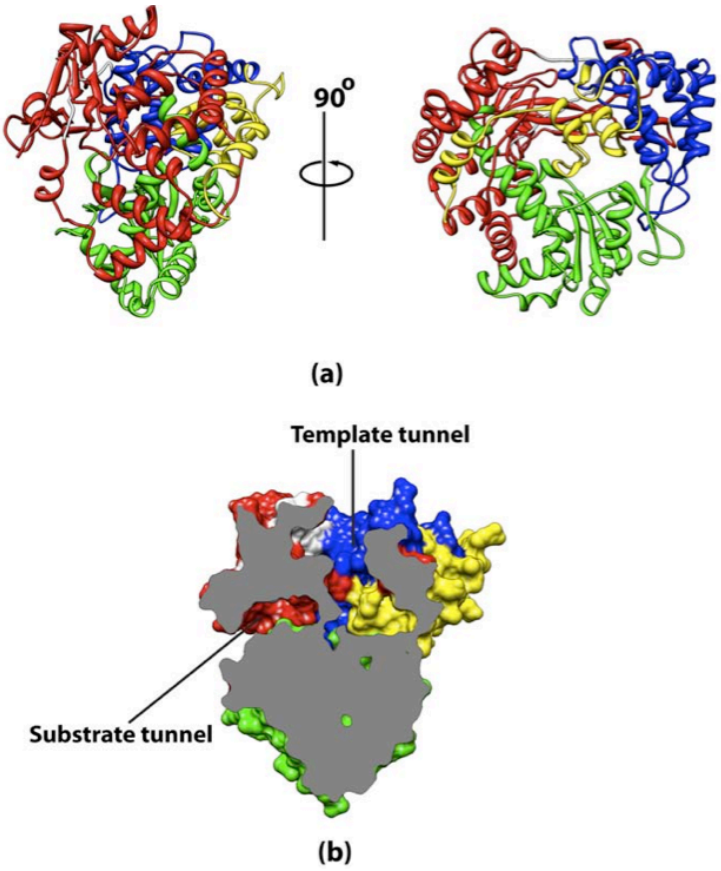


Figure 1.8. Structure of φ6 P2.
(a) Two views of a ribbon representation of the RNA-directed RNA polymerase, P2, from the cystovirus φ6. The fingers (1-30, 104-276, 333-397), thumb (37-91, 518-600), palm (277-332, 398-517) and C-terminal domains (601-664) are colored in red, blue, green and yellow respectively. The chains connecting the fingers and thumb domains (fingertips) are shown in light grey. (b) Side view of the φ6 P2

molecular surface. The surface has been cut to reveal the substrate and template tunnels. The specificity pocket lies at the region of intersection of the two tunnels. The orientation is the same as that on the left panel of (a).

1.3.4.3. Role of Divalent Metals in the Catalytic Activity of P2

The Mg^{2+} ions play a crucial role in phosphodiester bond formation during the catalytic process of P2. It has been shown that a pair of Mg^{2+} are involved in the daughter strand synthesis of replication complexes (40). One Mg^{2+} promotes the deprotonation of the 3'-hydroxyl of the nascent strand. At the same time the other Mg^{2+} facilitates the formation of the 5-coordinate transition state at the α-phosphate of the NTP. Mg^{2+} also stabilizes the transient by-product, inorganic pyrophosphate, before it is released via the substrate tunnel in P2 (see **Figure 1.9**) (40).

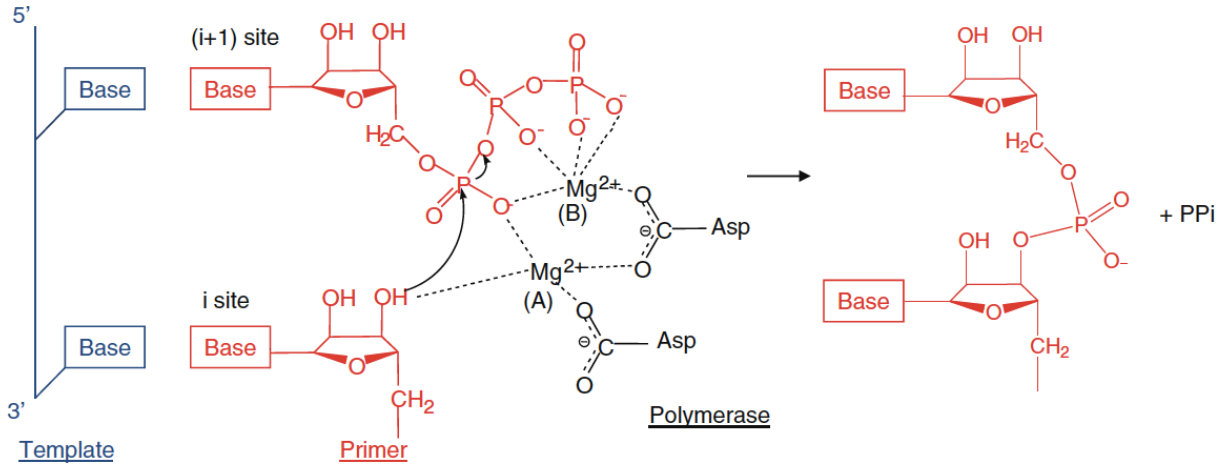


Figure 1.9. Catalytic Role of Mg^{2+} Ions.
A pair of Mg^{2+} is involved in the daughter strand synthesis of replication complexes. One Mg^{2+} (A) promotes the deprotonation of the 3' hydroxyl of the nascent strand. At the same time the other Mg^{2+} (B) facilitates the formation of the pentacovalent transition state at the α-phosphate of the NTP. Mg^{2+} (B) also stabilizes the transient by-product, PPi, before it is released (40). Figure taken from Choi, K.H; 2012(31).

Manganese ions (Mn^{2+}) are known to stimulate a number of RdRPs including HCV, Broome Mosaic Virus (BMV), poliovirus and cystoviruses (24). It was reported that Mn^{2+} stimulated the replication and transcription efficiency of φ6 P2 (44). A high-resolution structure

proved that Mn^{2+} binds in either the catalytic pocket (45) and/or at a specific allosteric position (39). Mutation of the Mn^{2+} binding site E491 to Q and the resultant decrease in the Mn^{2+} binding capability lead to a decreased RNA affinity and compromised elongation rate in φ6 P2 (46). However, under physiological conditions, the intracellular concentration of Mn^{2+} is low, its stimulation of the catalytic activity may not be biologically significant.

Unlike Mg^{2+} and Mn^{2+}, Ca^{2+} inhibits the catalytic activity of φ6 P2 (40). The origin of this metal ion inhibition was elucidated from crystallographic studies that suggested that Ca^{2+} ions inhibit the initiation and transcription of RNA synthesis of φ6 P2 by altering the geometry interactions in the catalytic pocket. In φ6 P2, two Ca^{2+} ions can substitute the Mg^{2+} ions in the initiation complexes. One of the Ca^{2+} ions occupies a position equivalent to the corresponding Mg^{2+} in the initiation complex (40). The other Ca^{2+} ion disrupts a stacking platform that involves the rotation of an aromatic residue (Tyr-630) that belongs to the C-terminal priming domain (39), leading to an unfavorable geometry with sub-optimal base stacking interactions.

1.4 Aims

In the absence of proofreading exonuclease activity, RdRPs have often been characterized as error prone polymerase (47). The fidelity of RdRPs is widely investigated due to its importance in virus genome variation (48) and the development of RdRPs targeted antiviral drugs (49). The structural basis of fidelity control is still incompletely understood, but current studies point towards conformational changes during nucleotide-addition cycle (47,50). One of the fidelity checkpoints for RdRPs involves a conformational change likely involving the reorientation of triphosphate moiety into a position suitable for efficient catalysis (47,50). There is evidence that changing a single amino acid residue (G64S), at a remote site distant from

catalytic site, increased fidelity control in poliovirus polymerase (3D) (50). Molecular dynamics simulations in G64S mutant also reconciled the result by revealed an altered dynamics observed globally in nanosecond timescale (51). Therefore, dynamics may play a role in fidelity control, bridging the gap between 'structure and function' (52).

Here we select φ6 P2 as a candidate from RdRPs, utilizing multiple-quantum spectroscopy, to investigate the role of dynamics along the catalytic cycle. The first step for us is to probe 25 Ile residues distributed over the conserved structural domains and sequence motifs, measure and analyze the dynamic changes along the catalytic cycle. Furthermore, in the case of φ6 P2, the efficiency of RNA polymerization *in vivo* depends on the sequence at the extreme 3'-ends of the RNA template especially with respect to the two penultimate bases (16). This difference in efficacy could be simply due to the affinity with which P2 recognizes the sequence at the 3'-end of the template. Or it could be structural or conformational changes resulting from the fidelity of recognizing the correct substrates and template sequences to initiate RNA polymerization. In this work, we further investigate the influence of the extreme 3'-end sequence of the single-stranded RNA templates and the nature of the substrate nucleotide triphosphates on the motional modes by using the measurement and analysis of multiple-quantum spin relaxation.

Chapter 2

DYNAMICS ON MULTIPLE TIMESCALES IN THE RNA-DIRECTED RNA POLYMERASE FROM THE CYSTOVIRUS φ6 P2

Most of the contents of this chapter have been published as:

Ren Z, Wang H and Ghose R (**2010)**. *Nucleic Acids Res.* **15**, 5105-5118.

2.1. Motion and Catalysis in RdRPs

As discussed previously, multiple high-resolution crystal structures exist for φ6 P2 in the apo-form, and in complex with NTPs (and NTP-analogs), single-stranded DNA (ssDNA) and single-stranded RNA (ssRNA) templates (24,39,40). While these structures of φ6 P2 and indeed of several other RdRPs provide atomic snapshots along the catalytic pathway, extensive flexibility both at the domain level and at the level of local motifs seems essential to reconcile biochemical data with structural information (52). For example, in φ6 P2 and in the homologous hepatitis C virus polymerase (HCVpol), the central catalytic channel is enclosed by additional loops ("fingertips") that connect the thumb and fingers (as discussed earlier in **Section 1.3.4.3.)** limiting the possibility of large-scale opening and closing motions as in DNA polymerases (53). Thus, in the absence of domain motion, the narrow template channel (**Figure 1.4b**) is capable of accommodating ssRNA but not RNA duplexes. However, these motions, though not large in amplitude, likely exist, since both enzymes are capable of processing circular templates (54). Reorientation of the fingers and thumb would then be necessary in order to accommodate the wider template. Indeed structures of HCVpol have been solved in both the "open" and "closed"

forms (55).

On the other hand, evidence of motion on a more local scale has also been found in RdRPs. Studies on the poliovirus polymerase (3Dpol) and also on other classes of nucleic acid polymerases (HIV-1 reverse transcriptase, RB69 DNA-dependent DNA polymerase and T7 DNA-dependent RNA polymerase) have suggested that a step that leads to significant rate enhancement during the nucleotide addition cycle involves protonation of the pyrophosphate product facilitated by the conformational changes in motif D (56). Other studies have hinted towards the influence of motions involving motif A on catalysis (50). An allosteric surface site that binds NTPs has also been located in HCVpol (7). This site, that is a target for a host of nucleotide-mimetic antivirals, is dynamically coupled to the catalytic site that is more than 30 Å away (57). Thus, dynamics both local and long-range, seem to be "the missing link between structure and function" (52) in these ubiquitous enzymes.

Measurement, quantitative analyses and interpretation of nuclear spin relaxation data in solution provides the most promising route towards characterization of dynamics on multiple timescales (58) to complement the large amount of structural and biochemical data available on RdRPs. These studies have proven to be quite successful in revealing functional dynamics in other proteins. For example, in the case of adenylate kinase, the lid opening motions were noted to be rate limiting in the catalytic cycle of the enzyme (59). Similar conclusions were drawn for cyclophilin A where the measured exchange rate (k_{ex}) obtained from relaxation dispersion experiments, was found to be close to the sum of the forward and backward rates of the cis-trans isomerization (60).

While the RdRPs are large (larger than 55 kDa) and generally difficult to solubilize in high enough quantities under conditions suitable for NMR studies, recent advances in

experimental methodology for both the backbone (61) and sidechains (62) coupled with the availability of cryogenic probe technology (63), provide great promise for the accurate measurement and interpretation of spin-relaxation data for these enzymes in solution.

Methyl groups are often in the hydrophobic core, and they are good probes for protein structure and dynamics studies (64). Because of good sensitivities and long relaxation times, methyl group NMR studies are particularly beneficial for high molecular weight system as φ6 P2. P2 has 25 Ile that are overall well distributed in 3 dimensional fold structure. Adding 2-keto-3-d_2-4-^{13}C-butyrate as precursor apply to a perdeuterated minimal medium composed of ^{12}C labeled glucose results perdeuterated Ile-Cδ1-^{13}C,^{1}H -labeled P2 without scrambling (see **Figure 2.1**) (65).

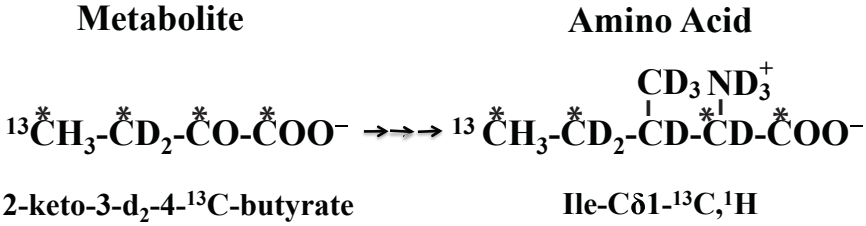


Figure 2.1. Chemical Structure of Metabolite 2-keto-3-d_2-4-^{13}C-butyrate.
2-keto-3-d_2-4-^{13}C-butyrate supplementing D_2O based minimal medium composed of ^{12}C labeled glucose and the resulting isoleucine (Ile) product. '*' represents the isoleucine carbons derived from 2-keto-3-d_2-4-^{13}C-butyrate.

In our attempts to elucidate the role of dynamics along the catalytic cycle of RdRPs, in atomic detail, we have commenced NMR studies of φ6 P2. While an analysis of the dynamics of all backbone and sidechain nuclei remains our goal, as a first step, we present here an analysis of the spin-relaxation of multiple quantum coherences involving ^{13}C and ^{1}H nuclei at the δ1 position of the Ile residues in P2 (66-68).

2.2. Materials and methods

2.2.1. Overexpression and Purification of φ6 P2

The φ6 P2 expression plasmid was constructed by PCR amplification of the pLM687 (30) plasmid that encodes of the cystoviral L segment with recombinant Pfu DNA polymerase (Stratagene) by using the following oligonucleotides—5' GGTAAGCGCCATATGCCGAGGA GA-3' and 5'TACGAATTCCGGCATGATTACCTAGGCATTACA-3' as upstream and downstream primers respectively. The P2 gene fragment was digested with NdeI and EcoRI, and ligated to a digested pET28a(+) vector (Stratagene). N-terminal His6-tagged full-length P2 was transformed into Escherichia coli BL21-DE3 (Stratagene) and grown overnight in 20 ml of unlabeled M9 medium prepared in 100% H_2O and supplemented with kanamycin (50 mg/l). Following overnight growth, the bacteria were adapted for growth in media containing progressively higher concentrations of D_2O from 10 to 100% at 37°C. The final overnight culture (20 ml grown in M9 medium prepared in 100% D2O) was used to inoculate 1 l of M9 medium in 100% D_2O containing 3 g/l of U-[^{12}C,^{2}H]-glucose (CIL) and 1 g/l of $^{15}NH_4Cl$ (CIL) as the principal sources of carbon and nitrogen respectively. When the optical density at 600nm reached 0.6, 50 mg of 2-keto-3-d_2-4-^{13}C-butyrate (CIL) was added into the growth medium. After 1 h of growth, the culture was placed on ice for 30 min and induced with 1mM IPTG. The growth was continued for 20 h at 16°C. The cells were harvested by centrifugation at 5,000 rpm and resuspended in lysis buffer containing 20mM Tris–HCl (pH 7.2), 300mM NaCl, 5mM Imidazole and two tablets of Complete Mini EDTA-free protease inhibitor (Roche). The suspension was lysed by sonication and the lysate was centrifuged at 16,000 rpm for 30 min at 4°C.

The soluble lysate was then loaded onto a nickel affinity column (Ni–NTA agarose, Qiagen), and washed extensively with buffer containing 20mM Tris–HCl (pH 7.2), 300mM NaCl, and 20mM imidazole. The bound recombinant P2 was eluted with buffer containing 20mM Tris– HCl (pH 7.2), 300mM NaCl, 500mM imidazole and the dialyzed against the NMR buffer that contained 20mM bis-tris, 400mM NaCl and 4mM DTT at pH6.5. The uniformly-^{15}N,^{2}H-Ile-δ1-^{13}C,^{1}H-labeled protein was concentrated by using Amicon Ultra-4 or Amicon Ultra-15 centrifugal concentrators (Millipore) for NMR analyses.

2.2.2. Assignment of Ile δ1 Resonances

All NMR experiments were performed using a Varian Inova 600MHz spectrometer equipped with a triple-resonance {^{1}H,^{13}C}-cryogenic probe capable of applying pulsed field gradients along the z-axis. Resonance assignment of the 25 Ile δ1 resonances was achieved using an Ile-to-Leu mutation strategy described below (**Figure 2.2**). Twelve Ile-to-Leu triple mutants were designed specifically avoiding simultaneous mutation of Ile residues that were sequentially or spatially proximal in the same set. In addition 13 single Ile-to-Leu mutants were also created. Mutations were introduced into the cDNA corresponding to wild-type φ6 P2 using the QuikChange Site-directed Mutagenesis and QuikChange Multisite-directed Mutagenesis kits (Stratagene) using primers designed utilizing the QuikChange Primer Design Program (Stratagene). The mutations were confirmed by sequencing. The mutated plasmids were transformed into E. coli BL21-(DE3)-RIPL competent cells (Stratagene) and each uniformly-^{15}N,^{2}H-Ileδ1-^{13}C,^{1}H-labeled mutant P2 was overexpressed and purified as described previously. The resonances corresponding to mutated Ile δ1 positions vanished in an ^{13}C,^{1}H HMQC spectrum. Of 25 Ile-δ1 resonances, 15 were assigned using the strategy employing triple-

mutants. However, some of these mutations resulted in protein misfolding and could not be used for assignments. Consequently, assignments for 8 of the remaining 10 Ile δ1 resonances were obtained using Ile-to-Leu point mutants. Resonances corresponding to Ile211 and Ile283 could not be unambiguously identified.

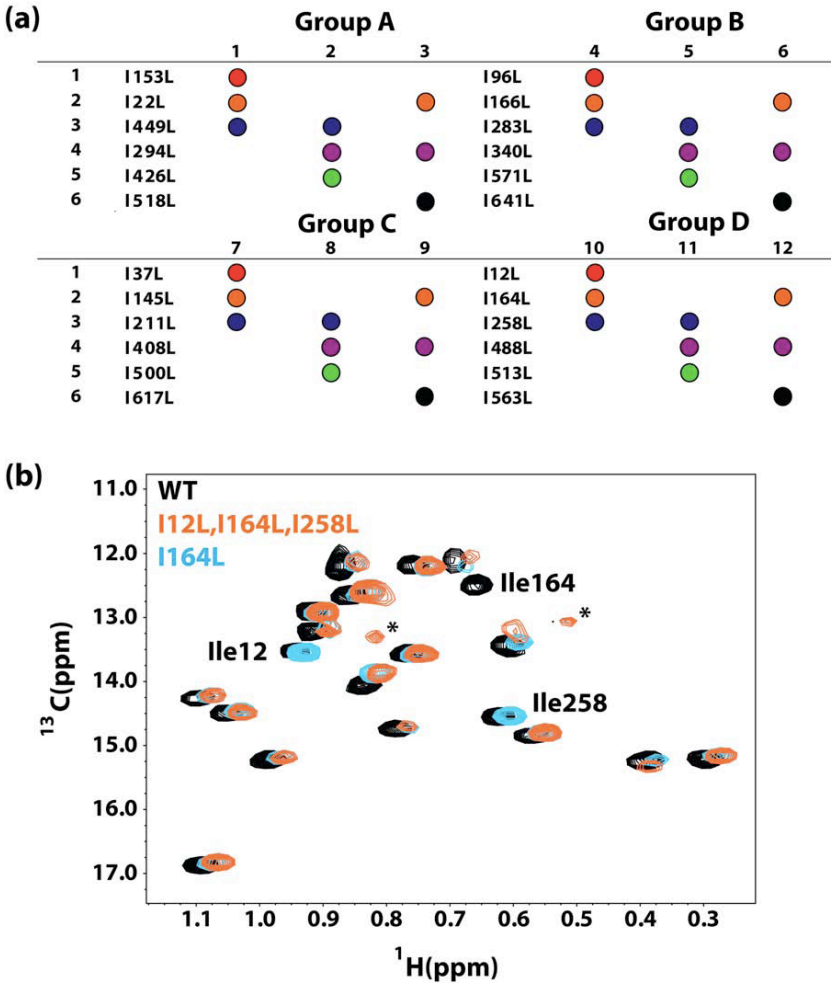


Figure 2.2. Strategy to Obtain Resonance Assignments for the Ile δ1 Positions in φ6 P2.
(a) Ile-to-Leu triple mutants. The 12 triple mutants were divided into 4 groups (Figure A, B, C and D) containing 3 triple mutants each. For example, the six single mutations I153L, I22L, I449L, I294L, I426L and I518L were specific to group A. Each triple 3 mutant belonging to group A shared only two mutations with the other group members. (b) Expansion of the central regions from representative spectra for single and triple mutants. ^{13}C, ^{1}H HMQC spectra for wild-type (WT) φ6 P2 (black) is shown along with those of triple mutant 10 from group D (orange) and the I164L point mutant (blue). The spectrum was acquired at a ^{1}H field strength of 600MHz at 25°C in a buffer containing 20mM Bis–Tris, 400mM

NaCl, 4mM DTT at pH=6.5. Spectra corresponding to wild-type P2 have been shifted to ease visualization. Impurities are indicated by '*'.

As **Figure 2.2** illustrates, group A consisted of mutants 1 (I153L, I22L, I449L), 2 (I449L, I294L, I426L) and 3 (I22L, I294L, I518L). The other sets of triple mutants were constructed in similar fashion. For a set of triple mutants within a group, the resonance corresponding to each Ile δ1 position was expected to disappear in the ^{13}C, ^{1}H HMQC spectra of one or two mutants (uniformly ^{2}H, Ile-δ1-^{1}H, ^{13}C labeled), as shown by color coding (filled circles), but never in all three. To assign resonances using the triple-mutants, the following strategy was used – e.g. for group A, the resonance corresponding to Ile449 (shown as a blue circle) vanishes from the ^{13}C,^{1}H HMQC spectrum of mutant 1 (I153L, I22L and I449L) and mutant 2 (I449L, I294L and I426L), while it remains in that of mutant 3 (I22L, I294L and I518L). The resonance corresponding to Ile294 (depicted by a purple circle) disappears from the spectra of mutants 2 and 3 and remains in that for mutant 1. On the other hand, the resonance corresponding to Ile426 (green circle) only vanishes in the spectrum of mutant 2. Considering this information together allows the assignment of resonances corresponding to Ile294, Ile426 and Ile449. Other assignments within each group were generated following the same strategy. In those cases where the triple mutants lead to protein destabilization/unfolding (10 of 25), assignments were carried out using Ile-to-Leu single point mutations. The resonances corresponding to Ile22, Ile96, Ile145, Ile164, Ile408, Ile446, Ile500, Ile571, Ile617 and Ile641 were confirmed using single mutants. The resonances corresponding to Ile211 and Ile283 could not be unambiguously assigned.

Impurities, as indicated by '*', can be removed by gel filtration, however this additional purification step resulted in a significant loss in protein yield, and was not performed for some mutants.

2.2.3. Chemical Shift Changes

The concentrations of Mg^{2+} (states **2–6**, **Table 2.1**), of GMPPNP (Sigma-Aldrich) (states 3, 4 and 5) and of GMPCPP (Jena Biosciences) (states **3A** and **6**), were held constant at 5, 4 and 4mM, respectively. The concentrations of the ssRNA were maintained in 3.5–5-fold excess of the concentration of P2 (250–350 μM).

Scaled chemical shift changes ($\Delta\delta_{i/j}$ in Hz) between states i and j (i, j=**1, 2, 3, 3A, 4, 5, 6**; Table 2.1) were calculated using the following equation:

$$\Delta\delta_{ij} = \nu_0\sqrt{\left(\delta_{i,H}-\delta_{j,H}\right)^2+\left(\frac{\gamma_C}{\gamma_H}\left(\delta_{i,C}-\delta_{j,C}\right)\right)^2} \quad i \neq j \tag{2}$$

where $\delta_{i/j,C/H}$ are the $^{13}C/^{1}H$ chemical shifts in states i/j in $^{13}C,^{1}H$ HMQC spectra and ν_0 is the spectrometer ^{1}H frequency in MHz (600 MHz in the present case).

Table 2.1. Complexes of φ6 P2 Utilized in This Study

State	Complex
State **1**	P2
State **2**	P2, Mg^{2+}
State **3**	P2, Mg^{2+}, GMPPNP
State **4**	P2, Mg^{2+}, GMPPNP, 5'-UUUCC-3'
State **5**	P2, Mg^{2+}, GMPPNP, 5'-UUUUC-3'
State **3A**	P2, Mg^{2+}, GMPCPP
State **6**	P2, Mg^{2+}, GMPCPP, 5'-UUUCC-3'

2.2.4. NMR Relaxation Measurements

All relaxation experiments were performed at 25°C and 600MHz using sweep-widths of 9000 and 1600 Hz in the direct (^{1}H, 512 complex points) and indirect (^{13}C, 32 complex points) dimensions respectively. The ^{13}C carrier was positioned at 13.15 ppm at the center of the Ile

^{13}Cδ1 resonances. P2 concentration varied from 250 to 350 mM for the cross-correlated relaxation measurements and was around 250 mM for the relaxation dispersion measurements. All fits to theoretical curves both for the relaxation analyses [Equations (2) and (4)] and for the fluorescence measurements [Equation (6)] were performed using in-house programs that utilize the ODRPACK subroutines (69). Errors in the fitted parameters were obtained from the inverse covariance matrices of the fits and included both the random as well as the model selection errors.

2.2.5. Measurement of Intra-methyl ^{1}H–^{1}H Cross-correlation Rates

In order to measure the intra-methyl ^{1}H –^{1}H dipole-dipole cross-correlated relaxation rate (η) (70) a set of two experiments were recorded. The signal intensity in the first experiment (A) during a variable relaxation delay (T) depends only on spin-dynamics in the I=3/2 manifold of the Ile δ1 methyl H_3 spin-system. In the second experiment (B), the signal intensity (I_B) is affected by spin-dynamics in the I =3/2 and I =1/2 manifolds. The cross-correlation rate (η) is related to the intensity ratio (I_A/I_B) in the following way (70):

$$\frac{I_A}{I_B} = \frac{-0.5\eta \tanh\left(\sqrt{\eta^2+\delta^2}T\right)}{\sqrt{\eta^2+\delta^2}T - \delta\tanh\left(\sqrt{\eta^2+\delta^2}T\right)} \qquad (2)$$

where δ depends on the ^{1}H–^{1}H cross-relaxation with external protons. Seven values of T =1, 2, 4, 6, 8, 10 and 14ms were used with 256 and 64 transients per t_1 point for experiments A and B, respectively. η and δ values were obtained by fitting the experimentally determined I_A/I_B ratio as a function of T to Equation (2). The order-parameters (S^2_{axis}) were calculated from the following equation (70) using the rotational correlation time (τ_C) (assuming that the rotational

diffusion was isotropic) estimated from hydrodynamic modeling:

$$S^2_{axis} = \frac{10}{9}\frac{r^6_{HH}\eta}{\gamma^4_H\hbar^2\tau_C P_2\left(\cos\left(\theta_{HH}\right)\right)^2}$$
$$P_2\left(\cos\left(\theta_{HH}\right)\right) = \frac{3\cos\theta^2_{HH} - 1}{2} \quad (3)$$

where r_{HH}=1.813 Å, θ_{HH} = 90º and all other symbols have their usual meaning.

2.2.6. Multiple–quantum Relaxation Dispersion Measurements

^{13}C, ^{1}H multiple quantum relaxation dispersion curves were obtained for states 1–6 (Table 2.1) at 600 MHz using the pulse sequences developed by Korzhnev et al. (71). Fifteen values of the RF–field strength (ν_1=1/2τ_{CPMG}, where τ_{CPMG} is the spacing between successive π pulses)=31.25, 62.5, 93.75, 125.0, 156.25, 187.5, 250.0, 312.5, 375.0, 437.5, 500.0, 625.0, 750.0, 875.0 and 1000 Hz were used with a constant relaxation delay of 32 ms. 128 transients were collected per t_1 point. Data at each field were acquired in an interleaved fashion with all ν_1 values collected prior to t_1 incrementation in order to minimize possible errors resulting from sample heating and/or precipitation and to increase the accuracy of the fitted exchange parameters. All relaxation dispersion data were analyzed assuming exchange between two chemically distinct sites A and B with an exchange rate $k_{ex} = k_{A\to B} + k_{B\to A,}$ where $k_{A\to B}$ and $k_{B\to A}$ are the rates for the forward and reverse process respectively. In the limits of fast exchange, k_{ex} >> $(\Delta w_H \Delta w_C)^{1/2}$, ($\Delta w_H$, Δw_C are the differences between the chemical shifts of the "ground" and "excited" states for ^{1}H and ^{13}C nuclei respectively) and when $\Delta w_H \to 0$, the effective multiple-quantum relaxation rate ($\Gamma_{MQ,eff}$) for a given RF-field, ν_1 is given by (71)

$$\Gamma_{MQ\,eff}\left(\nu_1\right) = \Gamma_{MQ}\left(\infty\right) + \frac{2p(1-p)\Delta\omega_H\Delta\omega_C}{k_{ex}}\left(1 - \frac{4\nu_1}{k_{ex}}\tanh\left(\frac{k_{ex}}{4\nu_1}\right)\right) \quad (4)$$

Where p is the population of the "ground-state" and $\Gamma_{MQ}(\infty)$ is the multiple-quantum relaxation rate at infinite field strength i.e. the value of $\Gamma_{MQ,eff}$ in the absence of chemical exchange. The exchange rates (k_{ex}) were obtained by fitting the experimental dispersion curves to Equation 4.

A point to be made here is that for practical considerations, using Equation (4) should not result in substantial errors in k_{ex} values provided that the product, $\Delta\omega_H\Delta\omega_C$, is small and the dynamic timescale lies in the fast exchange limit. However, errors in $\Gamma_{MQ}(\infty)$ and ξ, using this simplified approach, are larger and correlated with each other. Therefore we do not attempt to interpret $\Gamma_{MQ}(\infty)$ and ξ values obtained in the present study.

2.2.7. Fluorescence Measurements

All fluorescence anisotropy experiments were performed using single-stranded RNA oligos (5'-UUUCC-3' and 5'-UUUUC-3') labeled at the 5′-end with 6-carboxyfluorescein. Oligos were purchased PAGE purified from Dharmacon RNAi Technologies. Experiments were performed using a Panvera Beacon 2000 fluorescence anisotropy system. Fluorescence intensity data were acquired by passing the emission following an excitation at 490nm through a 520nm cutoff filter. A 10 s integration time was used for four combinations of the horizontal (H) and vertical (V) polarization filters. The measured intensities using parallel (I_{VV}) and perpendicular (I_{VH}) polarizer settings were converted into fluorescence anisotropy using the following equation:

$$A = \frac{I_{VV} - I_{VH}}{I_{VV} + 2I_{VH}} \tag{5}$$

Anisotropy values (A) were measured in the presence of increasing amounts of P2. Values of the dissociation constant (K_d) were obtained by fitting the dependence of A on P2 concentration using the following equation:

$$A(C) - A(0) = (A_\infty - A(0))\frac{K_d + C_{RNA} + C - \sqrt{(K_d + C_{RNA} + C)^2 - 4C_{RNA}C}}{2C_{RNA}} \tag{6}$$

where A(C) is the anisotropy when the concentration of P2 is C and $A\infty$ is the limiting anisotropy value. The concentration of RNA (C_{RNA}) was kept constant at 5 nM. All experiments were performed at pH 6.5 in the NMR buffer. Absence of non-specific binding to the fluorescein moiety was confirmed by measuring A values of fluorescein alone (300 nM) and in the presence 40 and 400 mM P2.

In order to measure the affinity of the GTP-analogs GMPPNP and GMPCPP towards P2, tryptophan (φ6 P2 contains 12 Trp residues) fluorescence spectra were obtained using a Spex Fluorolog-3 (Horiba Jobin Yvon) spectrometer with excitation and emission slits set to 1.6 nm. Fluorescence emission spectra (300–450 nm) were acquired in 0.5-nm steps and signal-averaged over three scans. Emission intensities were obtained by averaging the values at the maximum± 4 nm. The excitation wavelength was set to 282nm for each addition of either GMPCPP or GMPPNP to the protein sample in NMR buffer. Apparent K_d values were obtained by using fits to an analog of Equation (6) with the anisotropy values replaced by the corresponding emission intensities, the C_{RNA} by the concentration of P2 (held constant at 10 μM) and C by $C_{GMPNP/GMPCPP}$ representing the variable concentration (0–2.5mM) of the GTP analogs.

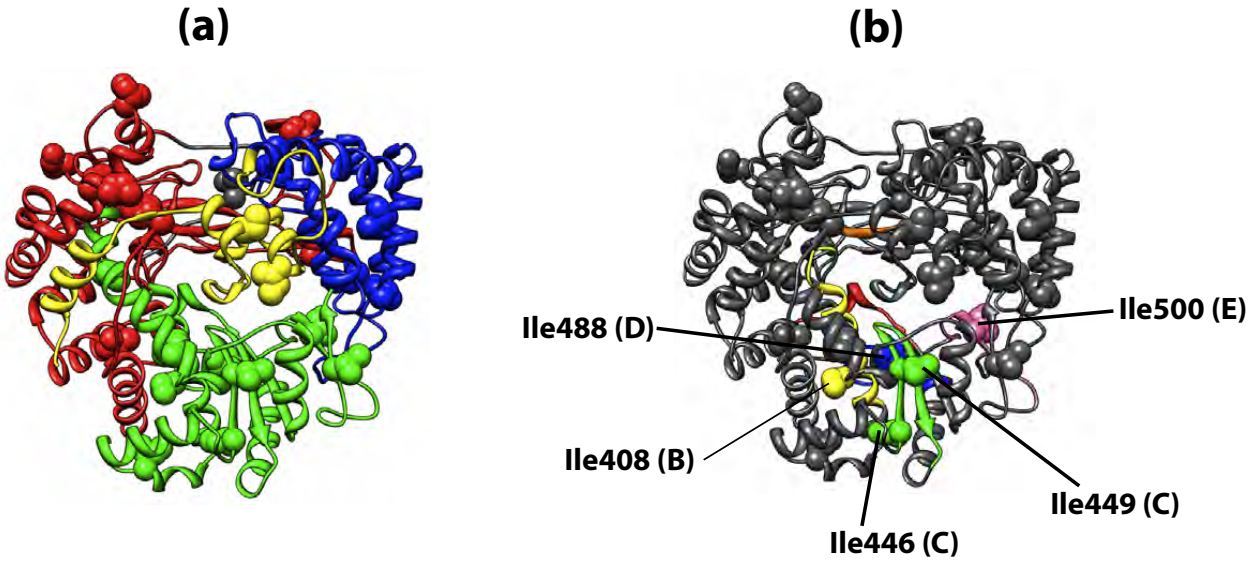
2.2.8. Hydrodynamic Modeling

The diffusion tensor for full-length P2 was estimated by utilizing the bead model as implemented in the HYDRONMR program (72). The effective radius of the atomic elements was set to 3.2 Å, the temperature to 298.15K and the viscosity to 9.11 cP. The minibead solvent shell radius was calculated automatically. The atomic coordinates of P2 were obtained from structures with the following PDB IDs: 1HHS, 1HHT, 1HI0, 1HI8, 1UVI, 1UVJ, 1UVL, 1UVM and 1UVN.

2.3. Results and Discussion

2.3.1. Chemical Shift Assignment of Ile in φ6 P2

$^{13}C,^{1}H$ HMQC spectra of uniformly-deuterated, Ile-δ1-[$^{13}C,^{1}H$]-labeled P2 showed excellent dispersion and all 25 expected Ile δ1 ^{13}C, ^{1}H resonances were seen/resolved (**Figure 2.3c**). Resonance assignments were carried out using a set of Ile-to-Leu single and triple mutants. The methodology employing triple-mutants has been described in **Section 2.2.2**. In all 23 of 25 Ile resonances could be unambiguously assigned using this strategy. Resonances corresponding to Ile211 and Ile283 (indicated by a '*" in **Figure 2.3c**) could only be tentatively assigned and relaxation data for these residues were not interpreted further.



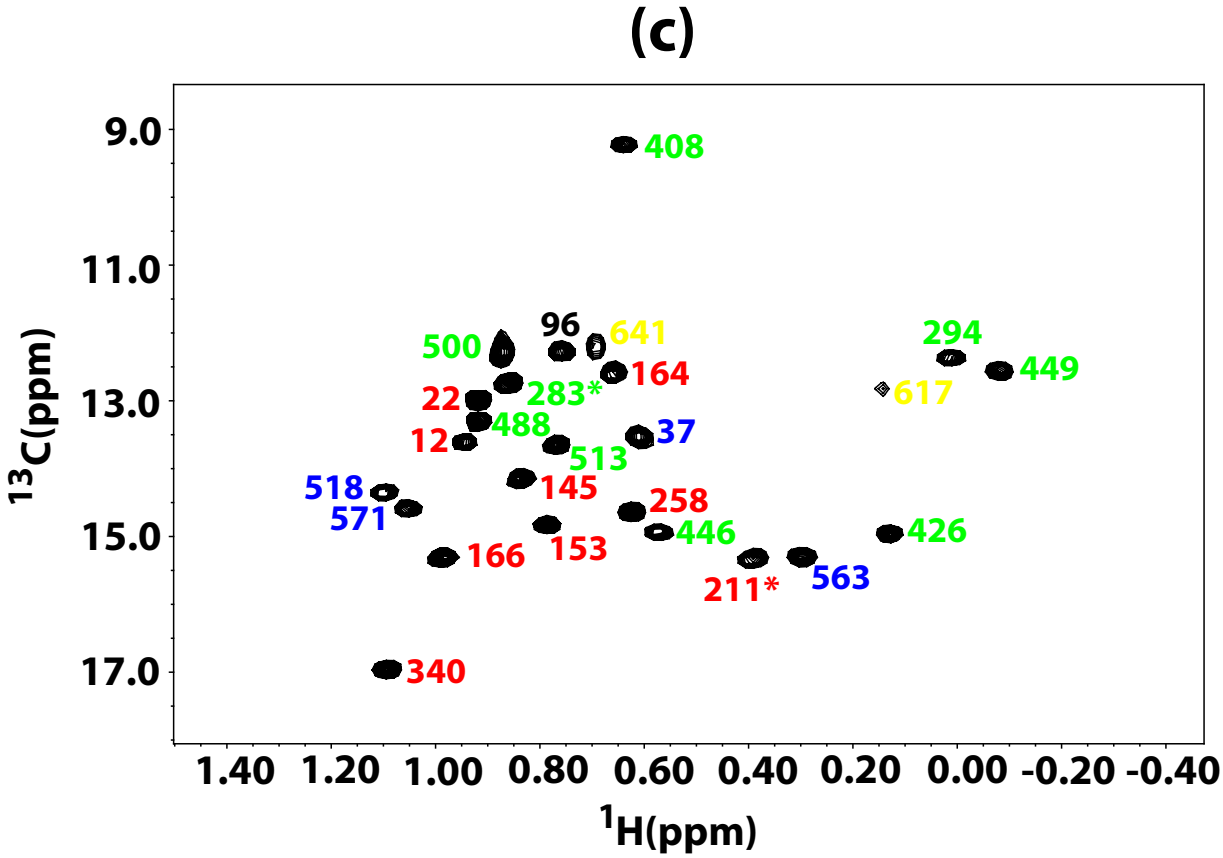


Figure 2.3. NMR Studies Using the 25 Ile δ1 Positions as Probes.
(a) Distribution of Ile residues over the P2 tertiary structure. The fingers (1-30, 104-276, 333-397), thumb (37-91,518-600), palm (277-332, 398-517) and C-terminal domains (601-664) are colored red, blue, green and yellow respectively. The Ile residues are shown as spheres and colored according to the domains they belong to: fingers (12, 22, 145, 153,164, 166, 211, 258 and 340); thumb (37, 518, 563 and 571); palm (283, 294, 408, 426, 446, 449, 488, 500 and 513) and C-terminal (617 and 641) domains. Ile96 that lies on the fingertips is colored dark grey. (b) Conserved RdRP sequence motifs in P2. Motif A (324-334, red), motif B (394-413, yellow), motif C (445-461, green), motif D (475-488, blue), motif E (493-505, pink) and motif F (268-272, orange) are shown. The Ile residues that lie on motifs A-F are colored accordingly and labeled. Motif C contains the ^{452}G/SDD^{454} sequence motif characteristic of RNA and DNA polymerases. (c) Assigned ^{1}H,^{13}C HMQC spectrum of uniformly deuterated Ile-Cδ1-^{13}C,^{1}H-labeled P2. The spectrum was acquired at a ^{1}H field strength of 600 MHz at 25°C in a buffer containing 20 mM bis-tris, 400 mM NaCl, 4 mM DTT at pH = 6.5. The resonances have been labeled according to the color scheme described in the (a). Resonances corresponding to Ile211 and Ile283 could not be unambiguously assigned using the Ile-to-Leu mutation strategy and these assignments indicated by '*' should be considered tentative.

2.3.2. Choice of Binary and Ternary Complexes of φ6 P2

We analyzed the spin relaxation rates of the Ile δ1 positions for P2 in seven states as following, the apo state, in complex with Mg^{2+}, with GTP analogs—GMPPNP and GMPCPP, and two different 5 nucleotides constructs ssRNA constructs (**Table 2.1**). The RNA construct, 5'-UUUCC-3', chosen for states **4** and **6**, corresponds to the final 5-nt from the 3'-end of the s^-, m^- and l^- RNA genomic strand. Previous biochemical study for P2 indicated that transcription is far more efficient for the s^- and m^- strands compared with that of l^- (that contains a 5'-UUUAC-3' sequence at the 3'-end). It suggested that the penultimate nucleotide from the 3'-end sequence played an important role when P2 processing s, m, and l ssRNA (17). Based on the biochemical study, the second 5-nt ssRNA construct, 5'-UUUUC-3', was chosen for spin relaxation study (state **5**), while introducing a minimal size differential.

The choice of the ternary complexes of P2 with substrate (Mg^{2+}/GTP analogs) and ssRNA warrants some discussion (**Figure 2.4**). P2 initiates the *de novo* RNA synthesis without the use of a primer. The initiation NTP acts as a one-nucleotide primer and base-pairs with the extreme 3'-nucleotide (2) of the template forming the first nucleobase (D1) of the daughter strand. The second substrate NTP (D2) base pairs with the penultimate template nucleotide from the 3'-end (T2). Formation of an initiation complex requires both T1-D1 and T2-D2 base-pairing to occur. The first phosphodiester bond is thus formed between D1 and D2 with the cleavage of the latter at the α-position with release of pyrophosphate. The process continues with each incoming substrate nucleotide (CTP, UTP, ATP or GTP) complementary to the corresponding template

nucleobase (38). Thus, among the ternary complexes studied here, state **4** (**Table 2.1**) represents a dead-end or stalled complex where the first phosphodiester bond between D1 and D2 can form since GMPPNP is a viable substrate for P2 given the fact that it can be cleaved between the α and β phosphates (29). However, additional polymerization beyond the one-nucleotide addition cycle and formation of a 2-nt daughter chain (PPP-G-P-G) cannot proceed, as the substrate nucleotide (A from ATP or an ATP-analog) complementary to T3 (U) is unavailable. Examples of stalling through nucleotide deprivation in φ6 P2 has been described in the literature (73) and characterized structurally (40). In state **5** (**Table 2.1**) a complex capable of forming a stable initiation platform cannot form since the complementary nucleotide D2 (A from ATP or analog) for T2 (U) is unavailable. State **6** (**Table 2.1**), on the other hand, represents a stable initiation complex, stabilized by proper base-pairing of D1 with T1 and D2 with T2 (D1,2 = GMPCPP and T1,2 = C). However the phosphodiester linkage between D1 and D2 cannot form since in GMPCPP the bond between α and β phosphates cannot be cleaved (**Figure 2.4**).

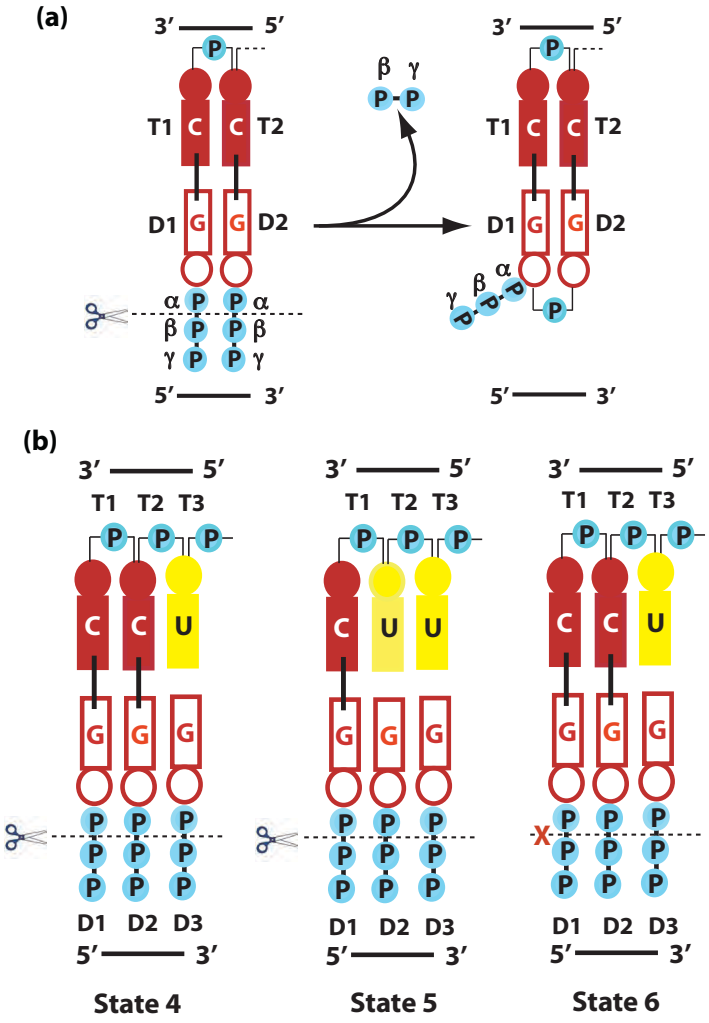


Figure 2.4. Choice P2/RNA/NTP Complexes.
(a) Formation of a stable initiation complex results from the proper base pairing of the template bases T1 (C) and T2 (C) (numbered 3'→5') with the substrate (GTP analog) bases D1 (G) and D2 (G). Nucleophilic attack by the 3'-OH of D1 and cleavage between the α and β-phosphate groups of D2 (represented by the dotted line), results in the formation of the first 3'-5' phosphodiester bond of the daughter chain between D1 and D2 (numbered 5'→3'), releasing pyrophosphate. The requirements for the reaction to proceed are thus, (2) proper T1-D1 and T2-D2 base pairing (indicated by solid lines) and (2) cleavability of the substrate (indicated by dashed lines) between the α, β-phosphates. (b) Schematic representation of the three ternary complexes (states 4, 5 and 6) between P2, substrate GTP analogs and 5nt ssRNA templates. State 4: T1(C)-D1(G) and T2(C)-D2(G) are correctly base-paired, the substrate GMPPNP is α - β cleavable and the reaction proceeds for one cycle generating 5'-PPP-G-P-G-3'. The reaction subsequently stalls since the base (D3, A) complementary to T3 (U) from ATP (or an analog) is not available. State 5: The substrate GMPPNP is α - β cleavable but proper base paring between T2 (U) and D2 (G) cannot occur. State 6: T1-D1 and T2-D2 properly base pair, however the substrate, GMPCPP, is not is α - β cleavable and the reaction cannot proceed. Schematic illustrations of different states are presented for easy visualization.

2.3.3. Fast Dynamics in φ6 P2

We investigated the ps-ns timescale dynamics of the Ile δ1 positions for states 1 through 4. Beginning with the apo enzyme (state **1**), we measured the difference between the relaxation rates of the fast and slow relaxing coherences corresponding to $I = 3/2$ manifold of the methyl H_3 spin-system at each Ile δ1 positions. The extracted difference relates to the intra-methyl ^1H–^1H dipole–dipole cross-correlation rate (η), which was proportional to the order parameter in Equation (3). When values are close to 0, residues are highly disordered on the ps–ns timescale. Whereas S^2_{axis} values are close to 1, residues are ordered on this timescale. 21 Ile resonances for the four states were analyzed. The representative fits to Equation (2) and the η values for the 21 Ile resonances are shown in **Figure 2.5b**.

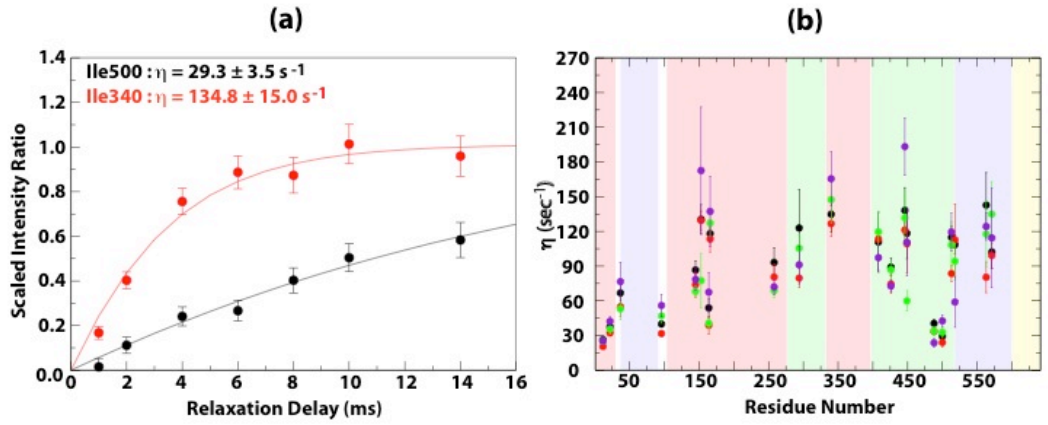


Figure 2.5. Fast Dynamics at the Ile δ1 Positions.
(a) Representative fits to Equation (2) for Ile340 and Ile500 are shown for P2 in the apo state (state **1**). The scaling factor used in the y-axis is the ratio (4) of the number of scans used in experiments A and B. (b) η values for the Ile residues in P2 are plotted against residue number. η values for states **1**, **2**, **3** and **4** are represented by black, red, green and purple circles respectively. The regions corresponding to the fingers, thumb, palm and C-terminal domains are shaded red, blue, green and yellow respectively. Data corresponding to Ile211, Ile283, Ile617 and Ile641 were not analyzed.

Prior to the availability of backbone ^{15}N, ^1H assignments and the analysis of backbone

^{15}N relaxation rates, we relied on hydrodynamic calculations (74) to estimate the nature of the rotational diffusion tensor of P2 (see **Section 2.2.8**). These calculations predicted a nearly isotropic diffusion tensor [$2D_{zz}/(D_{xx}+D_{yy})$ = 1.07 ± 0.03] with a corresponding rotational correlation time (τ_C) of 40.15 ± 0.76 ns. Thus using a τ_C value of 40 ns and isotropic diffusion we estimated S^2_{axis} values for states **1**-**4** using Equation (3).

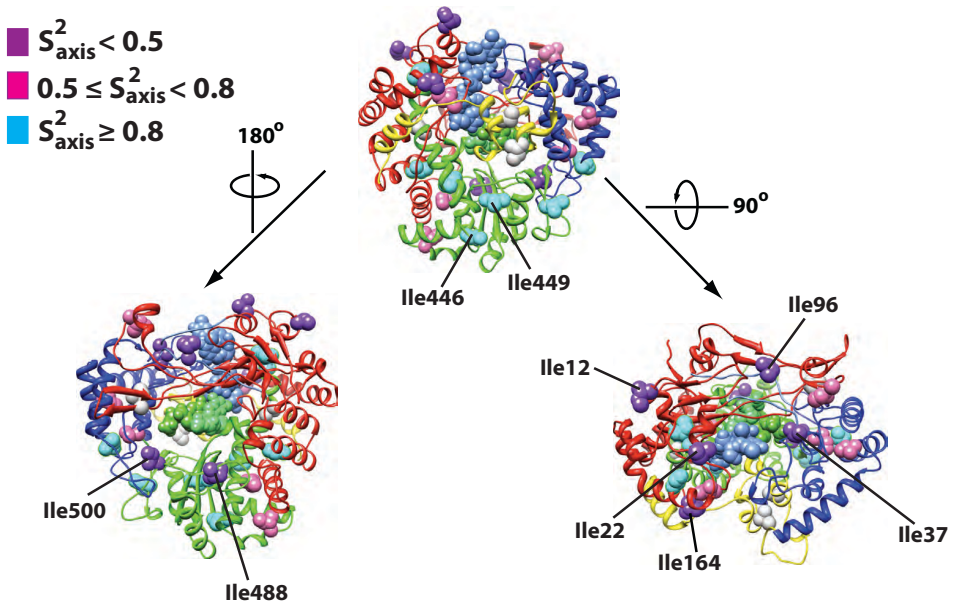


Figure 2.6. S^2_{axis} Values for the Ile Cδ1 Positions in apo-P2 (state 1).
The Ile residues are indicated by spheres colored according to their respective S^2_{axis} values. Three views of P2 are shown and key residues labeled. Residues for which data could not be analyzed (Ile211, Ile283, Ile617 and Ile641) are indicated by light grey spheres. The two substrate GTP molecules are shown in green and the template ssRNA in blue.

Apo P2 displayed an average S^2_{axis} value of 0.63 ± 0.27 (10% trimmed — 0.60 ± 0.22). The highest degree of order was seen for the palm domain residues located in and around the catalytic site that bears the conserved catalytic DD motif (445–461; Ile446, 0.96±0.14 and

Ile449, 0.82±0.15). Whereas the most dynamic Ile residues were located on the edge of the template tunnel at the top the fingers domain—Ile12 (0.18 ± 0.01), Ile22 (0.26 ± 0.01), and Ile164 (0.37 ± 0.06); Ile37 (0.46 ± 0.07) on the thumb domain; and Ile96 (0.28 ± 0.02) located on the fingertips. Compare to the Apo state, the overall patterns of fast dynamics in states **2–4** were similar. But there were some very large variations in fast dynamics timescale were seen in several residues, especially around the catalytic site for the motif C (residues Ile446 and Ile449). Ile449, located near the top of the β-hairpin bearing the catalytic Asp residues displayed increased disorder in the complex with substrate GMPPNP/Mg^{2+} (state **3**, S^2_{axis}= 0.42 ± 0.06 compared with 0.82 ± 0.15 for state **1**). This motion is indicative of local conformational change to facilitate stabilizing local interactions with RNA templates and substrates. Ile446, located at the base of the hairpin became highly ordered in state **4** (S^2_{axis}= 1.34 ± 0.27). This extreme rigidity near the catalytic site possibly reflects a distinct state for the stalled enzyme. Meanwhile, Ile449 (S^2_{axis}= 0.77 ± 0.33) is also ordered in state **4**. As will be discussed at length below, both the analysis of chemical shift perturbations as well as the dynamic characteristics on the μs-ms timescale also seemed to indicate that state **4** was quite different from all other states considered.

While the errors in the experimental η values (and consequently the S^2_{axis} values) were relatively large, the overall rigidity of the stalled complex (state **4**, S^2_{axis} = 0.64 ± 0.20, 10% trimmed — 0.57 ± 0.18) while comparable with that of the apo enzyme (state **1**) was higher than both state **2** (Mg^{2+}, S^2_{axis} = 0.54 ± 0.25, 10% trimmed — 0.51 ± 0.20) and state **3** (GMPPNP/Mg^{2+}, S^2_{axis} = 0.57 ± 0.27, 10% trimmed — 0.52 ± 0.21). The increased fluctuations on the fast timescale in the Mg^{2+}/GTP complexes could be an indication of conformational fine-tuning of the enzyme to facilitate the processing of a variety of ssRNA templates and substrate NTPs.

In states **1-4**, S^2_{axis} values showed highest degree of flexibility on the ps-ns timescale for Ile residues in close proximity to the entry portals for the substrate NTPs and the template ssRNA. In the contrast, two Ile residues located on the catalytic motif C showed high rigidity, except for the increase disorder at the top of this motif in the Mg^{2+}/GMPPNP loaded state (state **3**). The fast dynamics seems connected to loading of both substrates and templates. This increased dynamics likely helps modulate local interactions allowing a variety of substrates and templates to be processed efficiently.

2.3.4. Binding Affinity Towards Single-stranded RNA and GTP-analogs

The ternary complexes (P2/Mg^{2+}/GMPPNP/ssRNA) studied here involved two different ssRNA constructs. Prior to analyzing the dynamics on the μs-ms timescale it was necessary to confirm that perceived motional differences between the complexes was not related to the affinities of the two RNA constructs towards P2 i.e. to simple on/off processes. In order to measure the binding affinity of the two 5nt ssRNA constructs towards P2 we performed fluorescence anisotropy measurements using 5'-fluorescein tagged RNA.

Fluorescein, in its singly protonated form (pK_a = 6.4) has a reduced molar absorptivity (ε = 29000 M^{-1} cm^{-1}) compared with the corresponding dianionic form (ε = 76900 $M^{-1}cm^{-1}$) (75) that is the dominant species above pH = 7. However, we chose to perform these experiments using the NMR buffer (pH = 6.5) for consistency with the relaxation studies for states **4, 5**, and **6**. In spite of the reduced fluorescence, changes in fluorescence anisotropy could be monitored with a high degree of accuracy and the resultant binding isotherms could be fit to Equation (6) to obtain apparent K_d values (**Figure 2.7**). These studies revealed that P2 has a higher affinity for 5'-UUUUC-3' (K_d = 6.6 ± 0.4 μM) than for 5'-UUUCC-3' (K_d = 50.0 ± 1.7 μM). No significant

changes in binding affinity were noted in the presence of GMPPNP. Based on the binding affinities of the ssRNA templates and their concentration (~mM) in the NMR experiments one can estimate a $k_{ex} = k_{on}(C_{RNA}+K_d) \sim 10^5\ s^{-1}$ for exchange between the RNA-bound and free states of P2 assuming a diffusion limited on-rate. This value is about two orders of magnitude higher than the largest k_{ex} values ($\sim 10^3\ s^{-1}$) described below.

The binding affinities of the two GTP analogs towards P2, as estimated from tryptophan fluorescence quenching (data not shown) were quite similar (GMPPNP, K_d = 317 ± 31 μM, GMPCPP, K_d = 271 ± 31 μM) and comparable to that seen in other RdRPs (76). Here again, the expected k_{ex} values for simple on-off processes were expected to be 50- to a 100-fold higher that the highest k_{ex} values seen here.

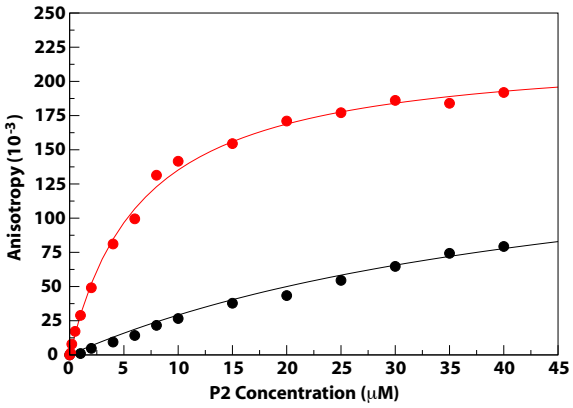


Figure 2.7. Binding Affinity of the 5-nt ssRNA Constructs Towards P2.
Data for 5′-UUUCC-3′ (black) and 5′-UUUUC-3′ (red) are shown. The experimental data points are represented by the solid circles and the fitted curves are depicted by the solid lines.

2.3.5. Slow Dynamics

In order to investigate conformer sampling on the μs-ms timescale in P2 we performed an analysis of the dependence of the Ile-Cδ1 $^{13}C,^{1}H$ multiple-quantum (MQ) relaxation rates ($\Gamma_{MQ,eff}$) on the strength (ν_1) of an applied the radiofrequency (57) field (MQ relaxation dispersion)

(71). These rates can be measured with high sensitivity in large proteins where traditional single-quantum (SQ) relaxation dispersion experiments are likely to fail (71). Further, MQ relaxation dispersion is uniquely sensitive to slow dynamics especially in the fast-exchange regime (77). The dependence of the effective MQ relaxation rates ($\Gamma_{MQ,eff}$) on the strength (ν_1) of the applied RF-field is extremely complex and cannot, in general, be treated using the Carver-Richards formalism used for SQ relaxation (71). However, such an exact treatment requires high-quality data at multiple static magnetic fields. In the fast exchange regime, when the chemical shift differences between the "ground" and "excited" states (assuming a simple two-state exchange model) for the two nuclei participating in the MQ coherence, is small, a simple functional form may be used to extract the exchange rates (k_{ex}) (78). We use this approach in the present case especially since the chemical shift changes seen (assuming that these changes are reflective of the difference between the ground and excited states) are sufficiently small so that the fast exchange condition $k_{ex} >> (\Delta\omega_H \Delta\omega_C)^{1/2}$ (where $\Delta\omega_H$ and $\Delta\omega_C$ are the differences between ground and excited state chemical shifts for ^1H and ^{13}C nuclei respectively) was maintained in all of the cases analyzed below.

For apo-P2 (state **1**), the k_{ex} values varied from 449 s^{-1} to 1690 s^{-1} (**Table 2.2**). However, these rates could be broadly classified into two distinct groups – group I that had higher exchange rates (k_{ex}) implying motion on a faster timescale (mean $k_{ex} = 1510 \pm 168$ s^{-1}) and group II consisting of residues that were mobile on a slower timescale (mean $k_{ex} = 600 \pm 87$ s^{-1}).

Group I consisted of two Ile residues that line one side of the entrance to the template tunnel (Ile37, Ile96) and one (Ile641) on the C-terminal domain (**Figure 2.8a**). k_{ex} values for the C-terminal Ile617 could not be extracted accurately due to extremely low S/N ratio for the corresponding resonance. However, visual inspection of the field-dependence of $\Gamma_{MQ,eff}$

suggested a k_{ex} value characteristic of a group I residue. In RdRPs, such as P2, that initiate RNA synthesis *de novo*, the absence of a primer requires additional interactions of the template ssRNA with the RdRP to stabilize the initiation complex. These interactions in P2 occur with a "priming loop" located on the C-terminal domain and containing a conserved aromatic residue (Y630 in φ6 P2) (39). In addition, the accepted model for the exit of the newly synthesized daughter strands from the catalytic cavity involves motion of the C-terminal domain (39). It thus stands to reason that the conformational dynamics involving the C-terminal domain would be synchronized with those that allow the entry of ssRNA into the template tunnel. We therefore hypothesize that these faster motions are designed to facilitate entry of template and exit of the daughter chains in a coordinated fashion in line with the current accepted model for RNA entry and egress (39). However, the effects of mutations in the both the rim of the template tunnel and the C-terminal domain are difficult to interpret in terms of functional motions in a simple way. The rim mutant K30A has reduced transcriptional activity but this may partially be explained by its reduced affinity for RNA. Replacing the priming loop ^{630}YSW^{632} on the C-terminal domain by a smaller and more dynamic GSG sequence leads to enhanced transcription, though this increase may also be attributed to back-priming as opposed to *de novo* RNA synthesis (79,80). A similar effect is obtained using the mutant E634Q, which is expected to lead to a more dynamic C-terminal domain due to disruption in a salt-bridge with K144. Since this mutant has smaller (~20%) back-priming activity compared to the GSG mutant, enhanced mobility of the C-terminal domain does seem to play a role in transcriptional efficiency (80).

Group II consisted of Ile164, Ile258, Ile488, Ile500 and Ile563. Of these residues, Ile488 (motif D), Ile500 (motif E) were located in and around the catalytic motifs on the palm and Ile563 on the base of the thumb **(Figure 2.8a)**. Ile164 is located in the so-called plough region

(Figure 2.8a). It has been shown that plough mutants enhance transcriptional activity though they do not have any significant effect on template binding (80). This is consistent with the nature of motions displayed by Ile164 with the k_{ex} being smaller than those of the group I residues. Additionally, given that the motion of Ile164 occurs on a similar timescale as the catalytic site residues suggests some degree of coupling of the plough region to the active site and therefore to the transcriptional activity. Ile258 is located on the diametrically opposite end of the fingers (see **Figure 2.8a**) suggesting that the group II motions represent dynamic coupling of the outer edges of the fingers to the active site. Given that this motional timescale was slower than that seen for group I residues (putatively involved in RNA entry and egress), they could be expected to be rate limiting, if indeed they are catalytically relevant. A recent kinetic study on the poliovirus RdRP 3Dpol (47), suggested several steps involving conformational changes that are likely to play a role in catalysis. These included: (i) a conformational transition in the RdRP/NTP/RNA ternary complex to an activated state primed for catalysis (expected k_{ex} = $k_{forward}$ + $k_{reverse}$ ~800 s^{-1}), (ii) a transition from the activated complex to an activated elongation product complex with the product pyrophosphate, the central reaction in the nucleotide addition cycle (expected k_{ex}~895 s^{-1}), and (iii) a deactivation of the product complex (k_{ex} ~500 s^{-1}). These steps are followed by the rapid release of pyrophosphate product (k_{ex} ~8100 s^{-1}). The timescales associated with the first three steps are consistent with the second, slower timescale seen in the present study (group II). The small differences in the predicted k_{ex} values between these three steps would make it difficult to pinpoint the exact conformational transitions corresponding to the slower timescale. It is most likely that all of these transitions contribute to the distribution in k_{ex} values for the group II residues. However, the conjecture that these motions could be rate limiting in catalysis is reinforced by the detection of k_{ex} values in the 500-800 s^{-1} i.

e. group II range for residues on several catalytic motifs, most notably in the Motif C residue Ile449 in the P2/Mg^{2+}/GMPPNP complex (state **3**) and in the stable initiation complex (state **6**) discussed below.

For state **2** (P2 + Mg^{2+}), as expected, no major changes in dynamic timescales are seen within group I (1516 ± 168 s^{-1}) or group II (379 ± 157 s^{-1}) residues. However, Ile500 that was a group II residue for state **1** now showed dynamics characteristic of group I (see **Table 2.2**), accompanied by the largest $\Delta\delta_{1,2}$ [see Equation (2)] between states **1** and **2** (19 Hz). This is not unexpected since Ile500 is within 4 Å of the divalent metal binding sites. Comparatively large $\Delta\delta_{1,2}$ values were also seen for Ile164 (13 Hz), Ile449 (11 Hz) and Ile488 (14 Hz). All of these residues with the exception of Ile164 were located at the catalytic center reinforcing the notion of long-range coupling of the plough region (where Ile164 is located) with the catalytic site.

For state **3**, a similar timescale separation was seen for group I (1134 ± 203 s^{-1}) and group II (427 ± 164 s^{-1}) residues. However, Ile164 now displayed a higher k_{ex} value that was closer in magnitude to that for Ile500 (Table 2.2). It is possible that these residues comprised an additional intermediate timescale. Ile164 and Ile500 also displayed comparatively large $\Delta\delta_{2,3}$ values of 12 and 14 Hz respectively. Here again, the correlated variation in k_{ex} and $\Delta\delta$ values suggested that conformational changes of the catalytic core were coupled to that of the plough region.

In the case of state **4**, P2+Mg^{2+}+GMPPNP+5'-UUUCC-3', is the dead-end complex. All conformational dynamics involving Ile96 **(Figure 2.8b)**, Ile164, Ile258, Ile449 and Ile563 were completely suppressed. Motion could only be detected for Ile488 and Ile500, which have a similar timescale as Ile164 and Ile500 in state **3**. The dead-end complex showed a different dynamic nature with other states both in the distribution as well as the magnitudes of the k_{ex} values. It has been suggested that P2 in this dead-end or stalled state can be reinitiated by

supplementation with the missing NTPs (38). However, this reinitiated complex has a great reduced polymerization rate than that prior to stalling. Another unique observation in state **4** was that the Ile37 resonance showed extensive line broadening due to an unresolved doublet; therefore the dispersion curve could no longer be reliably fit (**Figure 2.9**). But based on $\Delta\Gamma_{MQ,eff}$ = $\Delta\Gamma_{MQ,eff}$ (31.25Hz) – $\Delta\Gamma_{MQ,eff}$(1000Hz) = $15s^{-1}$ (for the major component of the doublet), slow conformational dynamics persisted in this state for Ile37.

For the unstable initiation complex in state **5**, P2+Mg^{2+}+GMPPNP+5'-UUUUC-3', the overall distribution is similar to state **3**— group I (Ile37, Ile96, and Ile500, k_{ex} = 1141 ± $149s^{-1}$) and group II (Ile164, Ile258, Ile488 and Ile563, k_{ex} = 453 ± $91s^{-1}$). Ile164 also displayed a relatively large $\Delta\delta_{3,5}$ value (20 Hz) (**Figure 2.10**).

In addition to the **1-5** states discussed above, we also analyzed the other two states (**3A** and **6**) involving GMPCPP, which is a non-cleavable GTP analog between α and β phosphate. For state **3A** (Mg^{2+}/GMPCPP, Table **2.1**), the motional modes and k_{ex} values similar to state **3** were expected. However, substantial differences were observed (**Table 2.2**). This indicated that the nature of perturbations induced by the GTP-analogs GMPPNP and GMPCPP were different, a scenario confirmed by inspection of the $\Delta\delta_{3,3A}$ values. Motions involving Ile164, Ile258 and Ile563 were completely suppressed, adding credence to the hypothesis that the motions of these residues were coupled to the conformational fluctuations around the catalytic cavity at the substrate NTP binding sites. Given the fact that the affinities of P2 towards the two GTP analogs were about the same, as mentioned above, it is likely that these differences arose from subtle variations in the orientation of the triphosphate moiety at the active site possibly resulting in altered coupling of the active site to remote regions of the protein.

However, the motional characteristics of Ile37, Ile96 and Ile488 were similar to state **3**

(**Table 2.2**), though small differences in k_{ex} values for Ile500 with this state was noted. While the k_{ex} value for Ile96 was similar to state **3**, the $\Delta\Gamma_{MQ,eff} = \Gamma_{MQ,eff}(31.25\ Hz) - \Gamma_{MQ,eff}(1000\ Hz) = 3.8\ s^{-1}$ (compared with 13.2 s^{-1} for state **3**, **Figure 2.8b**) most likely due to alteration in $\Delta\omega_H$ or $\Delta\omega_C$ values or both rather than a simple population shift (with respect to state **3**) given that $\Delta\delta_{3,3A}$ value for this residue was quite small.

The complex of Mg^{2+}/GMPCPP/5'-UUUCC-3' (state **6**) was selected to study the stable initiation state. In state **6** (**Figure 2.4b**), field-dependent $\Gamma_{MQ,eff}$ could only be detected for Ile37 (characteristic of a group I residue), Ile449 (group II), Ile488 (group II) and Ile500, while no such effects were seen for Ile96, Ile164, Ile258 and Ile563. Thus, the only clear difference between the motional characteristics of the GMPCPP/Mg^{2+} (state **3A**) and the stable initiation complex i.e. the GMPCPP/Mg^{2+}/ssRNA (state **6**) was the suppression of slow dynamics involving Ile96 (template entry) in state **6**. This can be explained as the ssRNA is now likely locked in a stable complex. Unfortunately the weak signals for the C-terminal Ile617 and Ile641 residues did not allow the quantitative analysis of these resonances. Thus the nature of the motions assigned to facilitate the egress of the daughter strands could not be analyzed.

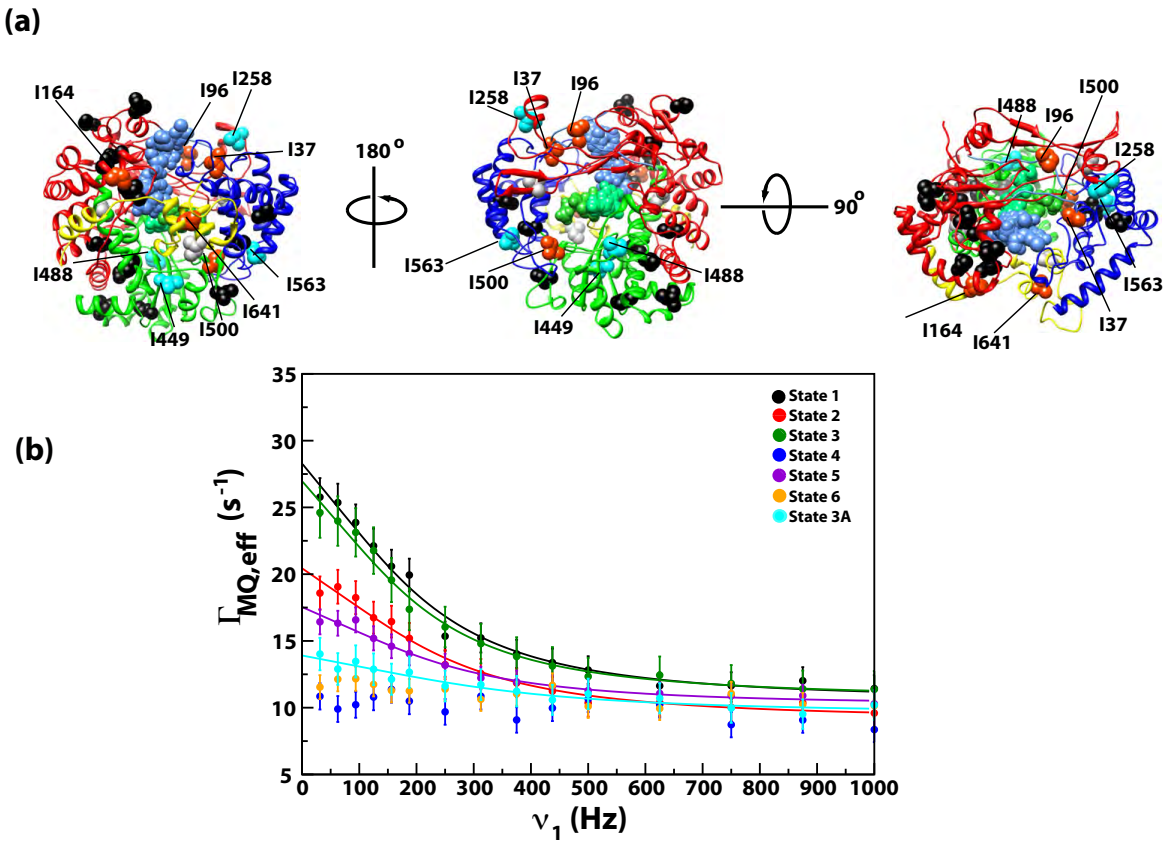


Figure 2.8. Multiple Quantum Dispersion Studies.
(a) Results of ^{13}C, ^{1}H multiple-quantum relaxation dispersion experiments. Ile δ1 positions that have detectable k_{ex} values are displayed on the tertiary structure of P2. Data for state **3** is shown as a representative example. Group I residues are shown in orange and group II residues are shown in cyan. Ile164 that has a k_{ex} value that is intermediate between group I and II for this state, is also shown in orange. Ile residues which had $\Gamma_{MQ,eff}$ values independent of the ν_1 field are depicted as black spheres. Data for Ile211, Ile283 and Ile617 were not analyzed in detail in any of the 7 states. These residues are indicated by light grey spheres. Two substrate molecules are shown in green and the template in blue. **(b)** Multiple quantum dispersion data for Ile96 for all seven states are shown. Experimental data are represented by the circles and theoretical curves resulting from fits to Equation (4) are shown by solid lines. Theoretical curves are not shown for states **4** and **6**, where $\Gamma_{MQ,eff}$ is independent of the applied RF-field strength (ν_1).

Table 2.2. Exchange Rates (k_{ex} in s^{-1}) Determined from Ile [$^{13}C\delta1,^{1}H\delta1$] Multiple Quantum Relaxation Dispersion

State	Ile37	Ile96	Ile164	Ile258	Ile449	Ile488	Ile500	Ile563	Ile641
1	1481 ± 118	1358 ± 109	600 ± 142	550 ± 104	ND	562 ± 179	690 ± 40	449 ± 123	1690 ± 301
2	1292 ± 86	1537 ± 142	210 ± 109	311 ± 74	[a]	577 ± 61	1719 ± 368	418 ± 241	[a]
3	1112 ± 88	1315 ± 92	855 ± 136	450 ± 54	414 ± 179	425 ± 51	1042 ± 139	417 ± 79	1344 ± 110
4	[b]	ND	ND	ND	ND	853 ± 118	1811 ± 185	ND	[a]
5	1253 ± 104	1549 ± 148	522 ± 226	336 ± 116	[a]	528 ± 49	1431 ± 96	424 ± 101	[a]
3A	1214 ± 207	2105 ± 430	ND	ND	[a]	679 ± 93	767 ± 35	ND	[a]
6	1811 ± 522	ND	ND	ND	451 ± 181	671 ± 145	972 ± 186	ND	[a]

[a] Analysis of these resonances was not performed due to poor signal-to-noise.
[b] Quantitative fits to obtain k_{ex} values could not be performed for Ile37 in state **4** due to extensive line broadening indicative of a departure from the fast exchange regime. $\Delta\Gamma_{MQ,eff} = \Gamma_{MQ,eff}$ (31.25 Hz) – $\Gamma_{MQ,eff}$ (1000 Hz) ~15 s^{-1}.
Residues for which dispersion profiles were flat i.e. those which had $\Gamma_{MQ,eff}$ values independent of the RF-field strength (ν_1) are indicated by "ND".

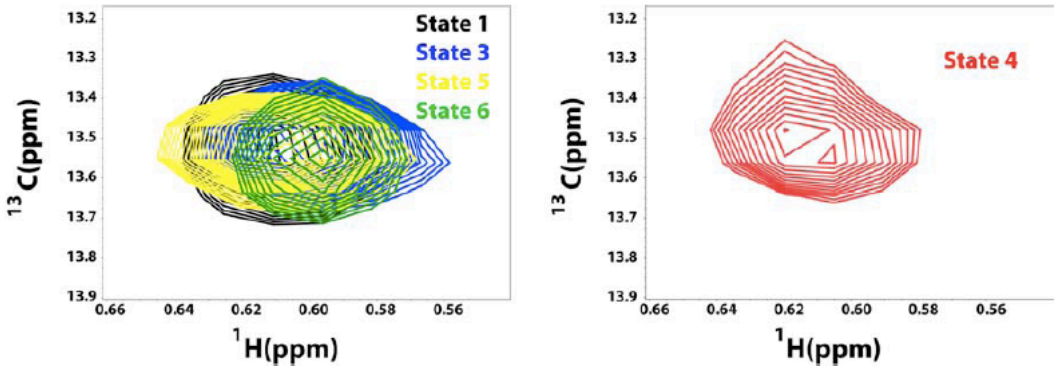


Figure 2.9. Lineshapes for the Ile37 Resonance in the Various States.
The peak corresponding to Ile37, in ^{13}C,^{1}H HMQC spectra of uniformly deuterated-Ile-Cδ1-^{13}C,^{1}H-labeled P2 at 600 MHz, displayed significant broadening for the stalled complex (state **4**) seemingly due to an unresolved doublet. No such characteristics were seen for the Ile37 resonance in any of the other states (including states **2** and **3A**, data for which has been omitted for clarity). MQ relaxation dispersion data for the Ile37 resonance corresponding to all states except **4** could be successfully fit to Equation (6) (fast exchange on the chemical shift timescale).

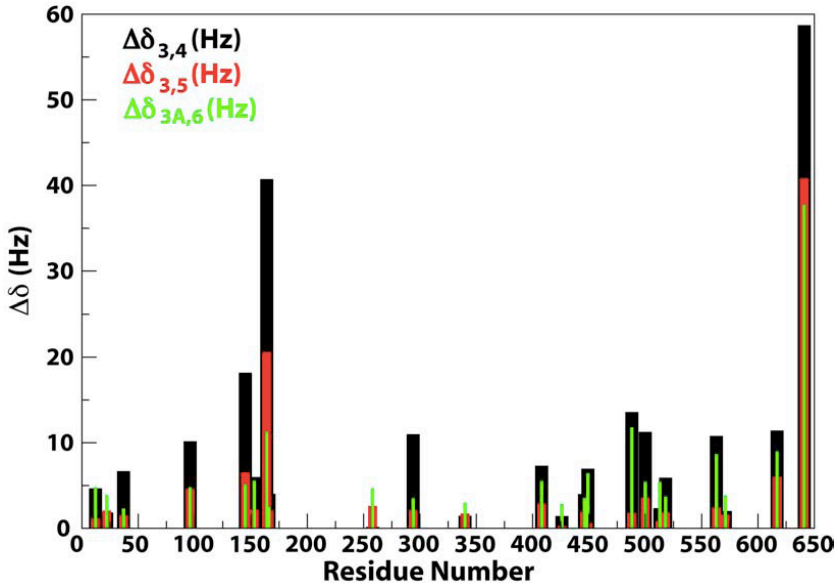


Figure 2.10. Chemical Shift Changes in the Ternary Complexes.
Scaled chemical shift differences of the Ile δ1 positions in P2 for ternary complexes with GTP analogs and 5-nt ssRNA templates, calculated using Equation (2), are plotted against residue number. Data are shown for the shift differences between states **3** (P2+Mg^{2+}+GMPPNP) and **4** (P2+Mg^{2+}+GMPPNP+5'-UUUCC-3') − $\Delta\delta_{3,4}$, **black**; between states **3** (P2+ Mg^{2+}+GMPPNP) and **5** (P2+ Mg^{2+}+GMPPNP+5'-UUUUC-3') − $\Delta\delta_{3,5}$, **red**; and between states **3A** (P2+ Mg^{2+}+GMPCPP) and **6** (P2+ Mg^{2+}+GMPCPP+5'-UUUCC-3') − $\Delta\delta_{3A,6}$, **green**. The ternary complex corresponding to the stalled state (state **4**) displayed the largest chemical shift differences with the state loaded with the corresponding GTP analog in the absence of RNA (Mg^{2+}/GMPPNP, state **3**). The following Ile residues are located near the template tunnel binding tunnel – Ile22 and Ile145 within 4 Å, Ile153 and Ile283 within 4-8 Å, and Ile294 within 8-10 Å (distances calculated for heavy atoms only).

2.4. Summary

Based on the characterization of the seven states discussed above, the following observations can be made — the slow dynamics in P2 could be generally characterized by two distinct timescales, a faster timescale (group I, k_{ex} ~1200-1500 s^{-1}) and a second, slower timescale (group II, k_{ex} ~500-800 s^{-1}).

The first, faster timescale (group I, k_{ex} ~1200-1500 s^{-1}) involved residues in close proximity to the template tunnel (Ile37 and Ile96) and on the C-terminal domain. These motions most likely facilitate the passage of the template through the tunnel and exit of the newly synthesized daughter strands by coordinated movements of the C-terminal domain. The nature of conformational dynamics in the C-terminal domain was also borne out by the large chemical shift changes in all the complexes with ssRNA – for Ile641, $\Delta\delta_{3,4}$ = 59 Hz, $\Delta\delta_{3,5}$ = 41 Hz and $\Delta\delta_{3A,6}$ = 38 Hz. Given the timescale of these motions, they are clearly not expected to be rate limiting. This is consistent with data obtained for both DNA (81) and RNA polymerases (47) where the translocation is not rate limiting in the process of polymerization. It has been suggested that the translocation step is coupled to pyrophosphate release from the enzyme. Rates of pyrophosphate release have been found to be as slow as 1200 s^{-1} in T7 DNA polymerase (81) and as fast as 8100 s^{-1} in 3Dpol (47). The k_{ex} values of group I residues were within this range of timescales. Formation of a stable initiation complex lead to suppression of the motion of the fingertip Ile96 (see **Figure 2.8b**) while that of the Ile37, that is closest to the ssRNA template, remained. This suggests that some motion around the template tunnel is required to thread the ssRNA through the RdRP even after the formation of an initiation complex. Ile96 was also no longer mobile in the stalled complex and the motion involving Ile37 occurred on a slower

exchange regime.

The second, slower motional mode (group II, k_{ex} ~500-800 s^{-1}) involved several residues located in the catalytic motifs (motif C – Ile449, motif D – Ile488 and motif E – Ile500) and in the outer edge of the template tunnel (Ile164, Ile258 and Ile563). The distribution of k_{ex} values within these groups of residues suggests that there may be multiple motional modes involved in producing them. This fact is supported by the observation of selective damping of motions at the outer edge of the template tunnel upon binding GMPCPP, a behavior distinct from GMPPNP binding. This also suggests that minor perturbations at the active site can cause activation or deactivation of a subset of these modes involving distant regions of P2. It has been proposed that one of the conformational changes that occur on the catalytic timescale likely involves the reordering of the triphosphate moiety and concurrent changes in the enzyme (47). Given the differences induced by GMPPNP and GMPCPP both in static chemical shifts (in Ile145, Ile164, Ile449 and Ile500) and dynamics changes (in Ile164, Ile258 and Ile563), it is possible that this reorientation contributes in part to one component of the slower timescale. Indeed the largest differences in the binding mode of these two GTP analogs most likely occur due to the orientation of the triphosphate moiety. An analysis of GTP-bound structures of P2 (e.g. 1H10) reveals that the O atom between the α– and β–phosphates (CH_2 in GMPCPP) participates in a hydrogen bond with the Arg270 sidechain, additional hydrogen-bonds are also seen for the β– and γ–phosphates (with basic residues – Arg268, Arg270 and His328) defining their orientation. The conserved motif F (268-272) that has evolved to co-ordinate the triphosphate moiety does not contain any Ile residues, so is invisible to the analysis presented here. Another key observation that emerged from these studies was evidence of long-range coupling of the plough region (as reflected by Ile164) with the catalytic core region reinforcing observation from

functional analyses (74).

Upon formation of a dead-end complex (state **4**), motions of Ile449 on motif C are completely suppressed while motion of the motif D residue, Ile488 occurs on a distinct, faster timescale. This dead-end complex also seems structurally distinct from all other RNA-bound states. Interestingly, Ile488 displayed k_{ex} ~600 s^{-1} in all states we analyzed, including the dead-end complex, where it had a slightly higher k_{ex} value compared with the other states. An analysis of the available structures of a range of RdRPs indicate that the conformation of motif D is amongst the most variable (52). It has been suggested that this motif D, that bears a basic residue functioning as a general acid, undergoes a conformational change in the that primes it to protonate the product pyrophosphate leading to a large enhancement in the rate of nucleotidyl transfer (52,56). There are several Lys residues (476, 479 and 487) on motif D in φ6 P2, which could serve this purpose.

The case of the motif E residue Ile500 was rather curious. It displayed the largest variation in k_{ex} values (from 690 to 1811 s^{-1}) over the several states, being slowest (and closest to the catalytic i.e. group II timescale) for states **1**, **3**, **3A** and **6**, and fastest (and closest to a group I timescale) for states **2**, **4** and **5**. The exact reason for this variation is unclear at the present time. Notably, Ile500 also showed significant chemical shift variations across the various states.

Another intriguing feature that was immediately evident was that several residues that were dynamic on the slow, μs–ms timescale (with significant k_{ex} values) across the states analyzed, were also dynamic on the fast, ps–ns timescale (with low η and consequently low S^2_{axis} values). These included Ile37, Ile96 and the catalytic site residues Ile488 and Ile500. As previously mentioned Ile449 at the top of motif C displayed enhanced fast dynamics in the Mg^{2+}/GMPPNP complex. It is likely that the motion on multiple timescales implies sampling of

conformational states as well as local conformational variations allowing the processing of a wide range of substrate NTPs and template RNAs.

Chapter 3

SLOW CONFORMATIONAL DYNAMICS IN THE CYSTOVIRAL RNA-DIRECTED RNA POLYMERASE φ6 P2: INFLUENCE OF SUBSTRATE NUCLEOTIDES AND TEMPLATE RNA

Most of the contents of this chapter have been published as:

Ren Z and Ghose R (**2011).** *Biochemistry* **50**, 1875-1884.

3.1. Introduction

As described in previous chapter, we characterized dynamics of φ6 P2 on the fast (ps-ns) and slow (μs-ms) timescales using the 25 Ile residues that are distributed over its structural domains and conserved sequence motifs as probes. We found that, on the fast timescale, the entrances to both the template and substrate entry portals were highly dynamic; the catalytic motifs were rigid, becoming more ordered in the presence of ssRNA. On the slow timescale, relevant for catalysis, two broad dynamic regimes were seen represented - a faster timescale (Group I: k_{ex}~1200-1500 s^{-1}) and a slower timescale (Group II: k_{ex}~500-800 s^{-1}) (**Table 3.1**). Motions on the faster timescale involving the template entry portal (Ile37 and Ile96) and the C-terminal domain (CTD, Ile641) seemed to be related to the entry of the template and exit of the daughter chain by motion of the CTD. Motions that involve RNA translocation have been shown not to be rate limiting in DNA (81) and RNA (47) polymerases. Interestingly, the timescale of the slower motions overlap with those suggested to being rate-limiting in kinetic analyses of homologous polymerases (47). These motions involve residues near the catalytic site (Ile449 on

motif C, Ile488 on motif D), at remote sites shown to be coupled to transcriptional efficiency (Ile164 on the "plough" region) (80), and additional residues located on the outer edge of the template tunnel (Ile258 on the fingers) and Ile563 (on the thumb). In addition, Ile500 that lies on motif E (palm) seemed to switch between timescales I and II depending on the nature of the complex.

Table 3.1. Residues That Exhibit Slow Microsecond to Millisecond Time Scale Dynamics[a]

Residue	Location	Slow motional time scale
Ile37	thumb, template entry portal	I
Ile96	fingertips, template entry portal	I
Ile164[b]	fingers, plough region	II
Ile258	fingers	II
Ile449[c]	palm, motif C	II
Ile488	palm, motif D	II
Ile500	palm, motif E	switches between I and II
Ile563	base of thumb	II
Ile641[d]	C-terminal domain	I

[a]These are the motional time scales seen for most of the states analyzed (82).
[b]The plough region is known to be coupled to transcriptional efficiency, i.e., catalysis (80).
[c]Motif C bears the conserved ^{452}S(G)DD^{454} motif.
[d]The motion of the C-terminal domain seems to assist in the exit of the daughter chain (39).

It has been shown that for φ6 P2, the efficiency of RNA polymerization *in vivo* depends on the sequence at the extreme 3'-ends of the RNA template especially with respect to the two penultimate bases (17). The 3'-ends of the minus strands of the small and medium genomic segments that contain a 5'…UUUCC-3' sequence are utilized more efficiently than the minus strand of the large segment or the 3'-ends of the genomic plus strands that contain 5'…UUUAC-3' and 5'…UCUCU-3' sequences respectively. This difference in processivity could be simply related to the affinity with which P2 recognizes the sequence at the 3'-end of the template. On the other hand it could also be the result of a more complex interplay between structural or conformational changes resulting from 3'-end recognition.

Therefore, in this chapter we mainly focus on investigating the influence of the substrate NTPs and extreme 3'-end sequence of ssRNA templates on the motional modes at specific catalytic steps by using multiple-quantum NMR spectroscopy at the 25 Ile δ1 positions in P2.

3.2 Material and Methods

Protein expression, purification and sample preparation, fluorescence anisotropy measurements, and multiple-quantum dispersion measurements were performed as previously described in **Section 2.2.**

3.3 Results and Discussion

3.3.1. Influence of Template 3'-Ends

We first tested whether this difference in template utilization could be attributed to the affinity with which P2 recognizes the 3′-end sequence. We measured the affinities towards P2 of five 5-nt templates (C1-C5, see **Table 3.2**) with varying 3′-end sequences, by monitoring the change in fluorescence anisotropy of these constructs bearing a 5′-end with 6-carboxyfluorescein in the presence of increasing amounts of protein. The affinities of two additional 7-nt templates were also measured (**Table 3.2**). The K_d values (**Table 3.2**) thus obtained varied from 1.2 ± 0.1 μM (C5) to 295.9 ± 15.2 μM (C2) with those corresponding to the 3′-segments of the genomic minus-strands (C1, C6, s^- and m^-; C2, l^-) binding with the lowest affinity (**Figure 3.1**). Based on our analysis, it appears that P2 prefers U over C at the 3′-ends for stronger binding. The greatly reduced affinity for the sequence corresponding to the 3′-end of l^- (C2) may be explained by the presence of a larger purine base (A) as previously suggested (39). Interestingly, sequences containing a UU (C5) or a CU (as in the genomic s^+, m^+ and l^+; C3 and C7 at their 3′-ends that have a substantial higher affinity towards P2 but are processed less efficiently compared to the CC sequence (as in s^- and m^-; C1 and C6). Thus, the affinity follows the trend (strongest to weakest binding) UU > UC > CU > CC > AC, while the template efficiency is CC > UC > AC > CU (83). It is also known that templates bearing a UU at the 3′-end are less efficiently utilized than those terminated with a UC sequence (17). Thus, template utilization and by inference formation of a stable functional initiation complex cannot be related in a simple way to binding affinity and is likely a more complex function of affinity and changes in structure and dynamics of P2 resulting from template and substrate binding.

Table 3.2. Apparent Binding Affinity of φ6 P2 Towards single-stranded RNA Sequences

Template	Sequence	K_d (μM)
C1	**5′-UUUCC-3′[b]**	**53.2 ± 3.5[a]**
C2	**5′-UUUAC-3′[c]**	**295.9 ± 15.2**
C3	5′-UUUCU-3′[d]	15.3 ± 1.4
C4	5′-UUUUC-3′	6.6 ± 0.4[a]
C5	**5′-UUUUU-3′**	**1.2 ± 0.1**
C6	5′-UUUUUCC-3′[b]	129.9 ± 4.4
C7	5′-UUUUUCU-3′[d]	15.0 ± 0.7

[a]Relaxation data were collected for P2 in the presence of templates that are indicated in boldface. [b]The 3'-end of the plus strands of the minus strands of the M(m^-) and S (s^-) genomic segments. [c]The 3'-end of the minus strands of the L (l^-) genomic segment. [d]The 3'-end of the plus strands of the L (l^+), M (m^+) and S (s^+) genomic segments. [e]In agreement with that reported by Ren et al. (82).

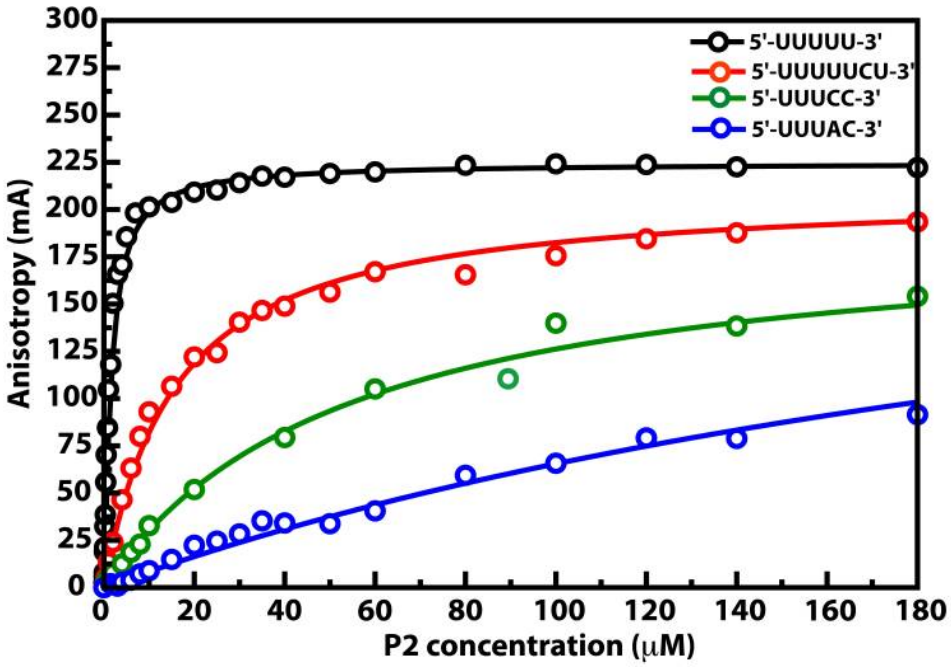


Figure 3.1. Representative Binding Curves Demonstrating the Affinities of Several ssRNA Constructs Towards φ6 P2.

Determined by fluorescence anisotropy measurements using 5'-carboxyfluorescein tagged ssRNA. Only a few binding curves are shown for ease in illustrating their differences.

3.3.2. Role of the Sequence at the 3'-end

Next, we tested whether the nature of ssRNA sequence plays a role in modifying the dynamics of P2 on the slow μs-ms timescale by measuring the RF-field (ν_1) dependence of the Ile-Cδ1 ^{13}C–^{1}H MQ relaxation rates. We collected data for three constructs C1, C2 and C5 that span the range of K_d values (**Table 3.2**) with the first representing the genomic s^- and m^- 3′-ends and the second the 3′-end of the genomic l^- segment, the two weakest binders and the third (C5) being the strongest binder to P2. Representative examples of experimental dispersion curves and the corresponding fits to Equation (4) section (details in **Section 2.2**) are depicted in **Figure 3.2**. Overlay of ^{13}C, ^{1}H HMQC spectra of P2 in the presence of the three constructs is shown in **Figure 3.3a** to illustrate the spectral quality.

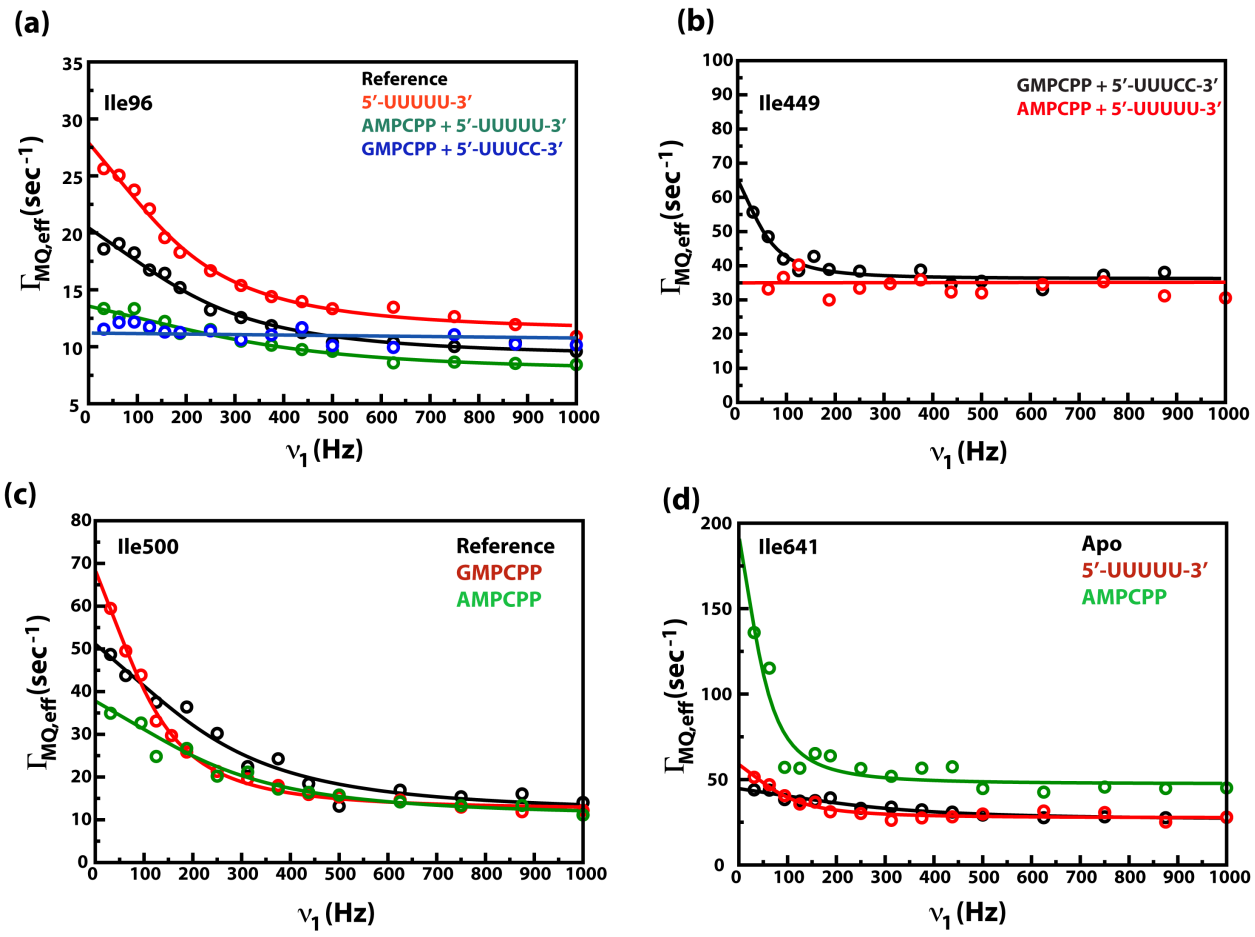


Figure 3.2. Multiple-quantum Dispersion Measurements.
Representative dispersion curves with experimental data depicted by circles and fits to Equation (4) by solid lines. **(a)** Data for the fingertips residue Ile96. **(b)** Data for the motif C residue Ile449. **(c)** Data for the motif E residue Ile500. **(d)** Data for Ile641 that lies in the C-terminal domain. The blue line in **(a)** and the red line in **(b)** are drawn to guide the eye and do not represent fits. These correspond to situations where the dispersion curves are flat i.e. $\Gamma_{MQ,eff}$ values are independent of ν_1. Data for only a small number of states are shown for visual convenience.

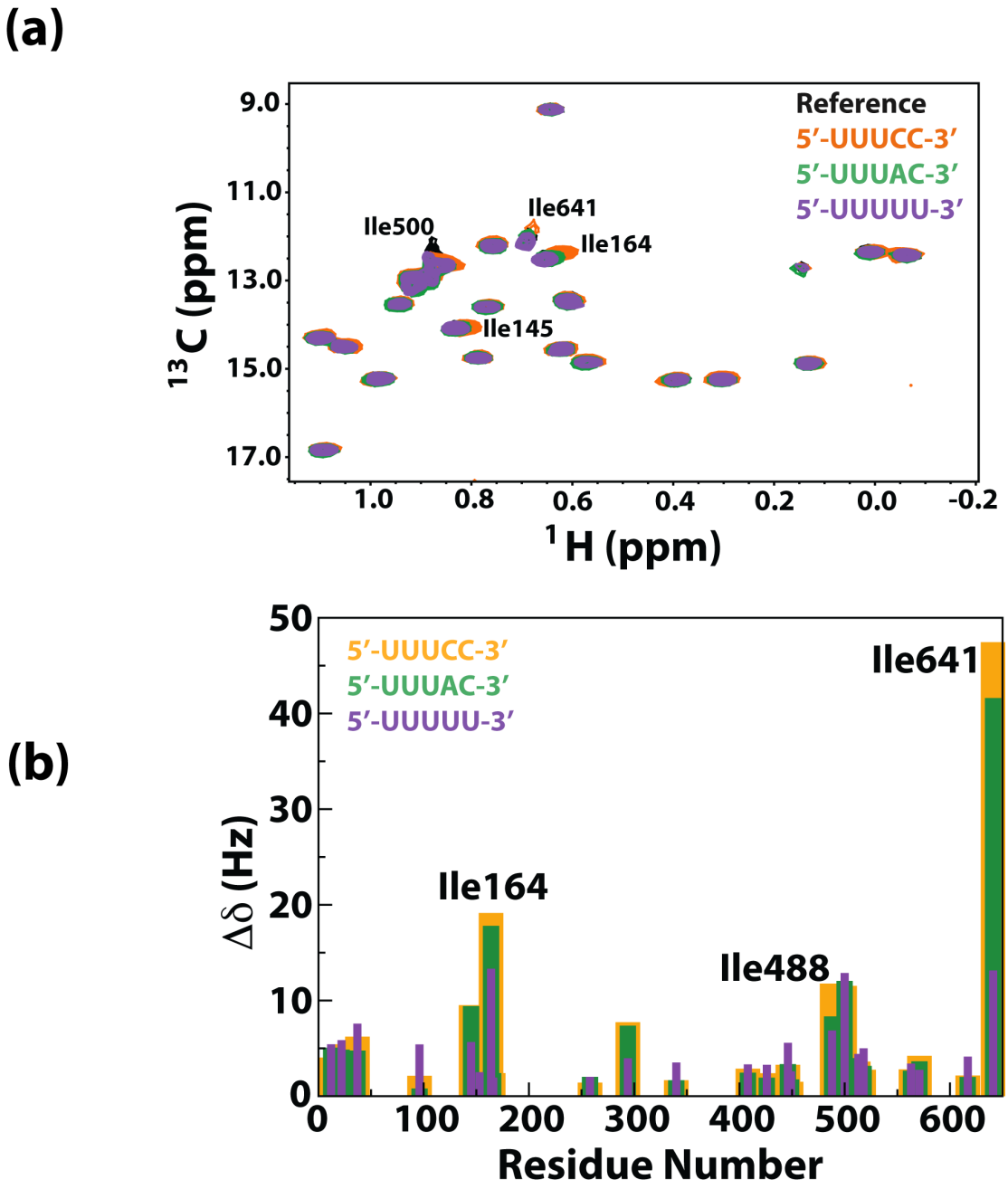


Figure 3.3. Chemical Shift Changes in the Presence of Different ssRNA Constructs.
(a) Overlay of the reference spectra (φ6 P2 in the NMR buffer, black) with each of three templates C1 (5'-UUUCC-3', 3'-ends of the genomic s^- and m^- strands, orange), C2 (5'-UUUAC-3', 3'-end of the genomic l^-, green) and C5 (5'-UUUUU-3', purple). **(b)** Scaled chemical shift changes (with respect to the reference) shown for each of the three templates (reference – black, C1 – yellow, C2 – green and C5 – purple).

The common feature of the influence on the slow dynamics of P2 for the three ssRNAs tested (**Table 3.3**) was the complete suppression of slow motion in the Group II residue Ile563 that lies at the base of the thumb domain. However, the correct genomic 3′-ends (C1 and C2) also resulted in a complete suppression of slow dynamics in the Group II Ile258 (on the fingers domain) while this motion persisted for C5. Ile641 that has been shown to be dynamic on the Group I timescale (quantitative data could not be obtained for this residue in the reference state, a k_{ex} value of 1358 ± 109 s^{-1} was found in the Apo state (82) seemed to display dynamics on a much slower timescale (k_{ex}= 580 ± 112 s^{-1}, see **Figure 3.2d**) in the presence of the non-physiological template C5, suggesting a decoupling of the motions of the CTD and the edge of the template entry portal and the fingertips. In addition to these perturbations in dynamics, significant differences in chemical shift perturbations (**Figure 3.3b**) induced by the genomic constructs C1, C2 and C5, were seen. Notably, Ile641 displayed comparable shift changes for C1 (47 Hz) and C2 (42 Hz), while the changes for C5 were much more modest (13 Hz). A marginally smaller chemical shift change was also seen in Ile164 for C5 (13 Hz) compared with C1 (19 Hz) and C2 (18 Hz). This suggests that the proper genomic 3′-ends elicit specific and unique changes in structure and dynamics compared with the non-physiological (C5) template in spite of having a 44-fold (C1) and 247-fold (C2) less affinity for P2. No significant differences in the dynamic timescales or chemical shifts were noted for any of the residues that lie near the catalytic site - Ile449 (motif C), Ile488 (motif D) and Ile500 (motif E).

Table 3.3. k_{ex} Values (S^{-1})of P2 in the Presence of Different single-stranded RNA Templates

State	Ile37	Ile96	Ile164	Ile258	Ile488	Ile500	Ile563	Ile641
Reference[b]	1292 ± 86	1537 ± 142	210 ± 109	311 ± 74	577 ± 61	1613 ± 232[c]	418 ± 241	X
C1	1535 ± 382	1132 ± 257	X	ND[a]	491 ± 176	1450 ± 274	ND[a]	X
C2	1140 ± 265	1007 ± 104	X	ND[a]	547 ± 157	1193 ± 284	ND[a]	X
C5	1145 ± 98	1287 ± 95	306 ± 116	734 ± 192	403 ± 174	1486 ± 248	ND[a]	580 ± 112

[a]ND implies that the relaxation rate ($\Gamma_{MQ,eff}$) was independent of the applied ν_1 field. [b]φ6 P2 in the NMR buffer containing 5 mM Mg^{2+} was taken as the reference state. Data for the reference state is from **Table 2.2**. [c]Data for Ile500 was refitted excluding two points with large errors in **Table 2.2**. The k_{ex} value obtained was unaltered within confidence bounds. X implies that the quality of the data did not allow quantitative k_{ex} values to be extracted. Data for Ile449 is not shown here and Table 3.4 since the states characterized either do not show field dependent relaxation rates (i.e. ND) or the quality of the data did not allow quantitative estimation of the k_{ex} values (i.e. X).

3.3.3. Influence of Substrate Purine Nucleotide Triphosphates

Next we tested the influence of the substrate NTPs on slow dynamics in P2 (in the absence of RNA). The purine nucleotides GTP and ATP, are required initially in φ6 given the pyrimidine-rich 3′-end of the genomic segments. Indeed, GTP is the preferred initiation nucleotide for φ6 and a host of other viral RNA polymerases that initiate RNA synthesis *de novo*, including those from bovine diarrhea virus (BVDV), HCV, broom mosaic virus (BMV) among others (38). This is especially important since the exact order of substrate and RNA binding to P2 is not precisely known and pre-binding of the initiation nucleotide (GTP) cannot be ruled out, as suggested by Choi et. al. (84). For the present studies, we utilized AMPCPP and GMPCPP, the α–β non-hydrolysable analogs of ATP and GTP respectively, to allow complex

formation but not catalysis in the ternary complexes discussed below. The overall differences in chemical shift changes induced were quite modest (**Figure 3.4b**), though some of the largest changes were localized near the active site. Ile449 (motif C) and Ile500 (motif E) showed larger chemical shift changes for GMPCPP – 13 and 16 Hz respectively compared to AMPCPP – 5 and 10 Hz respectively, indicating larger conformational changes around the active site for the GTP-analog. In contrast, a much larger chemical shift change for Ile641 (CTD) was noted for AMPCPP (36 Hz) compared with GMPCPP (20 Hz).

Differences in dynamic timescales were more pronounced (summarized in **Table 3.4**). In the presence of GMPCPP, the motion of Ile500 on motif E changed from a faster timescale (Group I, see **Table 3.1**) in the reference state to a catalytic timescale (Group II) (**Figure 3.2c**). AMPCPP induced no such change. Additionally, for AMPCPP (as in the case of template C5 discussed above), the dynamics of Ile641 (CTD) seemed to shift towards a slower timescale (k_{ex} = 336 ± 84 s^{-1}) (**Figure 3.2d**) around the Group II range rather than Group I type dynamics characteristic of this region (**Table 3.1**) (82). While the quality of data for this residue in the presence of GMPCPP did not allow us to extract a quantitative k_{ex} value, a k_{ex} value of 1344 ± 110 s^{-1} (82) was found for this residue in the presence of GMPPNP, indicative of Group I dynamics. Here too, as in C5 (above), the incorrect ligand leads to a decoupling of the motion of the template entry portal with the CTD.

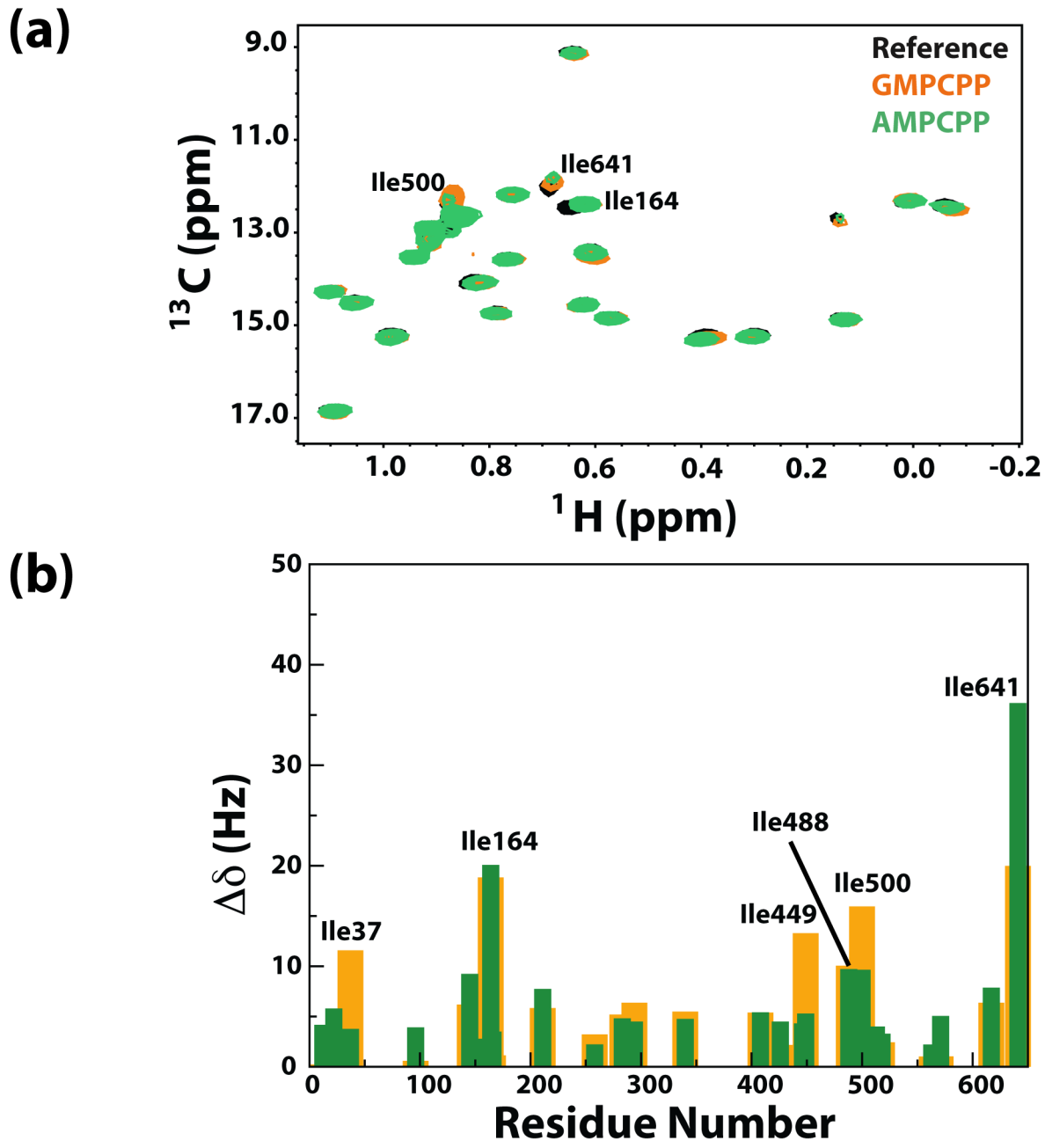


Figure 3.4. Chemical Shift Changes in the Presence of Substrate Analogs.
(a) Overlay of the reference spectra (black) with each of the two non-hydrolysable substrate analogs – GMPCPP (orange) and AMPCPP (green). **(b)** Scaled chemical shift changes (with respect to the reference) shown for the substrate analogs (GMPCPP – orange, AMPCPP – green).

Table 3.4. k_{ex} Values of P2 in the Presence of Substrate NTPs

State	Ile37	Ile96	Ile164	Ile258	Ile488	Ile500	Ile563	Ile641
Reference[a]	1292 ± 86	1537 ± 142	210 ± 109	311 ± 74	577 ± 61	1613 ± 232[b]	418 ± 241	X
AMPCPP	1405 ± 140	3010 ± 679	ND	ND	343 ± 107	1634 ± 166	ND	336 ± 84
GMPCPP[a]	1214 ± 207	2105 ± 430	ND	ND	679 ± 93	767 ± 35	ND	X

[a]Reference and GMPCPP data from **Table 2.2**
[b]Data for Ile500 in the reference state were refitted excluding two points with large errors. The k_{ex} value obtained was unaltered within confidence bounds from that obtained previously.

3.3.4. Ternary Complexes

Finally, we analyzed two ternary complexes that should form stable initiation platforms on the basis of proper complementarities between template 3′-end sequence and substrate NTPs. These were TC_A - P2−GMPCPP−C1, for which data are available from **Table 2.2** (**Figure 3.5a**) and TC_B - P2−AMPCPP−C5 (**Figure 3.5b**). Significant differences in dynamics were noted between TC_A representing the initiation platform formed during the transcription of the genomic s^- and m^- strands and TC_B that is nonphysiological in the context of φ6. The most striking difference was the complete suppression of the slow dynamics of Ile96 (fingertips, at the template entry portal) in TC_A (**Figure 3.5a** and **Table 3.5**).

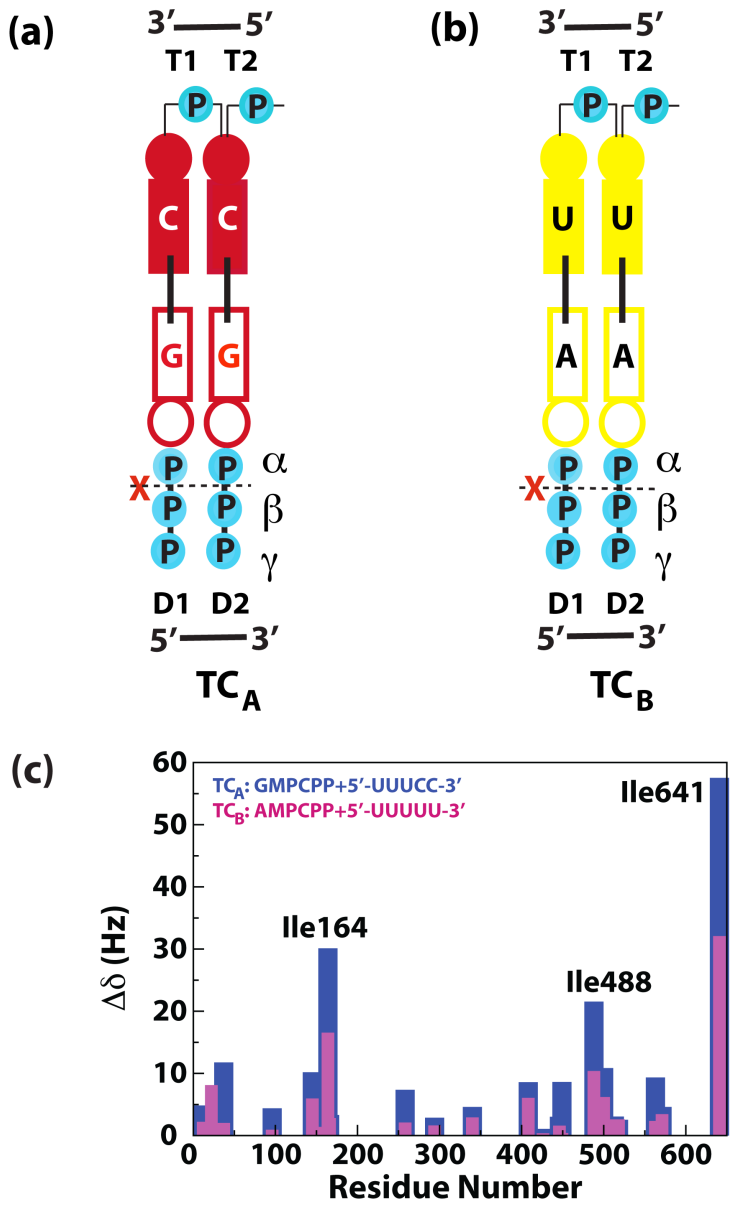


Figure 3.5. Ternary Complexes.
(a) Ternary complex TC_A forms between 5′-UUU**CC**-3′ (template C1), the non-hydrolysable GTP-analog GMPCPP, and φ6 P2. The template bases T1 and T2 (numbered 3′ to 5′) base pair with the substrate bases D1 and D2 (numbered 5′ to 3′) forming a stable initiation complex (based on base-pairing considerations alone). The first phosphodiester bond forms between D1 and D2 creating the daughter chain and releasing pyrophosphate. In TC_A, D2 (and D1) is the substrate GMPCPP that cannot be hydrolyzed between the α– and β–phosphates and the reaction cannot proceed, though the initiation complex forms. TC_A is physiologically relevant since template C1 corresponds to the 3′-ends of the s^- and m^- genomic segments. **(b)** TC_B the ternary complex between template C5 (5′-UUU**UU**-3′) and the non-hydrolysable ATP-analog AMPCPP, also allows proper base-pairing between T1:D1 and T2:D2, as in TC_A above, but C5 does not correspond to 3′-ends of any of the 6 genomic segments. **(c)** Scaled chemical shift changes (with respect to the reference state) shown for the ternary complexes (TC_A – blue, TC_B – pink).

Near the catalytic site, the motion of Ile500 in TC_A seemed to be shifted slightly towards the catalytic timescale (k_{ex} = 972±186 s^{-1}, **Table 3.5**) from a faster timescale. However, the most significant difference between the two complexes was the detection of slow motion on a catalytic timescale for Ile449 (k_{ex} =381 ± 95 s^{-1}, **Table 3.5**, **Figure 3.5b**) that lies on motif C (bearing Asp453 and Asp454 from the conserved DD motif) in TC_A but not in TC_B. These differences in dynamics between the physiological (TC_A) and non-physiological (TC_B) complexes were also reflected in the chemical shifts. The chemical shift changes seen in TC_A were also significantly larger than those seen in TC_B (**Figure 3.5c**) with the largest differences being at the plough (Ile164 – TC_A, 30 Hz; TC_B, 17 HZ), near the catalytic site (Ile488 – TC_A, 22 Hz; TC_B, 10 Hz) and at the CTD (Ile641 – TC_A, 58 Hz; TC_B, 32 Hz). This suggests that TC_A, the correct initiation complex was both structurally and dynamically distinct from the non-physiological TC_B.

Table 3.5. k_{ex} Values of P2 in Two Stable Initiation Complexes

State	Ile37	Ile96	Ile164	Ile258	Ile449	Ile488	Ile500	Ile563
TC_A	1811 ± 522	ND	ND	ND	381 ± 95	671 ± 145	972 ± 186	ND
TC_B	1166 ± 130	2273 ± 372	ND	ND	ND	X	1411 ± 326	ND

TC_A is a ternary complex between P2, GMPCPP and template C1 (5'-UUUCC-3'). TC_B is a ternary complex between P2, AMPCPP and template C5 (5'-UUUUU-3'). Data for TC_A was taken from **Table 2.2**. Data for Ile449 was refitted excluding two points with large errors. The k_{ex} value obtained was unchanged within the error bound. Data for Ile641 is not shown for the ternary complexes since the quality of the data did not allow quantitative estimation of the k_{ex} values (i.e. X).

3.4. Summary

From the present analysis it appears that different RNA templates affect dynamics on the slow timescale in φ6 P2 in a fashion that is not directly correlated to their relative affinities. RNA templates corresponding to the proper genomic 3'-ends have similar effects on structure and dynamics that are distinct from higher-affinity non-genomic RNA constructs. The effects of the substrate analogs AMPCPP and GMPCPP on the dynamics are also distinct, with the most important difference being the alteration of the dynamic timescale of Ile500, located near the catalytic site, from a faster to a slower one congruent with the catalytic timescale in the latter case. Additionally, in the presence of AMPCPP the motion of Ile641 moves to a slower timescale in essence decoupling the motion of the CTD from that of the template portal likely disrupting the coordinated motion of the template into and the daughter chain out of the catalytic pore (39,82). In the non-physiological ternary complex TC_B, residues around the catalytic site also do not show dynamics on the catalytic timescale as in the proper initiation complex, TC_A. Additionally in TC_A, motions of the catalytic motif C (reported by Ile449) on the catalytic timescale get activated. Thus, recognition of the correct genomic 3'-ends and the proper initiation NTP substrate, GTP, activates distinct changes in structure and dynamics in φ6 P2 to initiate RNA synthesis.

Chapter 4

CONCLUSION

Recent studies provide more and more evidence that dynamics seems to be the missing link between structure and function in polymerases (52,85). Our study that used 25 Ile as probes of molecular mobility represents the first step towards understanding these motions in a more quantitative way. We found that φ6 P2 was mobile on both the fast (ps-ns) as well as the slow (μs-ms) timescales (82). The timescale of the lower frequency motional modes that involve several conserved polymerase functional motifs was found to coincide with that relevant for catalysis (1-2 ms), determined from kinetic analyses of representative RdRPs (51). These studies strengthen the hypothesized role of dynamics in RdRP function. As RdRPs in general, and P2 in particular, function as intricate molecular machines with many moving parts working in unison.

We further investigated the influence of the extreme 3'-end sequence of the single-stranded RNA templates and the nature of the substrate nucleotide triphosphates on the slow motional modes using multiple-quantum relaxation dispersion. We found that P2, in the presence of templates bearing the proper genomic 3'-ends or the preferred initiation nucleotide (GTP), displays unique dynamic signatures that are different from those in the presence of non-physiological templates or substrates (86). This further suggests that dynamics may play a role in the fidelity of recognition of the correct substrates and template sequences to initiate RNA

polymerization.

While a definitive picture of the influence of motion awaits the availability of additional probe sites, the 25 Ile sites provide significant insight into the finely tuned molecular machine that is φ6 P2. The major contributors to the motions on the nanosecond timescale were conserved structural regions required for function. Motions involved in identical function were correlated regardless of location in the structure (82,86). We conclude that RdRp dynamics may be governed by long-range networks of interacting residues, which might have evolved for functioning (87), including polymerase fidelity.

A drawback of our approach and indeed that of others is that we interpret field (ν_1)-independent $\Gamma_{MQ,eff}$ values as lack of μs-ms timescale motion. RF-field- Analysis of the chemical shifts and populations of the major (ground) and minor (excited) states should also be highly informative (88-91), independent $\Gamma_{MQ,eff}$ may also occur if the differences in chemical shifts between the ground and excited states for ^{1}H or ^{13}C nuclei are vanishingly small. Thus, it is often useful to combine MQ dispersion measurements with corresponding SQ experiments. This, while easier to carry out in smaller and well-behaved proteins, is a non-trivial task in P2. Further, a definitive picture of the influence of motion in residue–by–residue detail awaits a complete or near-complete backbone assignment. This appears difficult in φ6 P2.

Therefore our laboratory is extending these studies to other methyl-bearing sidechains including Val, Leu, Met residues and the backbone ^{15}N nuclei in φ12 P2, which is a distinct relative to φ6 P2 and recognizing different initiation genomic strands, though it is better behaved and more amenable to solution NMR studies (8). We already solved φ12 P2 apo crystal structure at 1.7Å resolution, while obtaining co-crystallized φ12 P2 with an oligonucleotide ternary crystal structure remaining our goal. On one hand, φ12 P2 crystallographic structure studies will ease

our way for selective methyl-bearing sidechains labeling NMR structural assignment and eventually for backbone assignment. On the other hand, by comparing the similarities and differences between these two related cystoviral RdRPs, especially focus on structures and dynamics changes along their catalytic cycles, it will provide further insights on initiation reorganization and fidelity control. While our studies to provide a more informative picture of structure/dynamics/function correlations in RdRPs, are far from complete, to the best of our knowledge this work is the very first attempt to quantify dynamics along the catalytic cycle of viral RdRP, which is directly linking to fidelity control.

References:

1. Makeyev, E.V. and Grimes, J.M. (2004) RNA-dependent RNA polymerases of dsRNA bacteriophages. *Virus Res*, **101**, 45-55.

2. Evan Simpson1*, S.W., Josefina Bonilla2, Kateryna Gamazina3, Laura Cooley1 and Jennifer L Winkler1. (2007) Use of formative research in developing a knowledge translation approach to rotavirus vaccine introduction in developing countries. *BMC Public Health*, **7**, 281-292.

3. Maclachlan, N.J., Drew, C.P., Darpel, K.E. and Worwa, G. (2009) The pathology and pathogenesis of bluetongue. *J Comp Pathol*, **141**, 1-16.

4. Nakagawa, A., Miyazaki, N., Taka, J., Naitow, H., Ogawa, A., Fujimoto, Z., Mizuno, H., Higashi, T., Watanabe, Y., Omura, T. *et al.* (2003) The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure*, **11**, 1227-1238.

5. Poranen, M.M. and Tuma, R. (2004) Self-assembly of double-stranded RNA bacteriophages. *Virus Res*, **101**, 93-100.

6. Boerner, J.E., Lyle, J.M., Daijogo, S., Semler, B.L., Schultz, S.C., Kirkegaard, K. and Richards, O.C. (2005) Allosteric effects of ligands and mutations on poliovirus RNA-dependent RNA polymerase. *J Virol*, **79**, 7803-7811.

7. 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.

8. Gottlieb, P., Potgieter, C., Wei, H. and Toporovsky, I. (2002) Characterization of phi12, a bacteriophage related to phi6: nucleotide sequence of the large double-stranded RNA. *Virology*, **295**, 266-271.

9. Qiao, X., Qiao, J., Onodera, S. and Mindich, L. (2000) Characterization of phi 13, a bacteriophage related to phi 6 and containing three dsRNA genomic segments. *Virology*, **275**, 218-224.

10. Huiskonen, J.T., de Haas, F., Bubeck, D., Bamford, D.H., Fuller, S.D. and Butcher, S.J. (2006) Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure*, **14**, 1039-1048.

11. Jaalinoja, H.T., Huiskonen, J.T. and Butcher, S.J. (2007) Electron cryomicroscopy comparison of the architectures of the enveloped bacteriophages phi6 and phi8. *Structure*, **15**, 157-167.

12. Frilander, M., Gottlieb, P., Strassman, J., Bamford, D.H. and Mindich, L. (1992) Dependence of minus-strand synthesis on complete genomic packaging in the double-stranded RNA bacteriophage phi 6. *J Virol*, **66**, 5013-5017.

13. Mindich, L. (2012) Packaging in dsRNA viruses. *Adv Exp Med Biol*, **726**, 601-608.

14. Mindich, L. (1999) Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage phi6. *Microbiology and molecular biology reviews : MMBR*, **63**, 149-160.

15. Qiao, X., Qiao, J. and Mindich, L. (2003) Analysis of specific binding involved in genomic packaging of the double-stranded-RNA bacteriophage phi6. *J Bacteriol*, **185**, 6409-6414.

16. Makeyev, E.V. and Bamford, D.H. (2000) The polymerase subunit of a dsRNA virus plays a central role in the regulation of viral RNA metabolism. *The EMBO journal*, **19**, 6275-6284.

17. Onodera, S., Qiao, X., Qiao, J. and Mindich, L. (1998) Directed changes in the number of double-stranded RNA genomic segments in bacteriophage phi6. *Proc Natl Acad Sci U S A*, **95**, 3920-3924.

18. Qiao, J., Qiao, X., Sun, Y. and Mindich, L. (2003) Isolation and analysis of mutants of double-stranded-RNA bacteriophage phi6 with altered packaging specificity. *J Bacteriol*, **185**, 4572-4577.

19. 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.

20. Mindich, L. (2004) Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives. *Virus Res*, **101**, 83-92.

21. Sun, S., Rao, V.B. and Rossmann, M.G. (2010) Genome packaging in viruses. *Curr Opin Struct Biol*, **20**, 114-120.

22. Eryilmaz, E., Benach, J., Su, M., Seetharaman, J., Dutta, K., Wei, H., Gottlieb, P., Hunt, J.F. and Ghose, R. (2008) Structure and dynamics of the P7 protein from the bacteriophage phi 12. *J Mol Biol*, **382**, 402-422.

23. Sen, A., Heymann, J.B., Cheng, N., Qiao, J., Mindich, L. and Steven, A.C. (2008) Initial location of the RNA-dependent RNA polymerase in the bacteriophage Phi6 procapsid determined by cryo-electron microscopy. *J Biol Chem*, **283**, 12227-12231.

24. Poranen, M.M., Salgado, P.S., Koivunen, M.R., Wright, S., Bamford, D.H., Stuart, D.I. and Grimes, J.M. (2008) Structural explanation for the role of Mn2+ in the activity of phi6 RNA-dependent RNA polymerase. *Nucleic Acids Res*, **36**, 6633-6644.

25. Poranen, M.M., Butcher, S.J., Simonov, V.M., Laurinmaki, P. and Bamford, D.H. (2008) Roles of the minor capsid protein P7 in the assembly and replication of double-stranded RNA bacteriophage phi6. *J Mol Biol*, **383**, 529-538.

26. Kainov, D.E., Mancini, E.J., Telenius, J., Lisal, J., Grimes, J.M., Bamford, D.H., Stuart, D.I. and Tuma, R. (2008) Structural basis of mechanochemical coupling in a hexameric molecular motor. *J Biol Chem*, **283**, 3607-3617.

27. Huiskonen, J.T., Jaalinoja, H.T., Briggs, J.A., Fuller, S.D. and Butcher, S.J. (2007) Structure of a hexameric RNA packaging motor in a viral polymerase complex. *J Struct Biol*, **158**, 156-164.

28. Lisal, J., Lam, T.T., Kainov, D.E., Emmett, M.R., Marshall, A.G. and Tuma, R. (2005) Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states. *Nat Struct Mol Biol*, **12**, 460-466.

29. Kainov, D.E., Lisal, J., Bamford, D. H., and Tuma, R. . (2004) Packaging motor from double-stranded RNA bacteriophage phi12 acts as an obligatory passive conduit during transcription. . *Nucleis Acids Res.*, **32**, 3515-3521.

30. Eryilmaz, E., Benach, J., Su, M., Seetharaman, J., Dutta, K., Wei, H., Gottlieb, P., Hunt, J.F. and Ghose, R. (2008) Structure and dynamics of the P7 protein from the bacteriophage phi 12. *J Mol Bio*, **382**, 402-422.

31. Choi, K.H. (2012) Viral polymerases. *Adv Exp Med Biol*, **726**, 267-304.

32. Ferrer-Orta, C., Agudo, R., Domingo, E. and Verdaguer, N. (2009) Structural insights into replication initiation and elongation processes by the FMDV RNA-dependent RNA polymerase. *Curr Opin Struct Biol*, **19**, 752-758.

33. Sonntag, K.C. and Darai, G. (1995) Evolution of viral DNA-dependent RNA polymerases. *Virus Genes*, **11**, 271-284.

34. van Dijk, A.A., Frilander, M. and Bamford, D.H. (1995) Differentiation between minus- and plus-strand synthesis: polymerase activity of dsRNA bacteriophage phi 6 in an in vitro packaging and replication system. *Virology*, **211**, 320-323.

35. Kao, C.C., Singh, P. and Ecker, D.J. (2001) De novo initiation of viral RNA-dependent RNA synthesis. *Virology*, **287**, 251-260.

36. Ranjith-Kumar, C.T., Gutshall, L., Kim, M.J., Sarisky, R.T. and Kao, C.C. (2002) Requirements for de novo initiation of RNA synthesis by recombinant flaviviral RNA-dependent RNA polymerases. *J Virol*, **76**, 12526-12536.

37. Buck, K.W. (1996) Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv Virus Res*, **47**, 159-251.

38. 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.

39. 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.

40. 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*, **12**, 307-316.

41. 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.

42. 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.

43. Bruenn, J.A. (2003) A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res*, **31**, 1821-1829.

44. Ojala, P.M. and Bamford, D.H. (1995) In vitro transcription of the double-stranded RNA bacteriophage phi 6 is influenced by purine NTPs and calcium. *Virology*, **207**, 400-408.

45. 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.

46. Poranen, M.M., Salgado, P.S., Koivunen, M.R., Wright, S., Bamford, D.H., Stuart, D.I. and Grimes, J.M. (2008) Structural explanation for the role of Mn2+ in the activity of phi6 RNA-dependent RNA polymerase. *Nucleic acids research*, **36**, 6633-6644.

47. Arnold, J.J. and Cameron, C.E. (2004) Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg2+. *Biochemistry*, **43**, 5126-5137.

48. Barr, J.N. and Fearns, R. (2010) How RNA viruses maintain their genome integrity. *J Gen Virol*, **91**, 1373-1387.

49. Harki, D.A., Graci, J.D., Galarraga, J.E., Chain, W.J., Cameron, C.E. and Peterson, B.R. (2006) Synthesis and antiviral activity of 5-substituted cytidine analogues: identification of a potent inhibitor of viral RNA-dependent RNA polymerases. *J Med Chem*, **49**, 6166-6169.

50. Arnold, J.J., Vignuzzi, M., Stone, J.K., Andino, R. and Cameron, C.E. (2005) Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J Biol Chem*, **280**, 25706-25716.

51. Moustafa, I.M., Shen, H., Morton, B., Colina, C.M. and Cameron, C.E. (2011) Molecular

dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity. *J Mol Biol*, **410**, 159-181.

52. Cameron, C.E., Moustafa, I.M. and Arnold, J.J. (2009) Dynamics: the missing link between structure and function of the viral RNA-dependent RNA polymerase? *Curr Opin Struct Biol*.

53. Steitz, T.A. (1999) DNA polymerases: structural diversity and common mechanisms. *J Biol Chem*, **274**, 17395-17398.

54. Ranjith-Kumar, C.T. and Kao, C.C. (2006) Recombinant viral RdRps can initiate RNA synthesis from circular templates. *RNA*, **12**, 303-312.

55. 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.

56. Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Gotte, M., Konigsberg, W. and Cameron, C.E. (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol*, **16**, 212-218.

57. Di Marco, S., Volpari, C., Tomei, L., Altamura, S., Harper, S., Narjes, F., Koch, U., Rowley, M., De Francesco, R., Migliaccio, G. *et al.* (2005) Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. *J Biol Chem*, **280**, 29765-29770.

58. Palmer, A.G., 3rd. (2001) Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct*, **30**, 129-155.

59. Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z. and Kern, D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol*, **11**, 945-949.

60. Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E. and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438**, 117-121.

61. Tzakos, A.G., Grace, C.R., Lukavsky, P.J. and Riek, R. (2006) NMR techniques for very large proteins and rnas in solution. *Annu Rev Biophys Biomol Struct*, **35**, 319-342.

62. Tugarinov, V. and Kay, L.E. (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *Chembiochem*, **6**, 1567-1577.

63. Goger, M., McDonnell, J.M. and Cowburn, D. (2003) Optimization of Protein Assignment with Reduced Cycle Multidimensional NMR Experiments. *Spectroscopy*, **17**, 161-167.

64. Sibille, N., Hanoulle, X., Bonachera, F., Verdegem, D., Landrieu, I., Wieruszeski, J.M. and Lippens, G. (2009) Selective backbone labelling of ILV methyl labelled proteins. *J Biomol NMR*, **43**, 219-227.

65. Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. and Kay, L.E. (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J Biomol NMR*, **13**, 369-374.

66. Korzhnev, D.M., Kloiber, K. and Kay, L.E. (2004) Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: theory and application. *J Am Chem Soc*, **126**, 7320-7329.

67. Tugarinov, V. and Kay, L.E. (2006) Relaxation rates of degenerate 1H transitions in methyl groups of proteins as reporters of side-chain dynamics. *J Am Chem Soc*, **128**, 7299-7308.

68. Tugarinov, V., Sprangers, R. and Kay, L.E. (2007) Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J Am Chem Soc*, **129**, 1743-1750.

69. Boggs, P.T., Donaldson, J.R., Byrd, R.H. and Schanabel, R.B. . (1989) ODRPACK software for weighted orthogonal distance regression. *ACM Trans. Math. Software*, **15**.

70. Tugarinov, V., Ollerenshaw, J.E. and Kay, L.E. (2005) Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: an application to an 82-kDa enzyme. *J Am Chem Soc*, **127**, 8214-8225.

71. Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V. and Kay, L.E. (2004) Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc*, **126**, 3964-3973.

72. Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*, **147**, 138-146.

73. Vilfan, I.D., Candelli, A., Hage, S., Aalto, A.P., Poranen, M.M., Bamford, D.H. and Dekker, N.H. (2008) Reinitiated viral RNA-dependent RNA polymerase resumes replication at a reduced rate. *Nucleic Acids Res*, **36**, 7059-7067.

74. Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations [In Process Citation]. *J Magn Reson*, **147**, 138-146.

75. Nygren, J., Svanvik, N. and Kubista, M. (1998) The interactions between the fluorescent dye thiazole orange and DNA. *Biopolymers*, **46**, 39-51.

76. Arias, A., Arnold, J.J., Sierra, M., Smidansky, E.D., Domingo, E. and Cameron, C.E. (2008) Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic

analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *J Virol*, **82**, 12346-12355.

77. Wang, C. and Palmer, A.G., 3rd. (2002) Differential multiple quantum relaxation caused by chemical exchange outside the fast exchange limit. *J Biomol NMR*, **24**, 263-268.

78. Dittmer, J. and Bodenhausen, G. (2004) Evidence for slow motion in proteins by multiple refocusing of heteronuclear nitrogen/proton multiple quantum coherences in NMR. *J Am Chem Soc*, **126**, 1314-1315.

79. Laurila, M.R., Salgado, P.S., Stuart, D.I., Grimes, J.M. and Bamford, D.H. (2005) Back-priming mode of phi6 RNA-dependent RNA polymerase. *J Gen Virol*, **86**, 521-526.

80. Sarin, L.P., Poranen, M.M., Lehti, N.M., Ravantti, J.J., Koivunen, M.R., Aalto, A.P., van Dijk, A.A., Stuart, D.I., Grimes, J.M. and Bamford, D.H. (2009) Insights into the pre-initiation events of bacteriophage phi 6 RNA-dependent RNA polymerase: towards the assembly of a productive binary complex. *Nucleic Acids Res*, **37**, 1182-1192.

81. Patel, S.S., Wong, I. and Johnson, K.A. (1991) Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry*, **30**, 511-525.

82. Ren, Z., Wang, H. and Ghose, R. (2010) Dynamics on multiple timescales in the RNA-directed RNA polymerase from the cystovirus phi6. *Nucleic Acids Res*, **38**, 5105-5118.

83. Yang, H., Makeyev, E.V. and Bamford, D.H. (2001) Comparison of polymerase subunits from double-stranded RNA bacteriophages. *J Virol*, **75**, 11088-11095.

84. 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.

85. Yang, X., Welch, J.L., Arnold, J.J. and Boehr, D.D. (2010) Long-range interaction networks in the function and fidelity of poliovirus RNA-dependent RNA polymerase studied by nuclear magnetic resonance. *Biochemistry*, **49**, 9361-9371.

86. Ren, Z. and Ghose, R. (2011) Slow conformational dynamics in the cystoviral RNA-directed RNA polymerase P2: influence of substrate nucleotides and template RNA. *Biochemistry*, **50**, 1875-1884.

87. Dima, R.I. and Thirumalai, D. (2006) Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci*, **15**, 258-268.

88. Hansen, D.F., Vallurupalli, P. and Kay, L.E. (2009) Measurement of methyl group motional parameters of invisible, excited protein states by NMR spectroscopy. *J Am Chem Soc*, **131**, 12745-12754.

89. Vallurupalli, P., Hansen, D.F. and Kay, L.E. (2008) Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci U S A*, **105**, 11766-11771.

90. Vallurupalli, P., Hansen, D.F. and Kay, L.E. (2008) Probing structure in invisible protein states with anisotropic NMR chemical shifts. *J Am Chem Soc*, **130**, 2734-2735.

91. Hansen, D.F., Vallurupalli, P., Lundstrom, P., Neudecker, P. and Kay, L.E. (2008) Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *J Am Chem Soc*, **130**, 2667-2675.