DISSERTATION

LONG RANGE INTERACTION NETWORKS WITHIN $3D^{POL}$ AND THE ROLES THEY PLAY IN PICORNAVIRUS GENOME REPLICATION AND RECOMBINATION

Submitted by

Colleen L. Watkins

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2020

Doctoral Committee:

Advisor: Olve B. Peersen

Robert Cohen
P. Shing Ho
Jeffery Wilusz

ABSTRACT

LONG RANGE INTERACTION NETWORKS WITHIN $3D^{POL}$ AND THE ROLES THEY PLAY INPICORNAVIRUS GENOME REPLICATION AND RECOMBINATION

Picornaviruses contain a single-stranded positive sense RNA genome approximately 7.5kb in length. The genome encodes for a single polyprotein that can future be divided into three functional regions; the P1 region containing the viral capsid proteins, the P2 region whose proteins function primarily in membrane rearrangement during viral replication, and the P3 region which contains four protein responsible for RNA replication. The final protein in the P3 region is $3D^{pol}$, an RNA-dependent RNA polymerase (RdRP) whose structure is analogous to a "right hand" with fingers, palm and thumb domains, and around which this dissertation will be centered.

Section one of this work investigates the roles three regions within the fingers domain play in the catalytic cycle of $3D^{pol}$: "The kink" located within the index finger, "the gateway" found on the pinky, and "the sensor", which bridges the two beta-strands of the middle finger. This study demonstrates that the kink residues are involved in RNA binding as mutations to these residues result in decreased initiation time and elongation complex lifetime. The gateway residues are found to act as a molecular stop against which the template-RNA strand positions itself post-translocation, eventually resetting the active site for the next round of nucleotide incorporation. Lastly the sensor residues serve two key functions: 1) A final checkpoint to determine the correct nucleotide has entered the active site, and 2) As a possible source for proton donation to the pyrophosphate leaving group formed during catalysis. The inter-connected nature of the residues investigated in this section give rise to the idea that it is not individual

residues alone that control major steps during the catalytic cycle, but instead that long ranging interaction networks within the different polymerase domains are ultimately responsible for controlling different actions carried out by the polymerase.

Section two of this work looks at the long-range interaction networks within $3D^{pol}$ by dissecting the roles each polymerase domain plays in catalytic cycle. Through generation of chimeric polymerases it was determined that the pinky finger, with some influence by the fingers domain, controls RNA binding, the palm domain dictates nucleotide discrimination, and nucleotide capture and active site closure rates. It was also established that the thumb domain controls translocation, and an interaction between the palm and thumb domains was needed to generate a viable virus, supporting the idea of interface I, a protein-protein interface that was discovered in the first $3D^{pol}$ crystal structure. What is most striking about these findings is that unlike other single subunit polymerases that perform translocation by using a large swinging motion within the fingers domain, viral RdRPs use an entirely different domain altogether.

The last section of this work deals with viral recombination, an event that is carried out at a low frequency during virus replication. Recombination is proposed to be a mechanism by which mutations can be purged from the genome independent of polymerase fidelity. This study carries out a mechanistic investigation into how mutation of residue 420 from a leucine to an alanine affects polymerase replication kinetics. It also takes a look at the mutation of residue 64 from a glycine to a serine, a previously identified mutation that results in a high-fidelity polymerase, in the presence and absence of L420A. This work revealed that mutations L420A and G64S operate independently of each other by affecting different steps in the catalytic cycle with G64S increases in fidelity predominately from monitoring nucleosugar positioning while L420A affects nucleobase positioning and polymerase grip on the product RNA strand.

ACKNOWLEDGEMENTS

I have been lucky enough to have made four great decisions in my life, and they have predominately centered around my love for learning and science. My fourth greatest decision in life was to join Olve's lab. I have never felt more at home in a lab than this one, and for that I will always be grateful. I know in the years spent here I have been able to grow both scientifically and personally and I can say that I owe so much you for allowing me the space and security to do both. It may have been challenging at times, but I have never questioned the place those challenges have come from knowing that I have always been pushed me to my fullest potential. I also need to thank my work wife, Grace. I am thankful for the laughter and tears we have shared and will forever be amazed by how much you have helped me learn to go with the flow at some of my most stressful times.

The third greatest decision I have made was deciding to be a scientist and not a vet. And for that I must thank my family. Mom, Dad, and Stephen you have been with me from the start and never wavered in your love and support for me. I am thankful for all the hours you have spent listening to me vent about experiments, papers you have read, math homework you have check, trips to and from school, side of the highway pickups, and not leaving on mornings we were late to zero hour because I jumped in the shower late. This degree has been a long journey and you all have had one heck of a front row seat to it. I will probably never be able to express how truly grateful I am to all of you, but I know I never would have managed to get here without all of your help and love.

The single greatest decision in life I have made is marrying the love of my life (with dropping out of grad school the first time being a close second). And thank goodness Matt that you have been around for all of it. I have asked so much of you over the years chasing this dream

of mine and you have put up with everything. I will forever be in your debt. I will never know how I got through that semester in Atlanta, but I know I would not have been able to do it without you. Thank you for following me around the country, putting up with my constant needing us to move for my job, nights I was too tired to cook dinner, and way too many hours commuting to work over these past five years. You are truly the man of my dreams and in many ways have earned this degree too. Thank you and I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 A brief history of poliovirus discovery, infection, and eradication efforts

1.1.1 Discovery of poliovirus

Polio is an ancient disease with documented instances dating as far back as 1580 BC (Mehndiratta et al., 2014). Studies on Egyptian mummies as well as artifacts provide indications of polio's presence (Baicus, 2012). At that time in history outbreaks of poliovirus were relatively infrequent and likely due to poor sanitation and hygiene, which resulted from exposure to the virus early in childhood, generating a large population with immunity in a community (Minor, 2014; Nathanson and Martin, 1979). In 1789 Dr. Michael Underwood first clinically documented polio as a disease with a presentation that causes paralysis to the lower extremities of children (Underwood, 1789). Outbreaks of polio continued to rise throughout the 19th century, however the causative agent remained elusive. Finally, in 1908 Landsteiner and Popper concluded that poliovirus was the cause of the polio outbreaks (Landsteiner and Popper, 1908). Even with this discovery, poliovirus epidemics continued to occur during the early part of the 20th century with one of the earliest outbreaks in the United States occurring in June of 1916 (Mehndiratta et al., 2014). This epidemic resulted in over 27,000 cases of polio and more than 6,000 fatalities. From that point on, epidemics of poliovirus occurred each summer, reaching nearly 60,000 cases a year during outbreaks between 1940 and 1950 (Mehndiratta et al., 2014).

1.1.2 Poliovirus infection

1.1.2.1 Pathogenesis and transmission

Poliovirus (PV) is a member of the virus family *Picornaviridae* and genus *enterovirus*, and is transmitted via an oral fecal route, as are other enterovirus family members. The virus, in its virion form (RNA encapsulated by a capsid protein shell), is ingested by its host, often via a contaminated water source (Baicus, 2012). Only primates are naturally susceptible to the poliovirus, with humans being the only natural host. Susceptibility of non-human primates has been induced and varies by species (Koike and Nomoto, 2010). After ingestion, the virus begins to replicate in the oropharyngeal and intestinal mucosa (Koike and Nomoto, 2010; Racaniello, 2006). During this time, infected individuals shed live viruses through their feces, with 80% of them continuing to shed viruses for up to two weeks, prolonging the spread of the infection. It is also at this point the virus must first cross a physical barrier to gain access to the lymphatic system; the virus enters either the tonsils in the pharynx or the Peyer's patches found in the gastrointestinal tract. From there, the virus enters the bloodstream and establishes viremia (virus in the blood), and once that has occurred, symptoms such as sore throat, gastrointestinal upset, and fever may be associated with the relatively minor infection (Bodian and Horstmann, 1965). About a week after viremia establishment, antibodies against the virus have formed, and in 99% of cases, the viral infection is cleared from the human host. In the other 1% of cases, the virus is able to cross another physical barrier and gain access to the central nervous system, either by directly crossing the blood brain barrier or by retrograde axonal transport after viral uptake at a neuromuscular junction (Racaniello, 2006). If poliovirus gains access to the central nervous system it preferentially replicates within motor neurons, which leads to the characteristic flaccid paralysis, particularly in the lower limbs, associated with poliomyelitis infections.

1.1.2.2 Viral replication

Much research has gone into understanding poliovirus replication at the cellular and molecular levels (Figure 1.1) (De Jesus, 2007). Replication begins with viral uptake into a cell by virion binding to the human poliovirus receptor CD155 (Ren and Racaniello, 1992). Upon binding, the viral particle is endocytosed into the cell and two capsid proteins, VP4 and VP1, create pores in the newly formed vesicle membrane that allow the viral RNA to enter the cytoplasm, where viral replication ultimately occurs (Fricks and Hogle, 1990). The positive-sense RNA genome of the virus can be divided into three sections: the 5′-non-translated region (NTR), a single open reading frame, and the 3′-non-translated region (Figure 1.2) (De Jesus, 2007). The 5′-end of the genome is linked to the viral protein VPg as a means to protect the viral RNA. Following release of the virus into the cytoplasm, the host TBP2 protein, or unlinkase, cleaves the VPg and enables viral RNA translation to begin at the internal ribosome entry site (IRES), found within the viral RNA's 5′-NTR (Virgen-Slane et al., 2012). Open reading frame translation yields a single ~250 kDa polyprotein. The polyprotein is comprised of three functional regions: P1, P2, and P3. This polyprotein is processed by virally encoded proteases, generating approximately 12 viral proteins as well as functional intermediates (Palmenberg et al., 2010). P1 encodes the structural proteins that will ultimately form the viral capsid, P2 proteins are involved in membrane rearrangements required for genome replication, and the P3 proteins are directly responsible for genome replication, including the RNA-dependent RNA polymerase $3D^{pol}$.

Genome replication occurs at Golgi-derived membrane-encapsulated sites, called replication complexes (Belov and van Kuppeveld, 2012), and is carried out by the virally

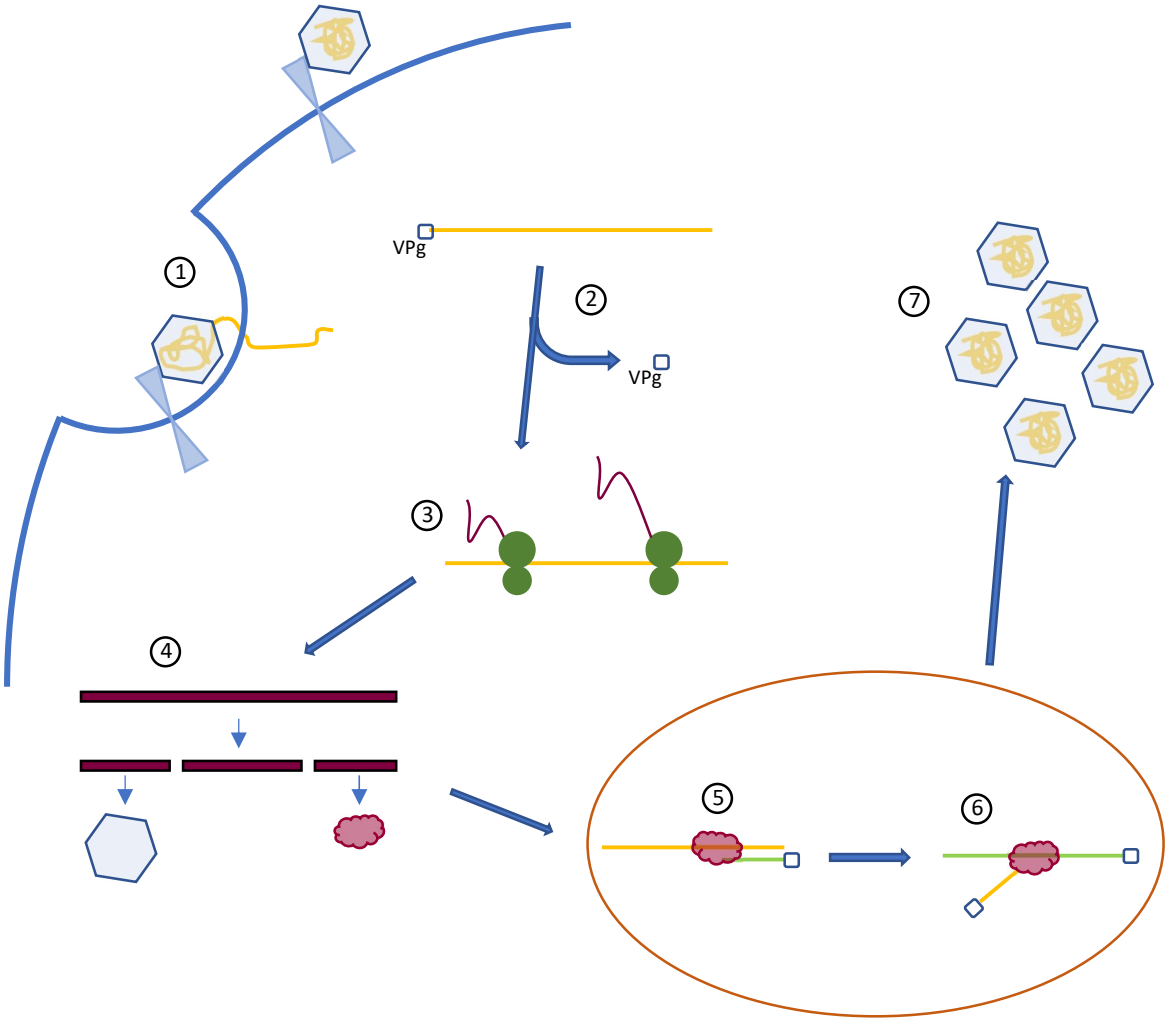

Figure 1.1: Picornavirus life cycle.

The poliovirus life cycle begins with binding of a viral particle to the extracellular CD155 receptor (1), which is followed by virus uptake and genome release (2). VPg cleavage occurs prior to the start of genome translation (3) to produce the viral polyprotein that is subsequently cleaved into multiple viral proteins (4). Genome replication occurs in membrane encapsulated replication complexes, first using the viral genome as a template for negative strand replication (5) followed by positive strand replication (6) to produce progeny genomes. Genomes are packaged into new viral particles that exit the cell and propagate the viral infection (7).

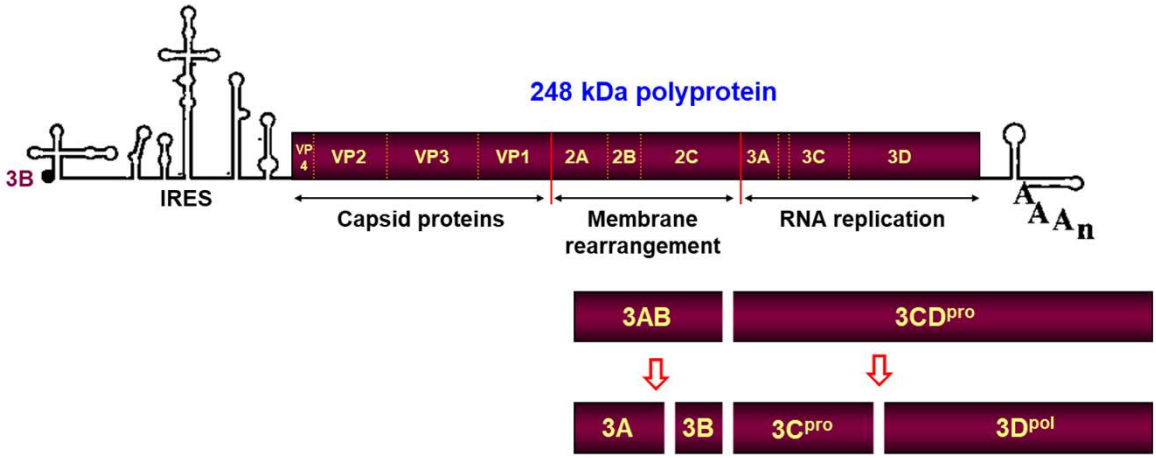

Figure 1.2: Structure of the picornaviral genome.

Cartoon of the poliovirus genome highlighting the 5′-NTR that includes the IRES, single open reading frame encoding for the viral polyprotein, and the 3′-NTR. The polyprotein contains three distinct regions, the P1 region containing the structural proteins, P2 that contains non-structural proteins involved in membrane rearrangement, and the P3 region whose non-structural proteins are used for genome replication. The polyprotein is cleaved by virally encoded proteases to generate both intermediate and mature viral proteins.

encoded RNA-dependent RNA polymerase $3D^{pol}$ (Flanegan and Baltimore, 1977). This process begins by using the viral genome as a template for negative strand synthesis, which is then in turn used for positive strand synthesis generating new viral genomes (Morasco et al., 2003; Murray and Barton, 2003). Both strands are synthesized by $3D^{pol}$ and require a virally derived protein primer called VPg, or 3B, that has been doubly uridylated through a process called uridylylation (Paul et al., 1998). These newly synthesized viral genomes are ultimately packaged with the capsid proteins into new progeny virions, which are released from the infected cell to further propagate the infection (De Jesus, 2007).

1.1.3 Eradication efforts

Vaccine development efforts for poliovirus began in 1935, but it was not until 1953 that a successful vaccine was produced. Jonas Salk introduced his formalin inactivated polio vaccine, IPV, which contained three wild poliovirus strains that vary primarily in their capsid proteins, and by April 1955 this vaccine was used throughout the United States (Monto, 1999). IPV is costly to make and, if the viruses are not inactivated properly, can result in poliovirus infections for the recipients, as was the case in the Cutter incident where children were administered live viruses that resulted in infections and fatalities (Nathanson and Langmuir, 1963). It should also be noted that this vaccine is administered intramuscularly as an injection, and predominantly provides humoral immunity, or immunity derived from antibody secretion in body fluids. Thus, this vaccine provides no gut immunity. An oral poliovirus vaccine, OPV, was introduced in 1956 by Albert Sabin (Sabin, 1960). It also contains all three poliovirus strains in a live attenuated form, which allows the virus to replicate and generate immunity, but the virus is not able to cause disease. OPV is much easier to administer in comparison to IPV, and provides exceptional

gut immunity to the recipient. By 1963, OPV was the vaccine type predominantly used in the United States and it remained the preferred choice until 2000 (De Jesus, 2007).

OPV, which was donated by Sabin in 1972 to the World Health Organization (WHO), was also the preferred choice when it embarked on a global polio eradication effort in 1988 (Baicus, 2012). There is, however, a major consequence of choosing the OPV vaccine as poliovirus itself is genetically unstable. This instability leads to cases of Vaccine-Associated Paralytic Poliomyelitis (VAPP) from now circulating vaccine-derived polioviruses (cVDPV) as a result of mutations in the poliovirus genome during replication or through genome recombination with other circulating enterovirus strains (Muslin et al., 2019). Today OPV is no longer used in countries that are deemed polio-free. As of 2019, poliovirus is found endemically only in three countries: Afghanistan and Pakistan, which are affected by both wildtype poliovirus type 1 and VDPV type 2, and Nigeria, which is only affected by VDPV type 2 (World Health Assembly, 2019). Both the genetic instability of OPV and the propensity for recombination with other enteroviruses derive from mechanisms centered around the viral RNA-dependent RNA polymerase (RdRP). Although the issues surrounding OPV can be circumvented by the use of IPV, the difficulties in IPV administration and higher cost to produce warrant further study of these polymerase involved mechanisms.

1.2 Poliovirus polymerase $3D^{pol}$: overview of polymerase structure and function

1.2.1 Classes of single subunit polymerases

Polymerases are responsible for one of the most ubiquitous processes for all forms of life, the replication of genetic material from a template. This genetic material can be either in the form of DNA or RNA and can be transcribed into either form depending on the class of

polymerase. There are four classes of polymerases: DNA-dependent DNA polymerases (DdDP), DNA-dependent RNA polymerases (DdRP), RNA-dependent DNA polymerases (RdDP), and RNA-dependent RNA polymerases (RdRP) (Cheetham, 1999; Hansen et al., 1997; Huang, 1998; Kohlstaedt et al., 1992; Sousa et al., 1993; Thompson and Peersen, 2004). Crystal structures of members of each class have shown a conserved folded structure analogous to a cupped right hand, with fingers, palm, and thumb domains, and the active site located within the palm domain (Figure 1.3) (Shu and Gong, 2017). Polymerases can also be found as either multi-subunit protein complexes, such as human RNA polymerase II (RNAPII), or as single subunit proteins like the *Thermus aquaticus* (Taq DNA) polymerase (Svetlov and Nudler, 2013). For the purposes of this discussion predominately the work described will be surrounding single subunit polymerases.

1.2.2 Mechanisms of action

1.2.2.1 Nucleic acid addition cycle

All polymerases use a general mechanism for nucleic acid incorporation (Steitz, 1998). This mechanism relies on two metal ions, almost exclusively magnesium, and two conserved aspartic acid residues to catalyze a nucleophilic attack by the 3′-hydroxyl on the α-phosphate of the incoming nucleotide triphosphate (NTP). Phosphodiester bond formation begins with deprotonation of the 3′-hydroxyl on the product strand by a general base. Although this general base has not been identified, molecular dynamics simulations using RNAPII have demonstrated the feasibility of this reaction being carried out by the bulk water present near the 3′-hydroxyl, which could also be used by other polymerases (Zhu et al., 2007). The now deprotonated 3′-hydroxyl is stabilized by metal A and poised for a nucleophilic attack on the incoming NTP's α-

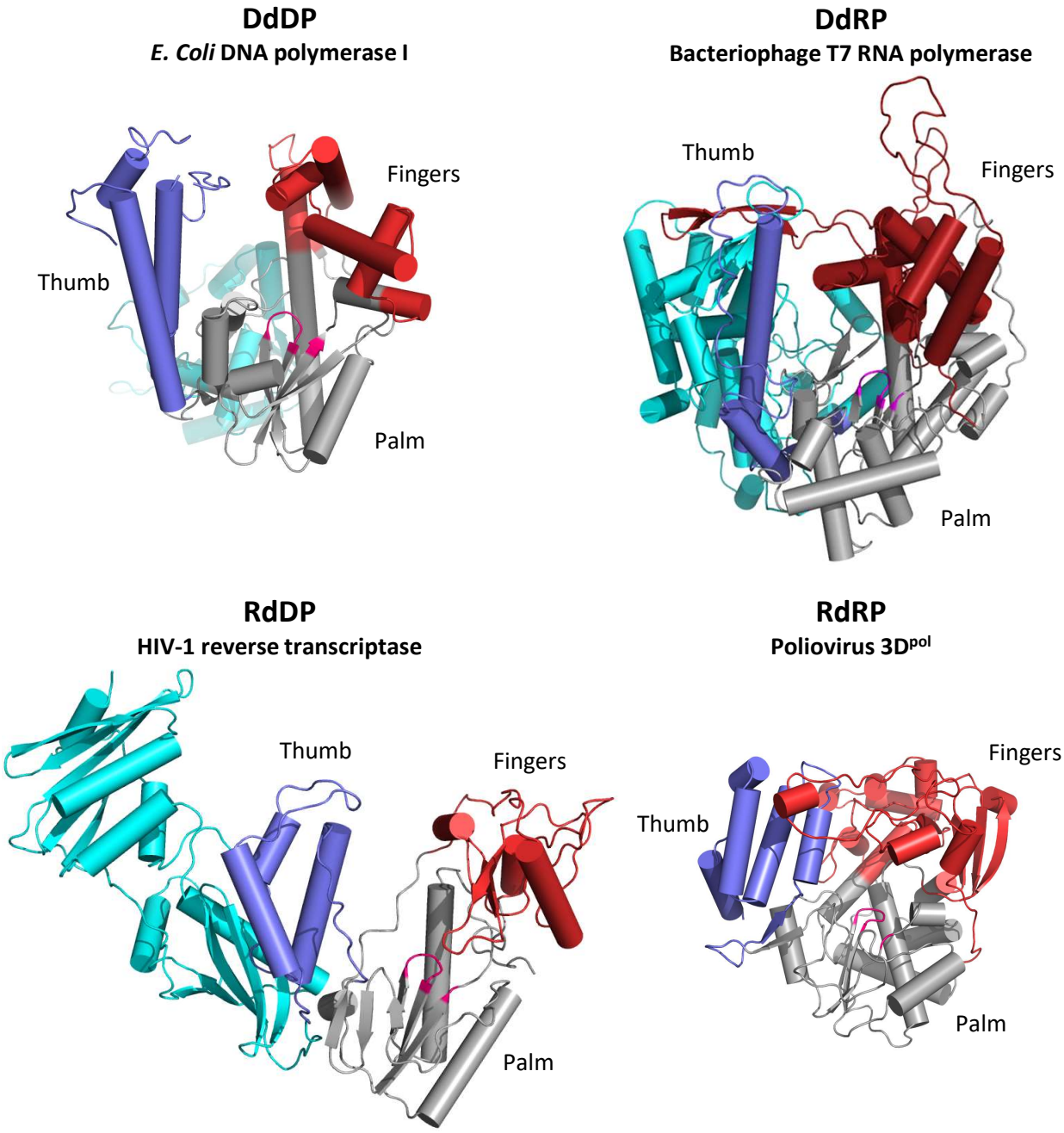

Figure 1.3: Comparison of crystal structures from representatives for each single subunit polymerase class.

Structures are oriented with the thumb domain in blue on the left, looking into the polymerases through the NTP/dNTP entry channel. Fingers domains are shown in red, and palm domains are in gray with the conserved magnesium ions shown as magenta spheres. Additional polymerase domains are shown in cyan.

phosphate. This proper positioning is due to coordination of the two metal ions by the two conserved aspartic acid residues.

Once the nucleophilic attack has occurred and the phosphodiester bond is formed, pyrophosphate is created and subsequently protonated by a general acid. This general acid has been identified for many different replicative polymerases, and while it may vary between residues, from a histidine in the multi-subunit RNAPII, to lysine in the single-subunit DNA-dependent polymerases, and an arginine in the HIV reverse transcriptase (HIV RT), it is always a basic residue in close proximity to the incoming NTPs α-phosphate (Figure 1.4) (Castro et al., 2009; Peersen, 2019a). A lysine residue in the poliovirus polymerase, lysine 359, has been suggested to be the proton donor for this polymerase (Castro et al., 2009), however it should be noted that it is not in a position to contact the α-phosphate. Molecular dynamics simulations instead suggest that K359 assists in the transport of NTPs into and pyrophosphate out of the polymerase active site, the former of which would also allow K359 to act as an indicator of correct NTP binding to trigger closure of the enzyme for catalysis (Shen et al., 2012a). This hypothesis could also explain the results that either mutation directly of lysine 359 to an arginine, or other mutations that favor K359 positioning within the polymerase (N370A or K228A) have direct effects on polymerase incorporation rate and fidelity (Shi et al., 2019).

1.2.2.2 Methods for nucleotide selection

Nucleotide selection in polymerases occurs via an induced fit mechanism. For most classes of polymerase, the incoming nucleotide enters into a pre-insertion site above the active site, in a position where it can hydrogen bond with the templating nucleotide, as well as additional residues to verify both the correct nucleobase (i.e. A,T,C,G, or U) and nucleosugar (ribose vs. deoxy-ribose) (Johnson, 2010; Kati et al., 1992; Moscato et al., 2016; Temiakov et al., 2004;

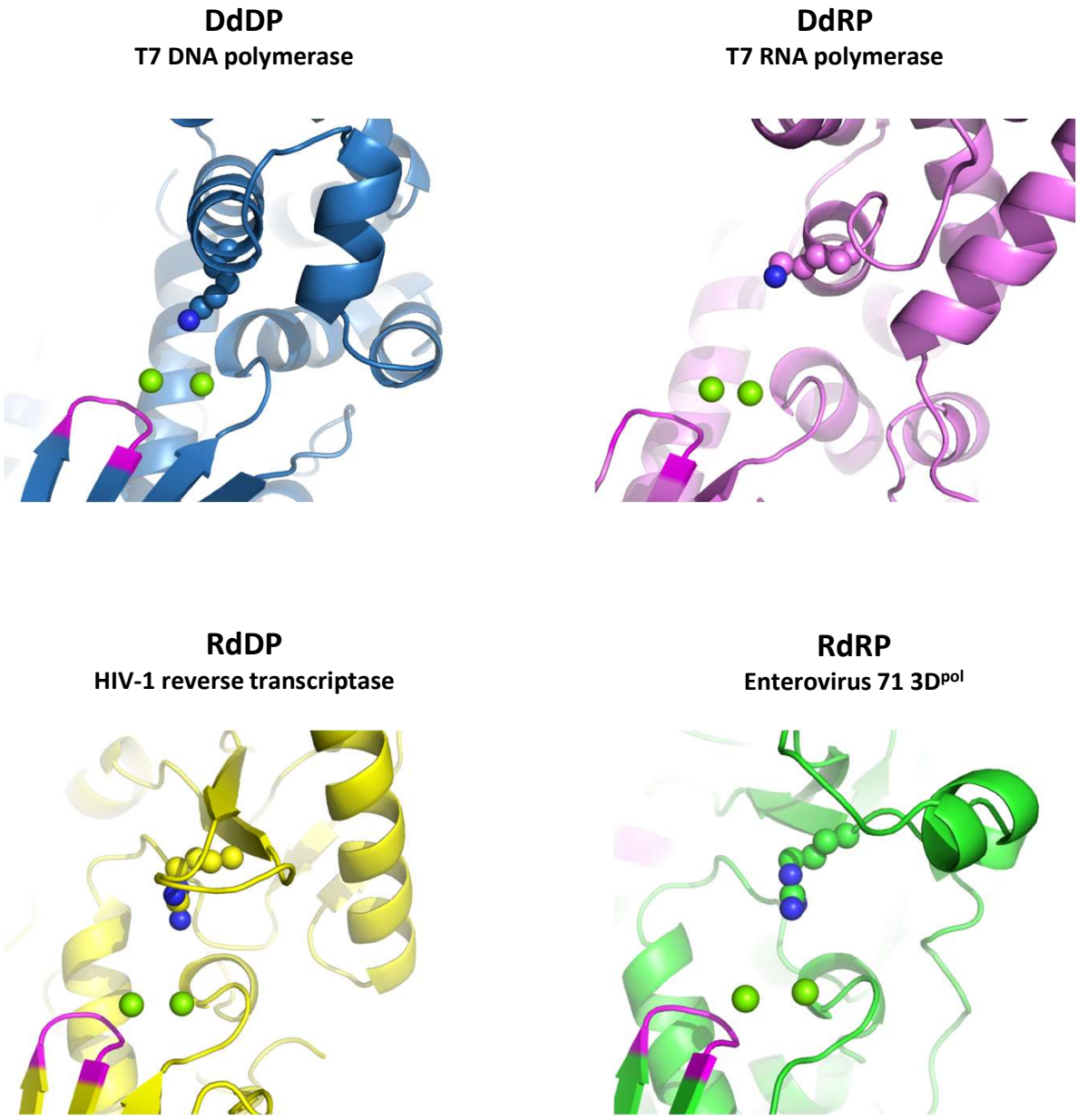

Figure 1.4: Structural comparison of the general acid use by single subunit polymerases.

Crystal structures from four single subunit polymerase representative members. Conserved magnesium ions are shown as spheres. Loop within motif C for each polymerase is highlighted in magenta. T7 DNA polymerase [1T7P], T7 RNA polymerase [1S76], HIV-1 reverse transcriptase [4PQU], Enterovirus 71 RdRP [5F8J]. Polymerase alignments used to create this figure:

Peersen, Olve. (2019). Polymerase Structure Alignments - v2a Update (October 2019) (Version 2a) [Data set]. Viruses. Zenodo. http://doi.org/10.5281/zenodo.3466420

Wong et al., 1991). If those criteria are satisfied, the incoming nucleotide and its templating base are pushed into the active site for catalysis, discussed in the previous section. This large movement is carried out by a swinging motion in the fingers domain, for most polymerases. By utilizing a preinsertion site, the active site of the polymerase remains static; both the prebound metal A and coordinating residues are in a near optimal geometry and can quickly catalyze the reaction once the correct nucleotide is bound.

Extensive steady state and pre-steady state kinetics work using HIV RT supports the idea of an active-site closure type of event serving as a fidelity determining step for the polymerase (Kati et al., 1992). When nucleotide incorporation kinetics were investigated in conjunction with elongation complex (EC) dissociation kinetics, it was shown that correct dNTP binding by the enzyme resulted in a faster dissociation rate of the polymerase from the DNA, which had been elongated by one nucloetide. However, if the ability to carry out catalysis was removed via a ddNTP, EC lifetime became longer, suggesting a conformational change step that increased stability upon correct nucleotide binding. It should also be noted that in the presence of an incorrect nucleotide, both the rates of enzyme isomerization and complex stability were reduced, thus giving rise to the idea this conformational change could serve as a fidelity checkpoint for the polymerase, with additional elemental effects analysis on HIV RT showing this isomerization step was occurring prior to catalysis (Kati et al., 1992). Crystallographic and NMR studies of DNA polymerase B show not only the large fingers domain movements associated with active site closure, but also the chemical shifts associated with different residues varied depending on whether the incoming nucleotide is correct or not, supporting the concept that structural changes to these polymerases are due to an induced fit produced only when the correct nucleotide is bound (Moscato et al., 2016). Kinetic analysis using the T7 DNA polymerase to compare the

rates of polymerization in the presence of correct and incorrect nucleotides shows a large decrease in rate not only for the incorporation of the incorrect nucleotide, but also in the incorporation rate of the next nucleotide, despite if the second nucleotide is the correct one (Wong et al., 1991). These findings support that the induced fit mechanism is able to be governed by both the incoming nucleotide and the 3′-end nascent nucleic acid strand to which the incoming nucleotide will be incorporated.

1.2.2.3 Catalytic cycle completion: nucleic acid strand translocation

Structural work using the bacteriophage T7 DdRP has elucidated information about the structural changes surrounding translocation (Yin and Steitz, 2004). This process is facilitated by movement of the O-helix tyrosine that stacks on the newly elongated product-RNA strand after catalysis has occurred. The O-helix must rotate by 22.5° to close the active site for chemistry and reverse by the same amount in order for translocation to occur. This movement places the new templating base into the preinsertion site and drives the base pair located in the active site out to reset for the next round of catalysis. This structural study also proposed that this movement does not occur until pyrophosphate release has happened. This observation has two consequences: first, this stepwise reaction allows for the coupling of translocation to pyrophosphate release to energetically drive translocation, and second, an energetic coupling would result in translocation occurring by a power stroke mechanism.

A second mechanism for translocation by the T7 DdRP has also been proposed in which the polymerase uses a passive Brownian ratchet. In this mechanism, small intermolecular motions result in a sliding of the polymerase between the catalytic cite and the post-translational site. Until NTP binding occurs, the polymerase is free to move between these two positions and establish an equilibrium between them . When the RNA is in the post-translational site, a new

NTP is able to bind, thus driving the equilibrium toward the post-translational state and, in turn, incorporating the next nucleotide (Yu and Oster, 2012). Single molecule studies with T7 DdRP have shown there is a bias in the equilibrium between the two states, favoring the post-translational state 3-to-1 over the pre-translational state (Thomen et al., 2005). Additional kinetics studies have shown a favoring of the pre-translocation site 4-fold over the post-translational site, however the authors note this may be a consequence of the RNA-DNA hybrid sequence being formed stabilizing the transcription bubble, resulting in a high energetic cost to disrupt (Guo and Sousa, 2006).

Most recently two molecular dynamics simulation studies have attempted to reconcile the two mechanisms for translocation (Da et al., 2015, 2017). These studies have used a combination of single-molecule data (Thomen et al., 2005), measuring the forces surrounding polymerase translocation, with structural studies (Temiakov et al., 2004; Yin, 2002; Yin and Steitz, 2004) dissecting the movements of the O-helix and its conserved tyrosine to generate a model that combines both translocation mechanisms. The data showed that translocation is carried out predominately by the Brownian ratchet mechanism which is used to detect changes in nucleic acid sequence encountered by the polymerase (Da et al., 2017). Full polymerase translocation commitment is dictated by the small power stroke movement associated with the O-helix tyrosine, which helps to energetically stabilize the post-translocation structure and assists with incoming nucleotide selection (Yu and Oster, 2012).

1.3 The catalytic cycle of $3D^{pol}$, a viral RdRP

1.3.1 Current information about the catalytic cycle

For viral RdRPs, there is a structural connection between the fingers domain and the thumb domain (Figure 1.5) (Ferrer-Orta et al., 2006). This causes a restriction in the movements by the fingers domain and, as a consequence, these polymerases have evolved to have different mechanisms for nucleotide selection and bringing the base pair into the correct proximity for catalysis (Gong and Peersen, 2010). The kinetics of the catalytic cycle have been well established and generally follow the reaction pathway established for other polymerases. Extensive work by the Cameron group has shown that, upon nucleotide binding, the polymerase undergoes a rate limiting conformational change prior to catalysis (Arnold and Cameron, 2000, 2004; Arnold et al., 2004; Castro et al., 2007, 2009; Gohara et al., 2004). Using a variety of rapid chemical quench and kinetic modeling, they were able to determine the rates for five steps in the catalytic cycle. Interestingly, these results show a clear decoupling of active site reopening from pyrophosphate release, which suggests that active site reopening and translocation are two distinct steps in the catalytic cycle (Arnold and Cameron, 2004; Arnold et al., 2004; Gohara et al., 2004). This decoupling would also limit the polymerase's ability to translocate strictly via a power stroke mechanism. In addition, a slow molecular isomerization step, prior to the first nucleotide addition step, was also detected, which could be either a repositioning of the RNA after binding or a closing of the polymerase after RNA binding, a step which would be very similar to the closing of the polymerase that is seen for other polymerase families (Arnold and Cameron, 2000).

To further determine the molecular interactions involved in the catalytic cycle, a series of EC crystal structures were solved to determine the molecular movements associated with

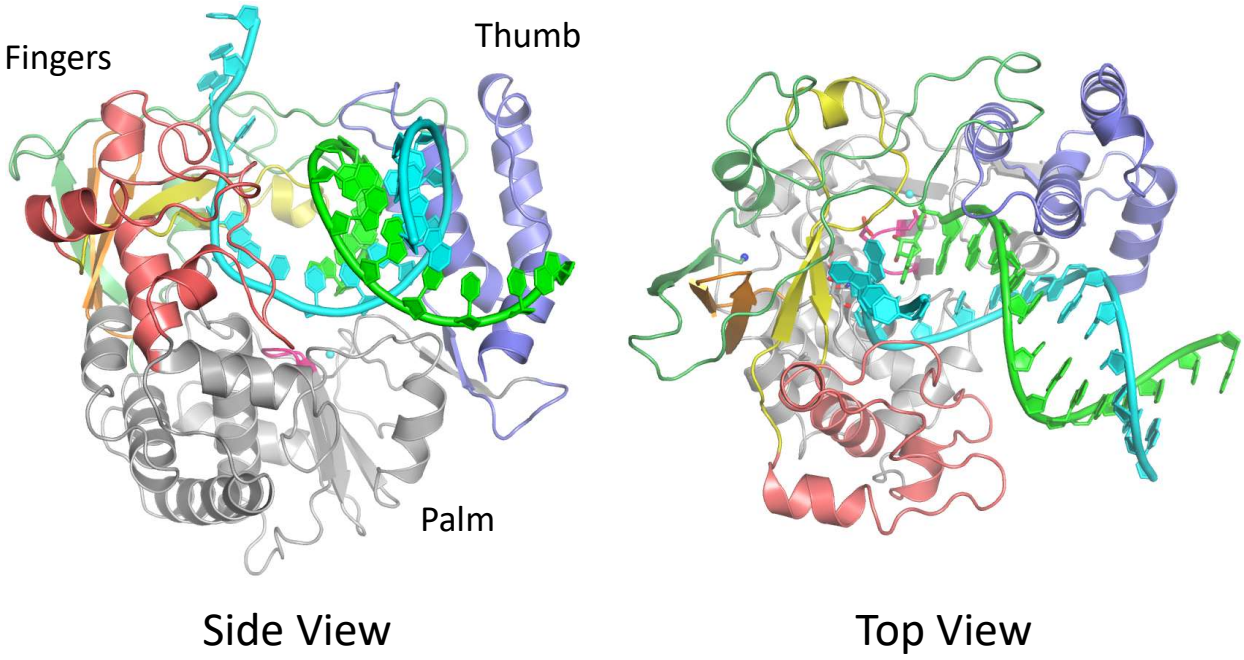

Figure 1.5: The picornavirus RdRP elongation complex.

(A) $3D^{pol}$ has a conserved fold similar to that of a cupped right hand with three domains: fingers shown in multiple colors on the left side of the polymerase to signify each finger, palm denoted in gray and containing the active site highlighted in magenta, and thumb colored slate. The template RNA strand is shown in cyan, and the product RNA in green. (B) The polymerase elongation complex structure looking down the template RNA entry channel to emphasize active site positioning of the two RNA strands. Crystal structure used for figure [3OL6]. Structure published in Gong and Peersen, 2010.

product-strand elongation by the polymerase (Gong and Peersen, 2010; Gong et al., 2013). This nucleotide incorporation process begins with the templating base in the active site, but the polymerase bound metal A is not in a position to coordinate the 3′-end for catalysis. Upon NTP binding, hydrogen bond formation between the 2′-hydroxyl on the incoming NTP and residues Asn 297 and Ser 288 within the polymerase starts a cascade that results in the movement of motifs A and D towards motif C to close the polymerase active site. Motif C forms a beta sheet with motif A, and as a result, metal A is moved up into the active site via movements of aspartic acid residues 233 and 329. Now this correctly structured active site is able to catalyze the nucleophilic attack by the 3′- hydroxyl on the incoming NTP α-phosphate using the same coordination as other polymerases for metals A and B to stabilize the deprotonated 3′-hydroxyl. Post-catalysis, the movements of motifs A and D are reversed, causing the movement of metal A out of the active site and reopening of the active site to allow for pyrophosphate release. It should be noted that no movements of the RNA strands occur in conjunction with pyrophosphate release.

1.3.2 Current hypothesis on how RNA strand translocation occurs

Presently the largest piece of information that is missing about the catalytic cycle in RdRPs is how the process of translocation occurs. Due to the tethering of the fingers domain and the lack of a structural homolog to the O-helix, the large swing motion of the fingers domain to drive translocation via a power stroke mechanism is not an option. However, there are two key pieces of evidence that have emerged that may shed light on the missing step. The first is a series of structures of the EV-71 elongation complex that show a clear half-step movement of the product-strand RNA prior to the movement of the template-strand RNA, a clear asymmetric RNA strand movement within the polymerase, which has not been the case with other

polymerases (Figure 1.6) (Shu and Gong, 2016). The second important piece of evidence is a series of crystal structures which highlight a loop structure in the polymerase's motif B, deemed the translocation loop (Sholders and Peersen, 2014). These structures show an in-and-out movement of a loop in motif B, and when it is in the out position the loop would sterically clash with the template-strand RNA. The combination of these two pieces of evidence suggest that RNA translocation occurs not entirely by a power stroke mechanism. Instead it is possible that RNA translocation is a combination of both the power stroke and Brownian ratchet mechanisms, much like the evidence for T7 DdRP is suggesting. These data point to the product strand first moving by a Brownian ratchet mechanism as no large movements were identified in the EV-71 structures, which would allow the product-strand RNA to move between the two observed positions freely until the template-strand is translocated. The second step in RNA translocation could have a power stroke-like mechanism during which the product-strand RNA completes its base step movement and the template-strand RNA moves in its entirety which. The motif B translocation loop may assist in the template-strand movement and prevent any backward movement, essentially serving as a committal step for catalytic cycle completion.

1.3.3 Polymerase dynamics and long-range interaction networks: the role they play in the catalytic cycle of 3D^{pol}

As viral RdRPs are limited by their connection between the fingers domain and the top of the thumb domain, it is conceivable that instead of large domain movements these polymerases may operate via smaller molecular movements within a large interaction network (Cameron et al., 2009; Moustafa et al., 2014; Yang et al., 2010). The existence of a dynamic long-range interaction network would support findings that residues distant from the active site have effects on polymerase fidelity (Arnold et al., 2005). The mutation of residue 64 from a glycine to a

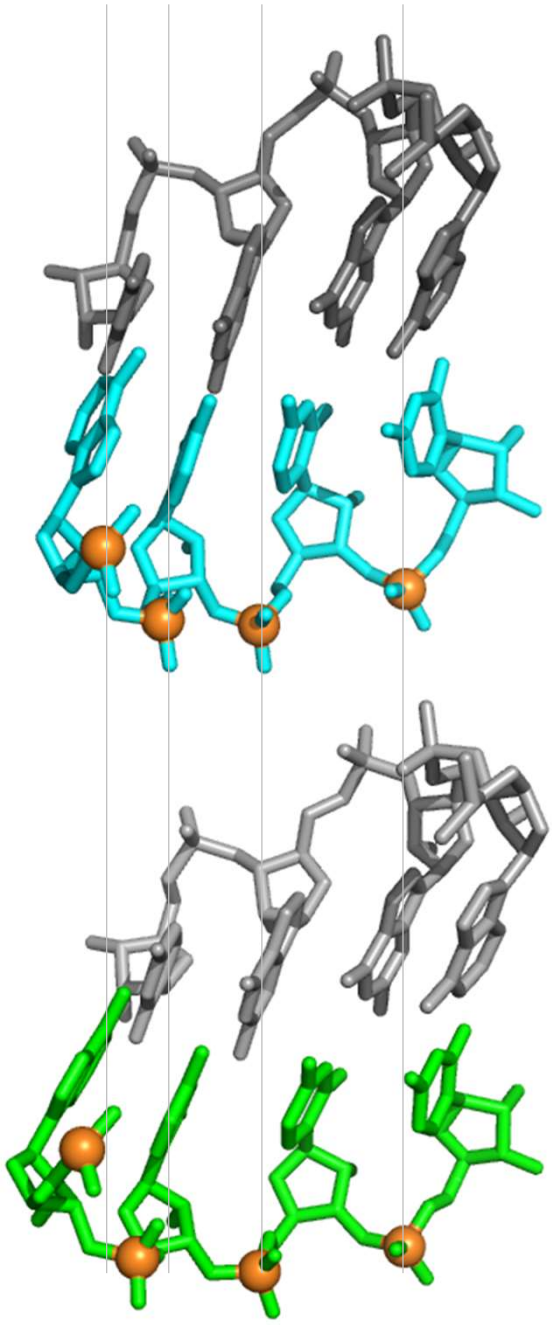

Figure 1.6: Asymmetric movement of the RNA strands during translocation by $3D^{pol}$.

Crystal structures showing the RNA template and product strands from the EV-71 elongation complex in the pre-translocated (top) and mid-translocation (bottom) conformations. The template strand is the top strand in each image shown in gray, with the product strand being denoted in color to highlight the base position, -1 to -4. Vertical lines are drawn to assist in visualizing the product strand movement relative to the template strand and backbone phosphate atoms colored orange.

serine results in a decrease in ribavirin sensitivity and, ultimately, an increase in polymerase fidelity, despite the fact it is not located within the polymerase active site while still having effects on the active site closure event prior to catalysis. Another residue, threonine 362, which is mutated in the OPV polymerase, also decrease the fidelity of $3D^{pol}$ through a role in active site closure and, like residue 64, is distant from the polymerase active site (Liu et al., 2015).

The Boehr and Cameron groups used NMR to detect molecular movements within $3D^{pol}$ at sites near and distant from the active site (Yang et al., 2012). One finding from this work is that mutations to $3D^{pol}$, such as G64S, which is involved in stabilizing the protein's N-terminus binding pocket, have been shown to affect the conformational environments of residues both in close proximity and across the polymerase, leading to the hypothesis that distant residues may be conformationally coupled together . While many interactions between residues can be identified via clearly visible electrostatic or hydrogen bonds in the crystal structures, the NMR analysis shows coupling between residues that are not able to be traced clearly to one another. These interaction networks may also have distinct functions, for example, discrimination between nucleobases or nucleosugars during nucleotide incorporation (Liu et al., 2015).

1.4 The interplay between polymerase fidelity and viral genome recombination

1.4.1 The low fidelity of $3D^{pol}$

Viral RdRPs are inherently low-fidelity polymerases with mutation rates ranging from 10^{-4} to 10^{-6} mutations per round of genome replication. This translates to 3.8 and 6.1 mutations per 10 kb of RNA replicated for coxsackievirus and poliovirus respectively, or approximately 2 to 4 mutations per viral genome replicated (Graci et al., 2012). The high error rate during viral genome replication gives rise to quasispecies formation, in which the overall viral genome

population has a consensus sequence, but individual viral genomes have small divergences in sequence (Andino and Domingo, 2015). This divergence can offer a fitness advantage to the virus, allowing it to adapt for optimal replication in many different cell types as a viral infection is propagated throughout its host (Lauring et al., 2012; Xiao et al., 2017). However, if too many mutations are accumulated, the fitness advantage is lost, yielding a virus that is no longer viable (Moratorio et al., 2017) (Gnadig et al., 2012).

Much work has been conducted to understand the mechanism by which polymerase fidelity is controlled. Ultimately it has been determined that this fidelity control is intimately linked with how viral RdRPs close their active site for catalysis, as mutations to residues in the coxsackievirus polymerase motif A results in RdRPs with lower fidelity than the wildtype polymerase, determined by deep sequencing analysis (Campagnola et al., 2015; Gnadig et al., 2012). Analogous mutations made to the poliovirus polymerase result in the opposite effect, where the polymerase fidelity is increased as compared to wildtype, suggesting that the coxsackievirus RdRP is likely at the upper limit of viral RdRP fidelity while the poliovirus polymerase resides at the lower end of the spectrum (Campagnola et al., 2015). This change in fidelity comes at a cost of polymerase replication rate, with high-fidelity polymerases becoming slower and low-fidelity polymerases having an increase in elongation rate (Fitzsimmons et al., 2018). It should also be noted that when mutations are made to the polymerase fingers domain, they have a greater effect on elongation rate than polymerase fidelity (Campagnola et al., 2015). Because a rigid-body movement of motifs A and D together is used to close the polymerase active site, mutations to motif D as well as motif A have resulted in changes to polymerase fidelity. A high-fidelity coxsackievirus polymerase is achieved through mutations to a select residue, phenylalanine 364 (residue 363 in poliovirus), found in motif D, that stabilizes the loop

at the top of motif D in an open conformation, which in turn leads to the stabilization of the active site in an open conformation. This open conformation results in slower active site closure and, in turn, higher nucleotide discrimination. It should also be noted that a planar amino acid in the 364 position is also needed to sustain viral growth (McDonald et al., 2016).

1.4.2 Exploiting the low fidelity of $3D^{pol}$ as an antiviral strategy, and how the virus combats it through genome recombination

The low fidelity of viral RdRPs allows for the incorporation of nucleotide analogues, which act as antivirals in many cases by causing an increase in mutation frequency in the viral genome, ultimately driving the virus to error catastrophe (Crotty et al., 2002). This accumulation of mutations to the viral genome drives the virus to the point of extinction. Ribavirin is a broad-spectrum antiviral nucleotide analog that can be incorporated into the viral genome across from either cytosine or uracil, causing mutations at the sites of incorporation upon further genome replication (Crotty et al., 2000). One strategy viral RdRPs have used to combat the incorporation of ribavirin is polymerase mutation to increase polymerase fidelity, as is the case with the poliovirus polymerase mutant G64S. This mutant is a high-fidelity polymerase that was discovered through multiple passages of the virus in cells in the presence of ribavirin (Pfeiffer and Kirkegaard, 2003). However, this strategy of increased polymerase fidelity can come at the cost of virus attenuation, and is therefore not a long-term approach that works for virus propagation (Kautz and Forrester, 2018).

Viruses possess a second mechanism for combating error catastrophe, genome recombination. There are multiple models for viral recombination, but the prevailing theory about how this occurs is through a copy choice mechanism (Figure 1.7) (Kirkegaard and Baltimore, 1986; Muslin et al., 2019). Briefly, this mechanism produces a progeny RNA

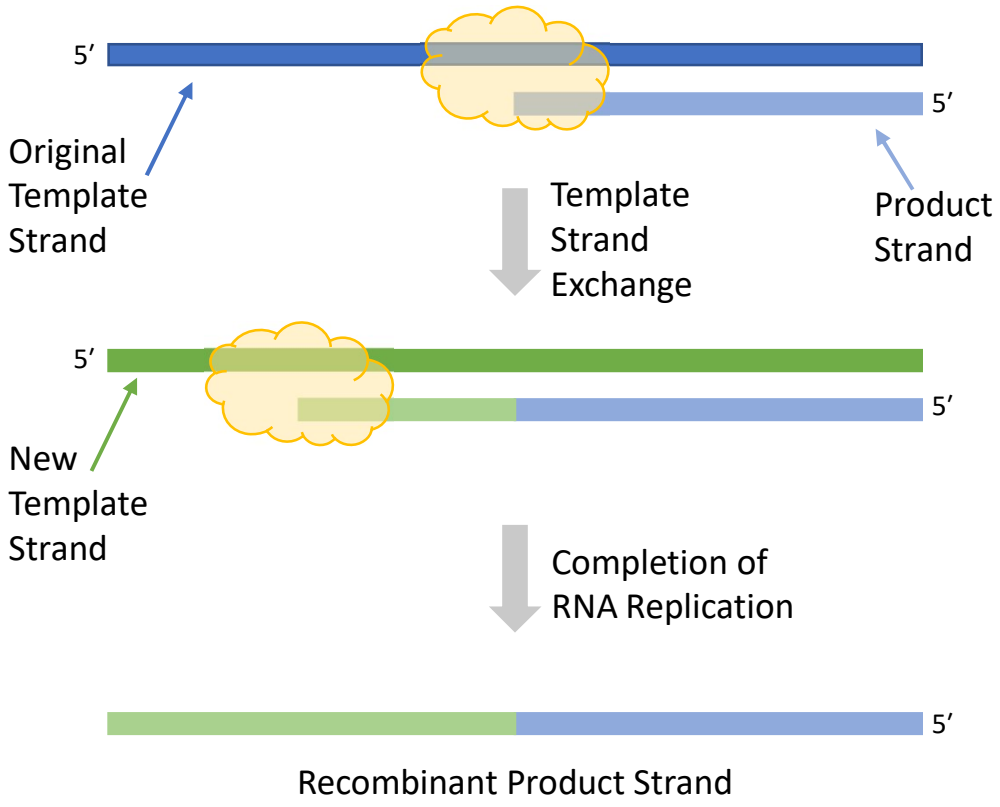

Figure 1.7: The copy choice mechanism of viral genome recombination.

Genome recombination occurs during viral replication by $3D^{pol}$. As the polymerase is replicating either the negative or positive sense RNA strands, the polymerase remains bound to the product RNA strand but dissociates from the template strand. Replication reinitiates after polymerase binding to a new template strand, ultimately producing a product RNA strand with complementary portions to each template strand.

molecule during replication that is a hybrid of two parental genomes because of the exchange of the template RNAs (parental) during replication while the polymerase remains bound to the progeny RNA for the entirety of replication. The production of a mixed-parental progeny RNA genome is classified as a form of sexual reproduction and is thought to be a strategy by which a large accumulation of mutations to the genome can quickly be purged to prevent viral extinction (Simon-Loriere and Holmes, 2011). It should be noted that this phenomenon has been observed between multiple enteroviruses of the same species, i.e. between enterovirus C members, which can ultimately result in the removal of introduced mutations to the viral genome and the reversion to an infectious virus form (Bessaud et al., 2016; McIntyre et al., 2013). Thus, a recombination event between a circulating enterovirus C family member and an inactivated poliovirus used for immunization can lead to a vaccine-derived poliovirus that is no longer attenuated and capable of spreading disease in an under vaccinated population. Currently this ability for viruses to recombine is one of the largest obstacles to a complete poliovirus eradication effort.

1.4.3 Current information about residues involved in picornaviral genome recombination

In the past few years, there has been a growing volume of work to determine the interplay between fidelity and recombination at the molecular level. This research has determined that recombination occurs in a two-step process, first by generating a hybrid genome that contains sequences from two parental virus strains, and then resolving this genome to remove any extraneous RNA sequence from the recombination event. It was also shown in this work that there appear to be no distinct recombination "hotspots" within the viral genome (Lowry et al., 2014). To date there have been three residues, aspartic acid 79, tyrosine 275, and leucine 420 (and their equivalents in other picornaviruses) within the polymerase that have been identified to

be involved in recombination in poliovirus (Acevedo et al., 2018; Kempf et al., 2016; Li et al., 2019; Woodman et al., 2018a; Xiao et al., 2016). In the residue 79 study, mutation of the aspartic acid residue to a histidine resulted in a significant loss in viral genome recombination while there was insignificant impact on polymerase fidelity (Xiao et al., 2016). As with residue 79, the mutation of tyrosine 275 to a histidine also showed impaired ability for genome recombination but showed no significant effects on polymerase elongation rate or fidelity (Acevedo et al., 2018).

Both residues 79 and 275 are located in the polymerase palm domain; residue 420, in contrast, is located in the thumb domain and unlike the other two residues it contacts the template RNA strand within the polymerase. The helix in the thumb domain, where residue 420 is located, has had multiple residues implicated in the regulation of the poly-A tail length, with the residue 420 leucine-to-alanine mutation producing some of the longest poly-A tails (Kempf et al., 2013). This helix contacts the minor groove of the template RNA as it is exiting the polymerase; mutation to residue 420 decreases the size of the hydrophobic side chain resulting in decreased interactions with the template RNA, thus weakening the polymerase's grip on the RNA as a whole, allowing for reiterative transcription to occur. This decrease in primer grip found for the leucine-to-alanine mutation also could have effects on genome recombination efficiency. There is a decrease in genome recombination observed for the residue 420 mutation, likely due to a loss of grip by the polymerase on the template strand, which may be required for the copy-choice mechanism of recombination. This loss in recombination efficiency also comes with an increase in ribavirin sensitivity, suggesting that the virus is now more susceptible to error catastrophe, despite having similar polymerase fidelity and elongation rates to that of the wildtype polymerase (Kempf et al., 2016). A study conducted using EV-71 showed comparable

findings to the poliovirus recombination study when the analogous mutation was introduced into the polymerase resulting in decreased recombination and increased ribavirin sensitivity (Woodman et al., 2018a).

1.5 Scope of the dissertation

This dissertation contains a summary of my graduate work, which has predominately focused on determining residues and polymerase domains that are involved in the final step in the catalytic cycle, translocation. During my time in graduate school, I also worked in collaboration with the Barton laboratory at the University of Colorado Anschutz Medical School investigating viral genome recombination, specifically looking at leucine residue 420, an established residue whose mutation to an alanine results in a decrease in the ability of $3D^{pol}$ to facilitate recombination. Chapter Two of this work summarizes the findings while investigating three distinct regions within the polymerase fingers domain and how mutation to residues in each of these regions affects the polymerase kinetics. Chapter Three takes a step back from individual residues and investigates the role each domain of $3D^{pol}$ plays in the overall function of the polymerase. Chapter Four gives an overview of the effects mutation to leucine 420 has on the polymerase kinetics. This chapter is a follow up to work described in Appendix A, in which is a virology study on the recombination mutants. Appendix B further expands on the work in Appendix A and Chapter Four to explore the region surrounding residue 420 in the hopes of establishing a polymerase region involved in recombination.

CHAPTER 2

FINGERS DOMAIN INTERACTION NETWORKS CONTRIBUTE TO DIFFERENT STEPS IN THE 3D^{POL} CATALYTIC CYCLE

Authors contributing to the chapter: X-ray diffraction data collection and solving of crystal structures in this chapter were solved by Ryan Czarny, with help from Grace Campagnola, during his rotation in the laboratory. Initial cloning, protein purification, enzyme kinetics, and crystallization of coxsackievirus crystals were done by me. Tables and figures in the chapter were also completed by me.

2.1 Introduction

The virus family *Picornaviridae* is comprised of single-stranded positive-sense RNA viruses whose ≈7.5kb genome encodes for a single polyprotein. After translation, the polyprotein is cleaved into approximately a dozen individual proteins by multiple virally encoded proteases. The polyprotein encompasses three functional regions: the P1 region containing the viral capsid proteins, the P2 region functioning in membrane rearrangement during viral replication, and the P3 region whose proteins are involved in viral RNA replication. The final protein in the P3 region is 3D^{pol}, an RNA-dependent RNA polymerase (RdRP). While the viral RdRPs share a conserved overall structure of a cupped right hand with other single subunit polymerases, they contain a unique fingers domain interaction the top of the thumb domain that is not seen in other polymerases (Ferrer-Orta et al., 2006). This interaction has led to adaptations by the viral RdRPs as to how they carry out their nucleic acid addition cycles (Arnold and Cameron, 2004; Arnold et al., 2004; Gohara et al., 2004; Gong and Peersen, 2010; Gong et al., 2013).

In general, replicative polymerases such as bacteriophage T7 RNA polymerase or *E. coli* DNA polymerase I rely on large swinging motions within the fingers domain to not only transition the incoming NTP from a pre-insertion site into the active site for catalysis, but also for additional movements to reopen the active site and translocate the nucleic acid template and product strands to allow for replication to continue (Yin and Steitz, 2004). Because of the contact between the fingers and thumb domains found in viral RdRPs, the large fingers domain movements are prevented. This has led viral RdRPs to have evolved a unique active site closure mechanism (Gong and Peersen, 2010; Gong et al., 2013). This mechanism relies on subtle movements motifs A and D within the palm domain to move motif A into a position to form a β-sheet with motif C. This movement brings the required aspartic acids and Mg^{2+} ion into the correct geometry to allow for incoming nucleotide catalysis. These movements are then reversed to reopen the active site and allow for RNA translocation. While the complete translocation step has not been elucidated, it is proposed to not be tightly coupled to active site reopening, and to occur via an asynchronous movement of the RNA stands (Shu and Gong, 2016), with the product strand moving by about a half-base step before the template strand, and additional residues 288-291 found in motif B may aid in template strand translocation (Sholders and Peersen, 2014).

In the work presented here we investigated the roles three other structural elements located within the fingers domain of $3D^{pol}$ play in the nucleic acid addition cycle (Figure 2.1A). These elements include the kink region (Figure 2.1B), located in the index finger, the gateway region (Figure 2.1C), which is part of motif G found in the pinky finger, and the sensor region (Figure 2.1D) contained in the ring finger as part of motif F. The findings of this investigation show that the kink mutants assist in binding to the template RNA strand and elongation complex maintenance, the gateway mutants act as a catch for the template RNA strand for positioning

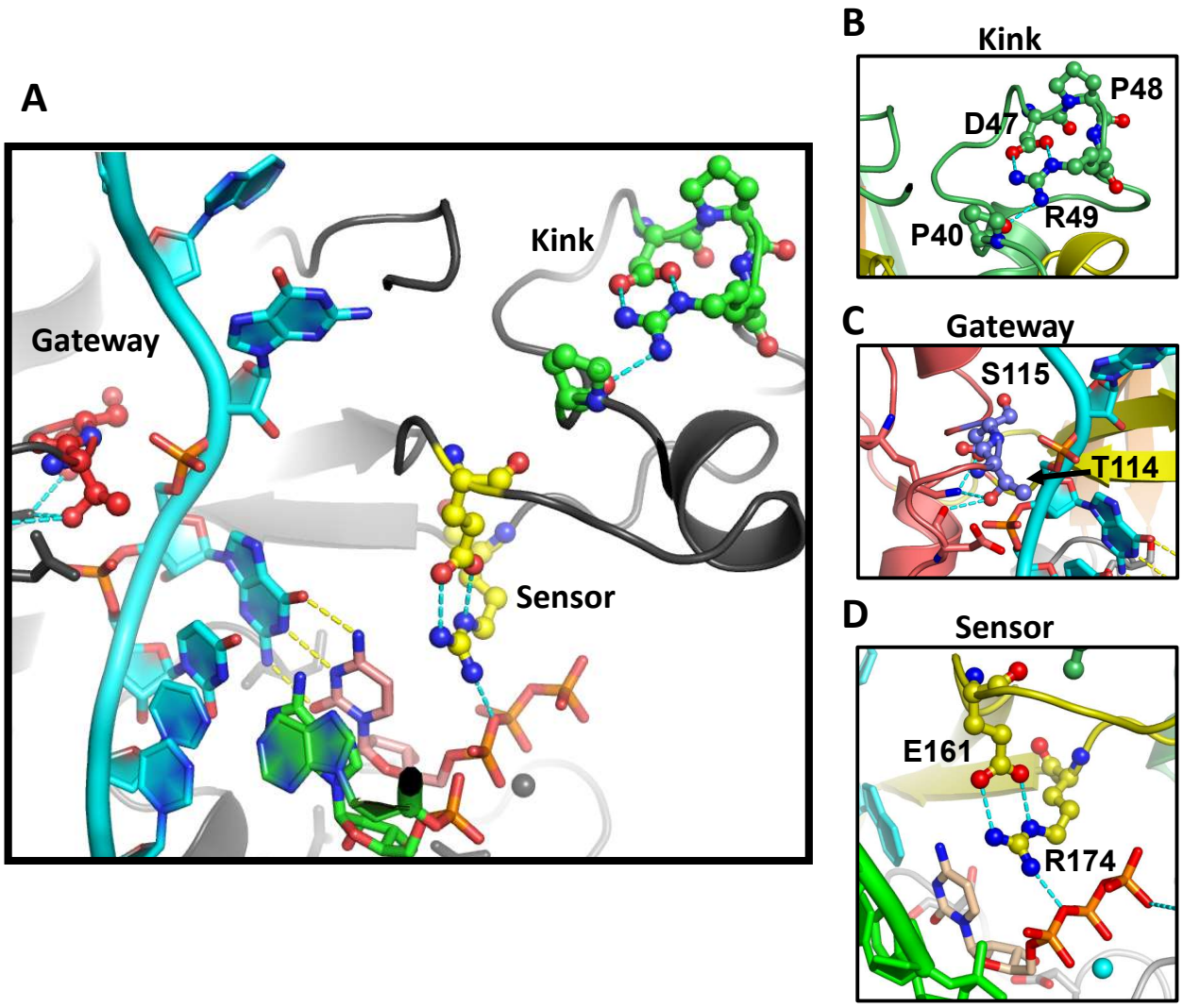

Figure 2.1: Targeted mutations broken down by polymerase region.

(A) Three regions within the polymerase fingers domain were targeted for investigation, all within close proximity to the polymerase active site. (B) The kink region is located within the index finger of $3D^{pol}$ and contains four interacting residues: P40, D47, P48, and R49. These residues form a kink structure through a series of denoted hydrogen bonds. (C) The gateway region interacts with the template strand RNA by positioning the methyl group side chain of T114 between the -1 and -2 bases phosphate backbone. This residue positioning is held in place through hydrogen bonding with residue S115. (D) The sensor region investigated contained two residues E161 and R174. These residues are found on opposite beta strands within the polymerase ring finger. Hydrogen bonding between E161 and R174 places R174 directly above the incoming NTP, positioning it to interact with the α-phosphate.

post translocation, and the sensor mutants serve to aid in both incoming nucleotide capture as a final fidelity checkpoint and chemistry on the incoming nucleotide. It is also clear from these findings that none of the above mentioned structural elements act completely independently from each other, suggesting that instead the fingers domain in its entirety acts a one functional unit through an intra-domain network to assist in the nucleic acid addition cycle.

2.2 Methods

2.2.1 Polymerase mutations

Three regions within the polymerase fingers domain were targeted for investigation, Figure 2.1A. Each region contained between two and four residues that were mutated, Figure 2.1B. The mutated polymerase genes were assembled into a T7-based expression vector with the In-Fusion cloning kit (Clontech) using a mutagenesis PCR protocol for each individual mutation, and the final constructs were verified by sequencing. Polymerase expression and purification were carried out as previously described (Hobdey et al., 2010) with the final purification step being size exclusion chromatography into a 200 mM NaCl buffer.

2.2.2 Initiation and stability

Initiation rates were assayed using 5 uM polymerase, 0.5 uM "10+1_12" RNA (Hobdey et al., 2010) and 40 uM GTP at room temperature in 50 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM Tris-(2-carboxyethyl)-phosphine (TCEP). 1 uL samples were removed at various time points up to 15 minutes and added to 19 uL of quench buffer containing 50 mM EDTA, 400 mM NaCl, 50 mM HEPES pH 6.5, and 2 mM TCEP. After the initiation reaction had proceeded for 15 minutes, the assay for the temporal stability of the resulting elongation complexes was

begun by adding 10 uL of the reaction to 90 uL of 300 mM NaCl, 4 mM $MgCl_2$, 50 mM HEPES pH 6.5, and 2 mM TCEP, wherein the high salt concentration limited further initiation or re-initiation by preventing RNA binding. The amount of competent elongation complex still present in this solution was tested at various points over a four-hour period by removing 5 uL aliquots and mixing them with 5 uL buffer containing 160 uM each of ATP, GTP, and UTP, followed by a two minute elongation reaction to produce full-length products before quenching with EDTA. Samples were analyzed by denaturing gel electrophoresis using 20% 19:1 polyacrylamide, 7M urea, 1X TBE gels. RNA bands were detected using a Li-Cor Odyssey 9120 infrared imager system to visualize the IRdye 800RS (Li-COR Biosciences) label on the RNA tetraloop, as previously described [23]. Single exponential equations were mathematically fit to the data using Kaleidagraph (Synergy Software) to determine time constants for the increased formation of +1 product in the initiation assay or the decreased formation of +7 chase product in the stability assay.

2.2.3 Enzyme kinetics assays

Nucleotide incorporation kinetics were determined using rapid stopped-flow kinetics methods previously described for both processive elongation (Gong et al., 2009) and single nucleotide incorporation assays (McDonald et al., 2016). Stalled elongation complexes were first generated by incubating 15 uM polymerase with 10 uM RNA in 50 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM TCEP, and 60 uM each of ATP and GTP for 15 minutes. The resulting complexes were diluted 200-fold to a final RNA concentration of 50 nM with 50 mM HEPES pH 7, 75 mM NaCl, and 4 mM $MgCl_2$, generating the sample that was loaded into the stopped-flow instrument, which was further diluted 2-fold when mixed with either CTP only or all NTPs in the reaction cell (i.e. 25 nM final). $MgCl_2$ was always maintained at 4 mM excess

over the total NTP concentration, and for the processive elongation assays the four NTP were present at equimolar concentrations. Kinetics data were collected at 30°C using a Bio-Logic SFM-4000 titrating stopped-flow instrument with a MOS-500 spectrometer. Single CMP incorporation and processive elongation reactions were fit using Kaleidagraph (Synergy Software), and double CMP reactions were modeled using KinTek Explorer (Johnson et al., 2009b, 2009a).

Rapid chemical quench experiments were done by first forming stalled elongation complexes with 15 uM polymerase and 10 uM 5′-fluorescein labeled RNA for 15 minutes at room temperature in buffer containing 75 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM TCEP, 200 uM ATP, and 60 uM GTP. The resulting complexes were diluted 20-fold to a final concentration of 500 nM (RNA with reaction buffer consisting of 50 mM HEPES pH 7, 50 mM NaCl, and 4 mM $MgCl_2$. Polymerase complexes were further diluted 2-fold when mixed with various concentrations of CTP, the single nucleotide substrate, were made in this same buffer. Quench-flow reactions were done in a Bio-Logic QFM-4000 instrument that uses variable rate continuous flow through a 3 uL reaction tube that is followed by mixing with the quench solution. Stalled elongation complex and CTP were mixed, allowed to react for 5 ms to 2 seconds in the aging loop before mixing with a quench solution of either 300 mM EDTA (100 mM final) or 3 M HCl (1M final). A total of 15 ul of each component was used in the reaction, yielding 45 uL final sample volumes. For the acid quench, 30 uL of each quenched sample were manually transferred to a secondary collection vial which had been preloaded with 30 ul of 1 M NaOH, 300 mM HEPES, pH 7 to neutralize the samples. Reaction products were analyzed by denaturing gel electrophoresis on 20% polyacrylamide (19:1), 7M urea, 1X TBE gels, and imaged using the 488 nm channel of an Amersham Typhoon Imager. Product formation rates

were obtained at eight different CTP concentrations using 8-12 timepoints in each quench experiment, and gel bands were quantified using the program PeakFit (Systat Software, Inc.) to analyze lane scans, fitting all the peaks in each lane with a single fixed peak width Gaussian across each lane to improve quantitation of weak bands.

Reactions that followed a concentration dependence used a variation on the conventional concentration dependence equation:

$$v = \frac{k_{cat}[S]}{K_M + [S]}$$

Instead the following equation was used:

$$v = \frac{k_{sp}[S]}{1 + (k_{sp}[S]/k_{cat})}$$

Where:

$$k_{sp} = \frac{k_{cat}}{K_M}$$

Unlike the conventional equation, the second equation directly fits the reaction rate and specificity. Rection specificity, k_{sp}, sets the lower limit for the reaction rate as it is a second order rate constant defined by the maximal rate of the reaction times the probability of a successful substrate binding event going forward in the reaction, and ultimately is an inherent property of the enzyme. By mathematically fitting the lower and upper bounds of the enzymatic reaction the determined errors are generally smaller for each parameter. When K_M is then derived, the propagated error associated with it also is generally less than if K_M were directly determined.

2.2.4 Protein crystallography

The structures were solved by molecular replacement using Phenix software and the wildtype CV 3Dpol structure (PDB entry 3DDK). The point mutations at residues 161 and 174 were introduced into the molecular structures and the models were passed through several cycles of standard Phenix refinement. In general, the lysine residue in the R174K mutations was not found to have strong electron density, suggesting this residue is fairly flexible, but the E161Q mutant did show density comparable to the nearby backbone.

2.3 Results

2.3.1 Mutated polymerase purification

Each mutated poliovirus polymerase expressed well in *E. coli* and could be purified using a series of Ni-affinity, anion exchange, and size exclusion chromatographies and finally was concentrated to ~250 uM in 200 mM NaCl buffer. These mutated polymerases can be categorized into three groups: the kink, the gateway, and the sensor mutations. All groups are located within the polymerase fingers domain. However, the coxsackievirus polymerases were less stable and were prone to aggregation in lower salt conditions. For this reason, final protein concentrations were generally lower at ~100 uM and required a 300 mM NaCl buffer to prevent precipitation.

2.3.2 Initiation rates and elongation complex stability

Rates for initiation were determined by +1 product accumulation over time as an elongation complex formation and first nucleotide incorporation. These reactions were monitored over a 15-minute period and fit to a single exponential equation to determine the

initiation time constant. Initiated reactions were then diluted into a high-salt buffer to prevent re-initiation by dissociated polymerases to determine the elongation complex lifetime (Figure 2.2). In general, the kink mutants, with the exception of P48G, showed an increased initiation time and decreased elongation complex (EC) stability (Table 2.1). The lack of effect by P48G may be due to no interactions between that residue and the others within the kink region (Figure 2.1B). The gateway mutants showed a similar pattern to the kink mutants of an increased initiation time coupled with a decrease in EC stability. In this case however, this tended to correlated with residue side chain size, with T114A and T114S having shorter initiation times and more stable ECs than T114L or T114V. T114S in fact had a shorter initiation time and more stable complex than T114A even though T114A possesses a smaller side chain, likely this is due to a preservation of a hydrogen bond between S115 and T114 (Figure 2.1C). As for the sensor residues, only E161Q containing mutants deviated from WT with an increase in initiation time, and all mutants showed a decrease in EC stability.

2.3.3 Rapid kinetics studies of processive elongation and nucleotide discrimination

In order to determine processive elongation rates for each polymerase, preformed elongation complexes were reacted with varying concentrations of NTPs and monitored via fluorescence change from the polymerase's approach to the 5′-end fluorescein on the hairpin RNA construct (Figure 2.3A and 2.3B). Processive elongation rate for each NTP concentration, determined as the number of incorporated NTPs divided by lag time prior to the increase in fluorescence, was graphed against reaction NTP concentration to determine the processive elongation concentration dependence (Figure 2.3C). As with the initiation and stability experiments, P48G did not differ from WT while the other three kink mutations all decreased in processive elongation rate (Table 2.1). Also of note, P40G and R49K both showed a two-fold

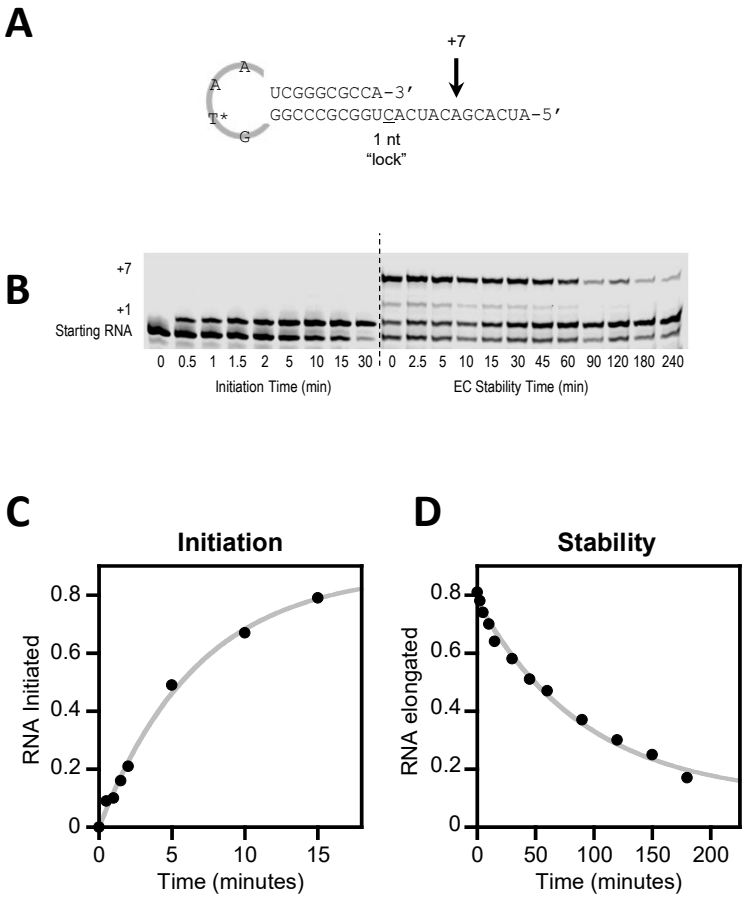

Figure 2.2: Initiation and elongation complex stability determination.

(A) PETE 10+1_12 RNA used for initiation and EC lifetime determination, denoted are a one nucleotide lock to form the EC, and the +7 position to which the polymerase extends to with NTP addition. (B) Example gel image showing accumulation of +1 product during the initiation reaction stage, and +7 product formation to test EC lifetime. +1 product formation is fit to a single exponential increase equation (C), decrease in +7 formation is fit to a single decreasing exponential (D).

Table 2.1: Summary of kinetics data for Kink, Gateway, and Sensor Mutants.

$3D^{pol}$	Initiation (min)	EC Stability (min)	Single Nucleotide k_{pol} (nt/sec)	Single Nucleotide K_M (μM)	Processive Elongation k_{pol} (nt/sec)	Processive Elongation K_M (μM)	Discrimination Factor
PV WT	4 ± 1	130 ± 20	25 ± 1	20 ± 2	20 ± 1	49 ± 4	120 ± 10
Kink Mutants							
P40G	13 ± 1	29 ± 2	22 ± 1	37 ± 3	13 ± 1	100 ± 10	170 ± 40
D47N	15 ± 1	19 ± 2	20 ± 1	46 ± 6	15 ± 1	55 ± 7	150 ± 30
P48G	6 ± 1	100 ± 20	27 ± 1	23 ± 3	20 ± 1	52 ± 4	130 ± 10
R49K	19 ± 2	29 ± 4	17 ± 2	40 ± 10	11 ± 1	90 ± 20	140 ± 20
Gateway Mutants							
T114A	10 ± 2	57 ± 10	25 ± 2	31 ± 5	14 ± 1	63 ± 6	100 ± 10
T114L	33 ± 7	5 ± 1	26 ± 3	90 ± 20	10 ± 1	100 ± 10	130 ± 10
T114S	4 ± 1	82 ± 9	26 ± 1	23 ± 1	20 ± 1	52 ± 4	110 ± 10
T114V	18 ± 4	9 ± 2	27 ± 6	150 ± 40	5 ± 1	80 ± 40	100 ± 10
S115A	5 ± 1	90 ± 20	28 ± 2	60 ± 7	11 ± 1	95 ± 6	160 ± 10
S115L	17 ± 3	15 ± 1	19 ± 4	110 ± 30	5 ± 1	30 ± 10	150 ± 20
S115T	5 ± 1	13 ± 1	23 ± 2	61 ± 7	8 ± 1	120 ± 10	130 ± 20
S115V	18 ± 3	25 ± 2	29 ± 1	38 ± 3	16 ± 1	74 ± 4	170 ± 10
Sensor Mutants							
E161D	5 ± 1	70 ± 20	24 ± 1	19 ± 2	10 ± 1	32 ± 2	100 ± 10
E161Q	10 ± 3	80 ± 20	22 ± 1	9 ± 1	17 ± 1	19 ± 2	38 ± 6
R174K	3 ± 1	60 ± 10	2 ± 1	69 ± 7	1 ± 1	39 ± 3	12 ± 2
E161Q + R174K	8 ± 1	80 ± 20	5 ± 1	16 ± 2	6 ± 1	34 ± 3	90 ± 10

increase in K_M compared to WT. We found the processive elongation rates for the gateway mutants decreased with respect to side chain size. The S115 mutants within the gateway group show more changes in K_M than the T114 mutants. The sensor mutants all had a slower processive elongation rate than that of WT, with the decrease in rate being dependent on which residue was mutated. Mutations to residue 161 had less impact on elongation rate than the R174K mutation. The E161D and E161Q polymerases showed only a slight decrease in processive elongation rate, while the R174K containing polymerases had a significant rate decrease. The E161Q R174K mutant had an intermediate rate relative to the polymerases containing each mutation alone. This observation likely points to a direct interaction with R174 that is perturbed by the lysine mutation but can be partly rescued by the additional E161Q mutation.

To further investigate the rate changes observed for processive elongation, single nucleotide incorporation rates were also determined. Rates were measured via fluorescence changes of a 2-aminopurine molecule located in the downstream template RNA (Figure 2.3D). Briefly, after adding the incoming NTP to the growing product strand RNA, the RNA is translocated to reset the active site for the next NTP addition. The RNA construct used for this assay places 2-aminopurine in an unstacked high fluorescent position after EC formation. After template strand translocation the 2-aminopurine is then stacked on to the product RNA strand, quenching its fluorescence. This decrease in fluorescence was determined for multiple nucleotide concentrations (Figure 2.3E), and an NTP concentration dependence was determined (Figure 2.3F). The kink mutants had a decrease in rate for both the single nucleotide incorporation reactions and their respective processive elongation rates. However, the gateway mutants showed almost no change in single nucleotide incorporation rates, only changes in K_M that correlated

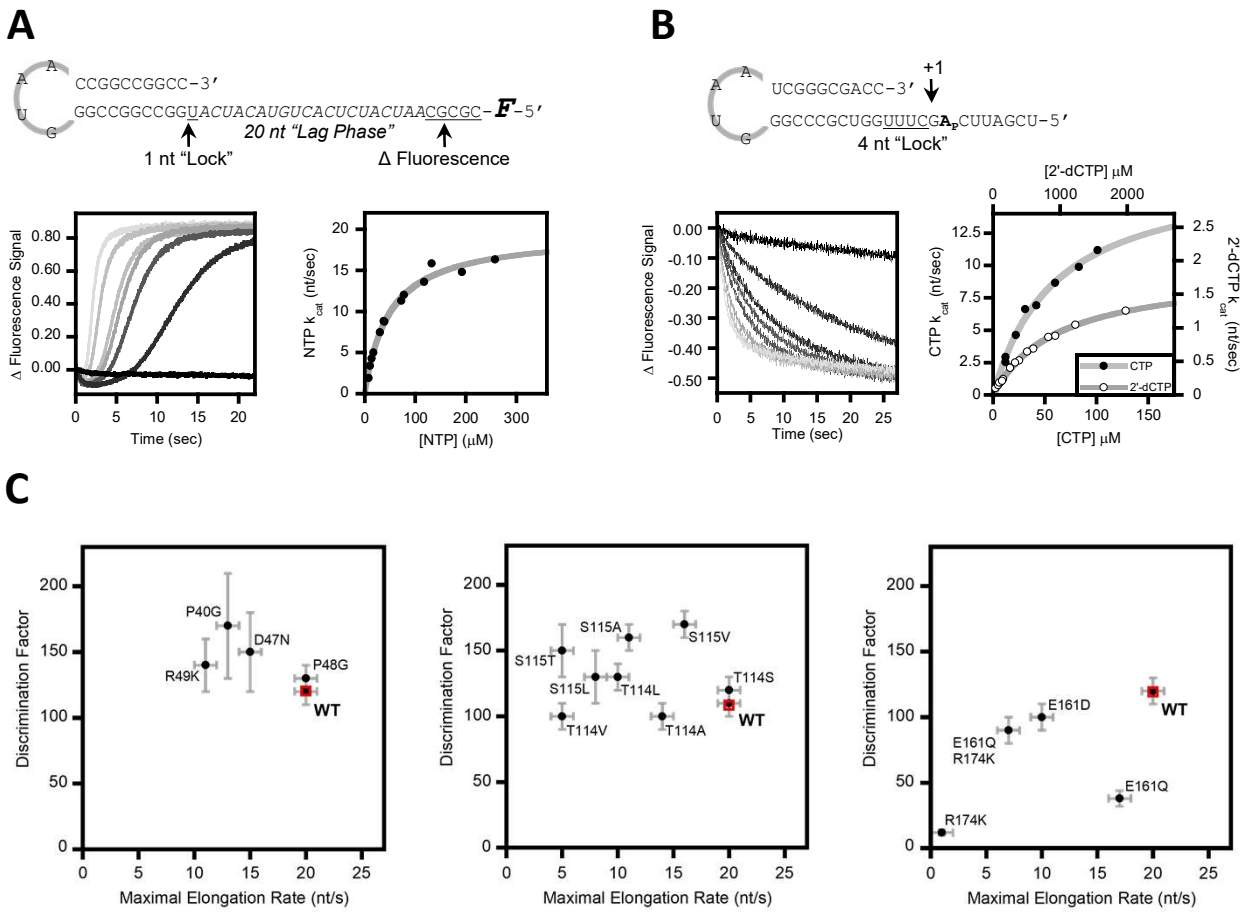

Figure 2.3: Discrimination factor and processive elongation analysis.

(A) PETE 10+1_25 RNA used to determine processive elongation rates. Denoted are single NTP lock sequence on RNA to form EC. The 5′-fluorscein molecule shows an environmental change in fluorescence five nucleotides before the polymerase reaches the 5′-end after having gone through the 20-nucleotide lag phase. Example processive elongation fluorescence traces at multiple concentrations, left, with the determined concentration dependence curve, right. (B) PETE RNA 10+3/4_2Ap used for single nucleotide incorporation rates and nucleotide discrimination factor determination. Denoted are the four-nucleotide lock to from the EC. An internal 2-aminopurine was used to show a change in fluorescence after CMP incorporation and RNA translocation. Example single CMP incorporation fluorescence traces at multiple concentrations, left, with the determined concentration dependence curve, right. (C) Processive elongation rate verses nucleotide discrimination factor. Left, kink mutants; center, gateway mutants; right, sensor mutants. WT polymerase highlighted in red.

with residue side chain size. This discrepancy between rates for a single NTP incorporation and multiple incorporations points to a step after the completion of an NTP addition prior to the start of another round of incorporation that has an effect on processive elongation rate and is directly proportional to the side chain size of residue 114. As for the sensor mutants, they followed the same rate changes that were observed for processive elongation.

Single nucleotide incorporation experiments were also conducted using 2′-dCTP in place of CTP to determine the nucleotide discrimination factor. Reactions were conducted as described above and nucleotide discrimination was determined by dividing the CTP incorporation catalytic efficiency, k_{cat}/K_M, by the 2′-dCTP incorporation reaction catalytic efficiency. Previous work from our laboratory has predicted that mutations within the fingers domain generally have little effect on nucleotide discrimination factor. While all the kink and gateway mutants did not have any significant discrimination factor deviations from that of WT, in line with our previous results, the sensor mutants of E161Q and R174K individually did have decreased nucleotide discrimination factor values. It should also be noted that when processive elongation rate is plotted against discrimination factor, mutations that result in increases in processive elongation rate generally do so at the expense of nucleotide discrimination and vice versa. The two sensor mutations that caused changes to nucleotide discrimination also decreased processive elongation rate. This finding suggests that sensor residues affect these two parameters independently, even though elongation rate and discrimination are tightly coupled.

2.3.4 Two-cycle elongation assays

To further dissect the changes in elongation rate observed for the T114 gateway mutants, a two CMP incorporation assay was conducted. A stalled elongation complex was formed by incubating each mutant polymerase with the hairpin RNA construct shown in Figure 2.4A, and

once formed, an internal 2-aminopurine (2AP) is placed in the +3 template RNA position. Upon addition of CTP two sequential incorporation events transition the 2AP from +3 to +2 to +1, resulting in an increase and then decrease in fluorescence, and varies in rate depending on CTP concentration (Figure 2.4B). Three titration series were simultaneously modeled in Kintek Explorer using the reaction shown in Figure 2.4A. This analysis established a rate for each CTP incorporation for WT at 40±1 s^{-1}, K_D of 8.4±0.2 uM for the first CTP and 56±2 s^{-1}, K_D of 27±1 uM for the second CTP, and a reposition step between the first and second catalytic cycles of 10.8±0.12 s^{-1} (Table 2.2). Mutation of T114 to a serine resulted in essentially no change in K_D, however the rate decreased slightly for the first CTP and increased dramatically for the second rate. The repositioning step also decreased to 8.7±0.09 s^{-1}, likely it is this decreased positioning rate that results in the decreased processive elongation rate. The alanine mutation resulted in a slightly faster polymerase rate of 31±1 s^{-1} for the first event and an increase to 69±2 s^{-1}, with only a slight change in CTP K_D for either binding event. The nucleotide repositioning rate was further decreased to 5.8±0.07 s^{-1}, which even with the increased reaction rate still would probably cause the observed decrease in processive elongation rate, more noticeably than that of T114S. The T114L mutation did not vary much from WT at a first reaction rate of 44±3 s^{-1}, with a faster second rate of 75±5 s^{-1}, there was however a significantly higher first CTP K_D at 163±9.8 uM but a comparable second K_D and slower repositioning rate at 4.2±0.05 s^{-1}. A combination of the high initial K_D and slow repositioning are strong candidates for the observed slow processive elongation rate despite the faster second incorporation rate. The final T114V mutant had the fastest incorporation rate when compared to WT at 50±5 s^{-1}, and 75±3 s^{-1} for the second event. This faster rate is countered by a slower reposition rate of 7.3±0.22 s^{-1}. Although this reposition rate is not as slow as either T114A or T114L, the K_D for CTP for the first and second events are

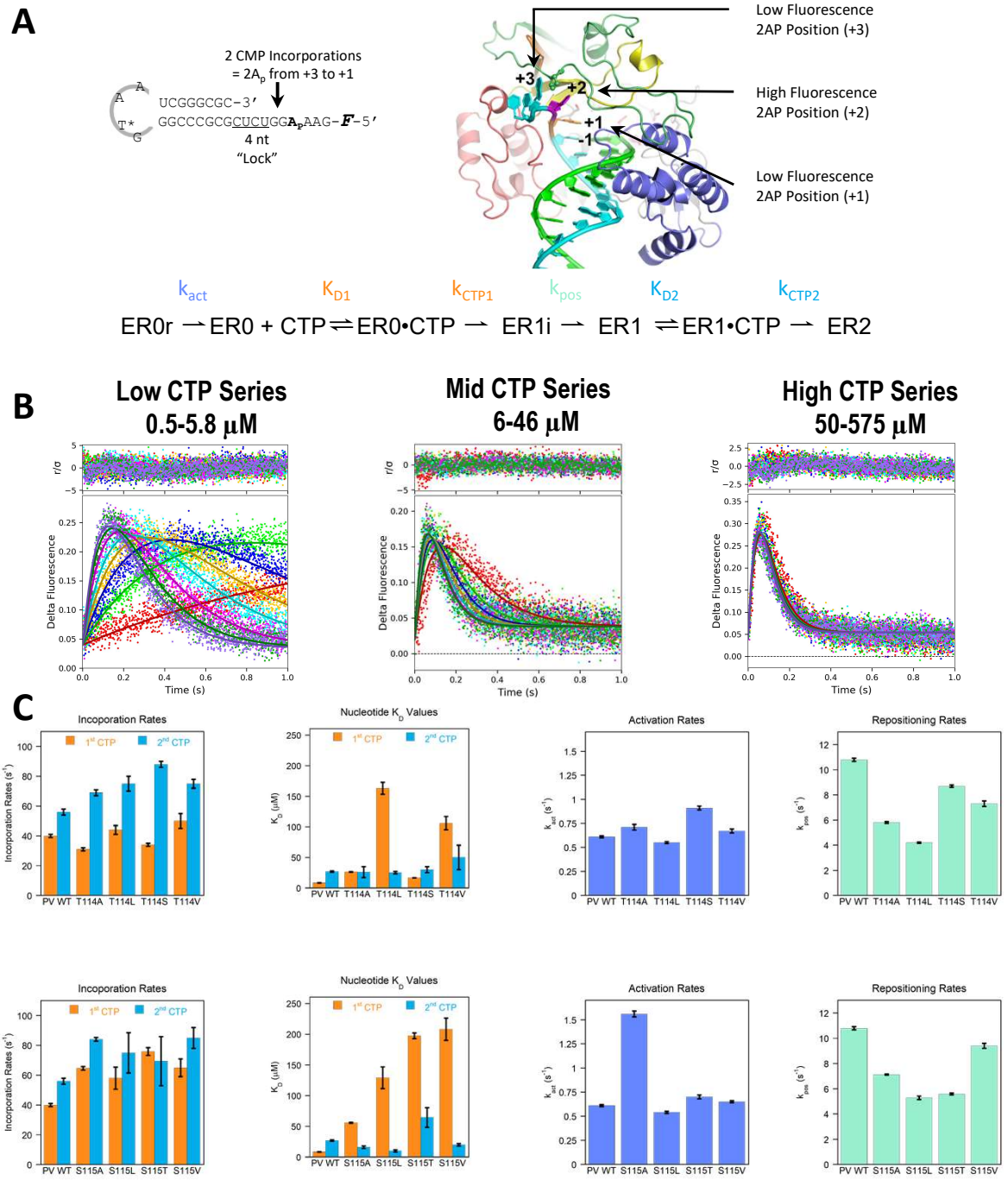

Figure 2.4: Double CMP incorporation.

(A) PETE 8+4_2_2Ap_3 RNA construct used for double CMP incorporation assay. Noted are the four-nucleotide lock to form the EC. Two CMP incorporations result in a fluorescence change from the internal 2-aminopurine with each translocation event, detailed on polymerase EC structure. Finally, the kinetic model used for fitting is shown with approximate polymerase starting populations and all kinetic parameters fit highlighted. (B) Fluorescence traces for three CTP concentration series fit to the kinetic model denoted in A. (C) Bar graphs showing rates and CMP affinities determined for each polymerase.

Table 2.2: Double CMP incorporation kinetics summary table.

Note only the two higher concentration series were used in fitting T114L and T114V due to the high K_D values causing no detectable signal at lower concentrations.

$3D^{pol}$	% EC in Active Form (ER0)	Polymerase Activation (s^{-1})	1st CMP Incorporation		NTP Reposition/ Transition (s^{-1})	2nd CMP Incorporation	
			k_{cat} (nt/s)	K_D (μM)		k_{cat} (nt/s)	K_D (μM)
PV WT	75	0.61 ± 0.01	40 ± 1	8.4 ± 0.2	10.8 ± 0.12	56 ± 2	27 ± 1
T114A	65	0.71 ± 0.03	31 ± 1	26.3 ± 0.5	5.8 ± 0.07	69 ± 2	26 ± 9
T114L	55	0.55 ± 0.01	44 ± 3	163 ± 9.8	4.2 ± 0.05	75 ± 5	25 ± 2
T114S	60	0.91 ± 0.02	34 ± 1	16.7 ± 0.2	8.7 ± 0.09	88 ± 2	30 ± 5
T114V	60	0.67 ± 0.02	50 ± 5	106 ± 11	7.3 ± 0.22	75 ± 3	50 ± 20
S115A	60	1.56 ± 0.03	65 ± 1	55.8 ± 0.7	7.1 ± 0.04	84 ± 12	16 ± 2
S115L	50	0.54 ± 0.01	58 ± 7	129 ± 18	5.3 ± 0.13	75 ± 14	10 ± 2
S115T	50	0.66 ± 0.02	76 ± 3	198 ± 5	5.6 ± 0.07	69 ± 17	64 ± 16
S115V	50	0.65 ± 0.01	65 ± 6	208 ± 18	9.4 ± 0.2	85 ± 7	20 ± 2

106±11 uM and 50±20 uM, respectively. The combination of weaker affinity coupled with slower repositioning is a likely why this mutant is observed to have the slowest processive elongation rate.

2.3.5 Rapid quench kinetics

To further investigate the observed changes in single nucleotide incorporation rate and nucleotide discrimination with the sensor mutants a series of rapid chemical quench experiments were conducted. Each reaction was conducted using the same RNA hairpin construct (Figure 2.5A) and either quenched with EDTA or HCl before polyacrylamide gel analysis (Figure 2.5B and 2.5C), or monitored for completion using 2-aminopurine fluorescence (Figure 2.5D) over a range of CTP concentrations. Concentration dependence for nucleotide capture, active site closure and chemistry, and catalytic cycle completion were used to determine the influence of each sensor mutation on the catalytic cycle (Figure 2.5E-2.5G, Table 2.3). As previously seen, the only significant differences in K_M were observed for the R174K sensor mutation, which showed almost a two-fold increase in K_M value when compared to WT. The rate of nucleotide capture for WT is 52±4 s^{-1}, and while the E161Q mutation did not have an effect on nucleotide capture, the E161D mutation increased the capture rate to almost 80 s^{-1}. It is possible that by shortening the side chain of residue 161 a steric constraint that was present in the WT polymerase was released. Adding the R174K mutation to the E161Q mutation slowed nucleotide capture to 21±1 s^{-1}, and the R174K mutation alone had the slowest nucleotide capture rate of 7±1 s^{-1}. This decrease in capture rate that was only observed in polymerases containing the R174K mutations suggest that of the two sensor residues R174K is the dominant residue in nucleotide capture but the removal of an electrostatic interaction between these residues can partially compensate for the lysine mutation. All the sensor mutants showed a decrease in active site

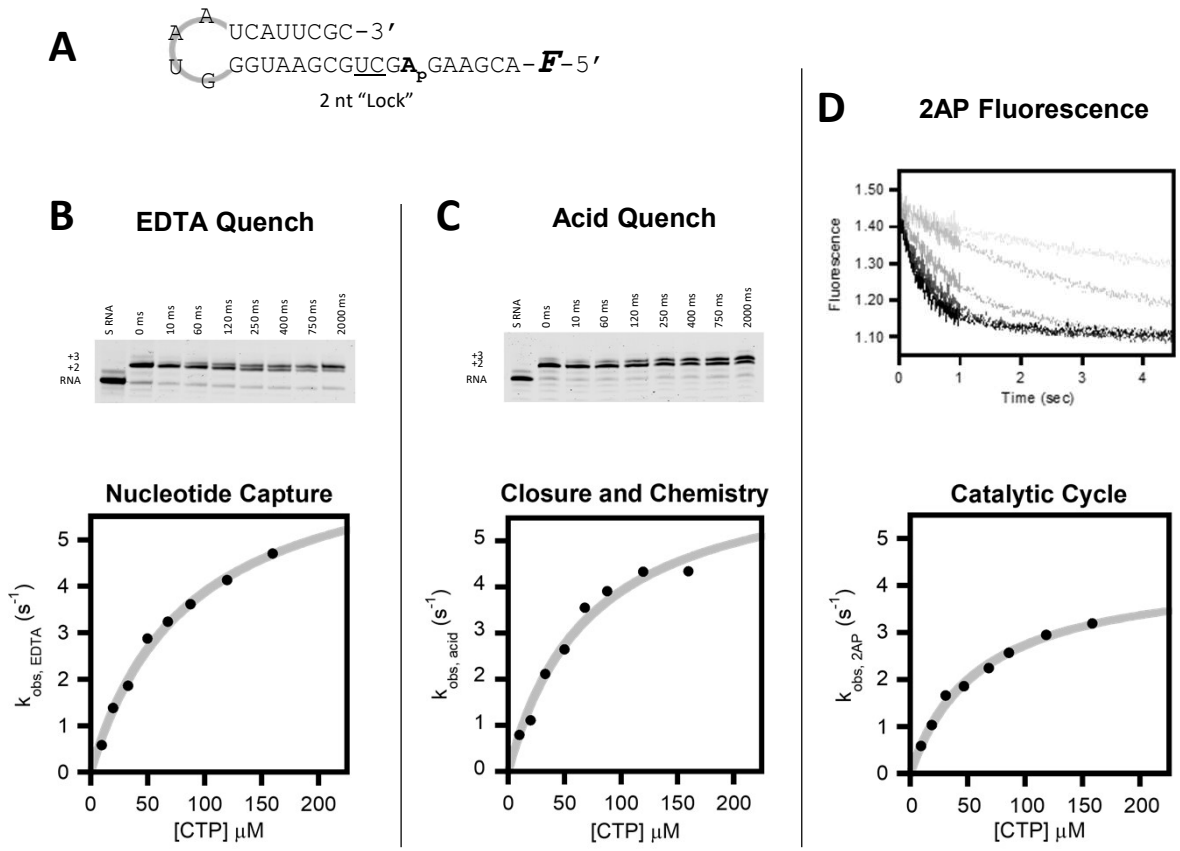

Figure 2.5: Rapid chemical quench and stopped-flow analysis.

(A) PETE 8+2_3_5_2Ap RNA construct used for rapid chemical quench analysis. Denoted are the two-nucleotide lock, an internal 2-aminopurine used for stopped-flow analysis to determine the catalytic cycle completion rate, and a 5′-end fluorescein as a marker for gel analysis of rapid chemical quench samples. (B) Representative data collected for EDTA quench to detect nucleotide capture, (C) acid quench for determination of active site closure and chemistry, and (D) fluorescence traces at multiple nucleotide concentrations to determine the rate for catalytic cycle completion by translocation. Concentration dependence plots for each reaction type are shown. A hyperbolic concentration dependence equation was fit to the data to determine maximal rate and catalytic efficiency.

Table 2.3: Catalytic cycle kinetics summary table.

Values for K_M were derived by dividing the maximal rate by catalytic efficiency.

$3D^{pol}$	Nucleotide Capture			Active Site Closure and Chemistry			Complete Catalytic Cycle		
	Catalytic Efficiency (μM^{-1} s^{-1})	k_{EDTA} (s^{-1})	K_M (μM)	Catalytic Efficiency (μM^{-1} s^{-1})	k_{ACID} (s^{-1})	K_M (μM)	Catalytic Efficiency (μM^{-1} s^{-1})	k_{2AP} (s^{-1})	K_M (μM)
PV WT	0.85 ± 0.07	52 ± 4	65 ± 8	0.9 ± 0.1	37 ± 2	41 ± 6	1.0 ± 0.1	35 ± 2	34 ± 5
E161D	2.1 ± 0.2	78 ± 4	38 ± 4	1.1 ± 0.2	25 ± 2	24 ± 5	0.53 ± 0.05	26 ± 2	50 ± 8
E161Q	2.1 ± 0.4	48 ± 4	23 ± 5	1.8 ± 0.1	25 ± 1	14 ± 1	1.3 ± 0.1	28 ± 2	22 ± 3
R174K	0.084 ± 0.006	7 ± 1	86 ± 8	0.093 ± 0.01	7 ± 1	70 ± 10	0.073 ± 0.004	4 ± 1	60 ± 6
E161Q + R174K	0.54 ± 0.06	21 ± 1	39 ± 5	0.26 ± 0.02	13 ± 1	51 ± 4	0.35 ± 0.05	12 ± 1	33 ± 7

closure and chemistry rate when compared to WT at 37±2 s^{-1}. E161D and E161Q each had a rate of 25 s^{-1}, thus by either shortening the residue side chain or removing the electrostatic interaction the overall effect on active site closure and chemistry was the same. The E161Q R174K mutant had an additional rate decrease to 13 s^{-1} while the R174K mutation alone had no further decrease in rate from the nucleotide capture step. This finding again suggests that removal of the electrostatic interaction between residues 161 and 174 can compensate for the lysine mutation. In all cases, the rate of catalytic cycle completion did not significantly differ from that of active site closure and chemistry indicating that the two events are tightly coupled.

2.3.6 Coxsackievirus sensor mutation kinetics

To verify the observed trends with the sensor mutations, the same mutations were introduced in to the coxsackievirus polymerase, with the hope that they would have the same observed changes in polymerase kinetics as those determined for the poliovirus polymerase. The same experiments detailed in Figures 2.2 and 2.3 were conducted for these polymerases and summarized in Table 2.4. The only observed deviation from the poliovirus results was that the R174K and E161Q R174K mutants had unstable elongation complexes, which made it impossible to determine some kinetic parameters. For the parameters that could be measured, they were in line with the results observed for the poliovirus polymerase mutants. It should also be noted that the slight drop in K_M observed in poliovirus for the E161Q mutation for both single nucleotide and processive elongation was exacerbated in the coxsackievirus polymerase. This also accounts for the drop in nucleotide discrimination factor for the E161Q R174K mutant that was not observed in poliovirus.

Table 2.4: Summary table of kinetics data for analogous coxsackievirus sensor mutants.

$3D^{pol}$	Initiation (min)	EC Stability (min)	Single Nucleotide k_{pol} (nt/sec)	Single Nucleotide K_M (μM)	Processive Elongation k_{pol} (nt/sec)	Processive Elongation K_M (μM)	Discrimination Factor
CV WT	1.3 ± 0.1	160 ± 20	22 ± 1	35 ± 2	15 ± 1	150 ± 10	300 ± 20
E161Q	8 ± 2	74 ± 4	22 ± 1	8 ± 1	11 ± 1	22 ± 4	130 ± 20
R174K	1.3 ± 0.1	< 5	-	-	-	-	-
E161Q + R174K	9 ± 1	< 10	1.7 ± 0.1	41 ± 7	-	-	80 ± 20

2.3.7 Sensor mutant crystal structures

As verification for the observed rate changes for each sensor mutation, a series of coxsackievirus crystal structures were solved to determine the positioning and interactions of residues 161 and 174. Mutation of residue 161 from a glutamate to glutamine did not show any significant structural changes to the polymerase (Figure 2.6A). A slight change in the positioning of R174 may have occurred, but in the apo structure it is difficult to determine how significant this change is relative to the positioning and nucleophilic attack on the incoming NTP. Mutation to residue 174 showed a much larger structural change (Figure 2.6B). In this structure the 174 residue is not fully resolved due to the flexibility of the lysine side chain. It is also possible that this flexibility hinders both the nucleotide capture and chemistry reaction efficiencies, thus supporting the slow rate for each step determined by the chemical quench reactions. This slow reaction rate could also be explained by a tight electrostatic interaction between residues 161 and 174, especially when the lysine mutation changes the delocalized positive charge of the arginine residue to the point charge of the lysine residue. It is conceivable that the lysine residue could occasionally occupy a position wherein it is interacting with the E161 residue and cannot break the electrostatic interaction, thus trapping residue 174 in an upward position and rendering it unable to interact with the incoming NTP. In either scenario, the lysine residue would not always be clearly resolved and would again result in decreased nucleotide capture and chemistry rates. Mutations to both residues 161 and 174 showed a positioning of each residue similar to that of the individually mutated residues (Figure 2.6C). This structural finding that the R174K mutation is no better resolved with the E161Q mutation suggests that side chain flexibility is the ultimate driving force behind the observed slower rates. However, there must also be some instances in

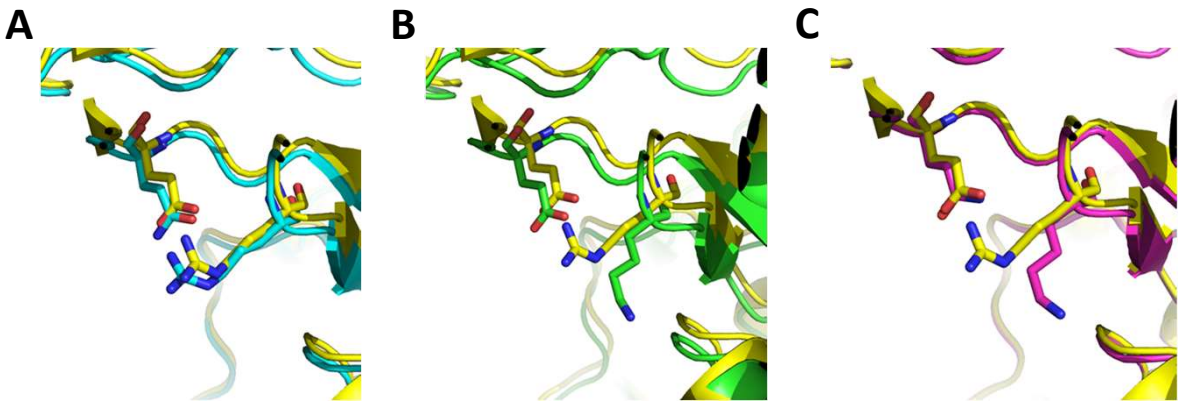

Figure 2.6: Mutations to sensor residues result in structural changes above the incoming NTP binding location.

Crystal structure alignments of WT polymerase (yellow) with the E161Q mutant polymerase (A, cyan), the R174K mutant polymerase (green, B), and the E161Q R174K double mutant polymerase (magenta, C). All structures were solved in the apo form.

which the lysine residue becomes trapped by the glutamate residue because the partial rescue of reaction rates cannot be explained by structural observations alone.

2.4 Discussion

The picornaviral polymerase $3D^{pol}$, while being one of the smallest replicating polymerases still maintains the conserved structural features of other single subunit polymerases, a cupped right-hand structure with fingers, palm, and thumb domains (Peersen, 2019a). However, unique to these polymerases is an interaction between the fingers and palm domains wherein the fingers domain contacts the top of the thumb domain and consequently restricts fingers domain movements. This restriction prevents $3D^{pol}$ from completing its catalytic cycle as other single subunit polymerases, such as the RNA polymerase T7 does; T7 relies on a swinging motion of the O-helix located within the fingers domain to drive the incoming NTP from a pre-insertion site to the catalytic cite, and a reversal of this movement results in the translocation of the template and product nucleic acid strands (Yin and Steitz, 2004). Because in $3D^{pol}$ this movement is not possible, these polymerases instead rely on a unique active site closure mechanism in order for catalysis to occur (Gong and Peersen, 2010; Gong et al., 2013). While this mechanism solves how these polymerases preform catalysis, the question how this polymerase completes the catalytic cycle and preforms steps such as RNA translocation remains incompletely answered. One such proposed mechanism is a long-range interaction domain within the fingers domain that could serve the function carried out by the previously mentioned O-helix (Shu and Gong, 2017). Thus, working in tandem individual structural elements within the fingers domain, such as the kink, gateway, and sensor residues, could aid in catalytic cycle completion. Work from this study supports this idea, with the kink residues primarily affecting RNA binding and maintained complex stability, gateways residues serving to assist in the template RNA strand

positioning post-translocation, and sensor residues playing a role in nucleotide capture and catalysis.

2.4.1 Kink residues play a role in RNA template binding efficiency

The increase in initiation time along with the decrease in elongation complex stability that was observed for the kink mutants suggest that these residues are involved in maintaining the polymerase's grip on the RNA. This is rather striking given that these residues appeared to no direct contact with either the template or product RNA strands in the original EC structures (Gong and Peersen, 2010; Gong et al., 2013). A recent publication by the Gong group has since shown the kink structure does indeed contact the template RNA strand, and plays a role in RNA binding (Shi et al., 2020), supporting our findings here. It should also be noted that this decrease in RNA binding efficiency made it not possible to determine individual rates within the catalytic cycle or two sequential CTP incorporation events. The inability to form an elongation complex structure in those cases suggests that by removing the hydrogen bonding network that makes up the kink structure, the polymerase's RNA binding is reduced. This finding could also support why single nucleotide and processive elongation rates are reduced by these mutations, as these residues do not interact with any necessary components within the active site. When these findings are taken together, they suggest that the kink structural element is likely part of a much larger intradomain hydrogen bonding network. Previous work modeling the flexibility of different regions within the polymerase identified the residues containing the kink structure to be relatively mobile (Shen et al., 2012b). Additional molecular dynamics simulations corroborated these findings and suggest that these residues may function in the NTP entry channel opening and closing (Moustafa et al., 2011), which could explain the changes in K_M that were observed for kink mutations P40G and R49K. Thus, not only would these residues be involved in

maintaining the polymerase's grip on the RNA, but also allow for NTP entry and closure back to the competent elongation complex upon nucleotide capture.

2.4.2 Gateway residues serve to position the RNA template before NTP binding and catalysis

Increases in initiation time, decreases in elongation complex lifetime, and processive elongation rate correlated with residue 114 side chain size suggest a steric interaction between the template RNA strand and the residue. Mutation to these residues also showed no changes observed for single NTP incorporation rate and the necessity of an addition reposition step between two sequential CTP additions. Taken together these findings suggest that residue 114 with the aid of residue 115 acts as a positioning factor for the template RNA post-translocation. Previously, the residues located in motif B have been implicated in RNA translocation. These residues, 288-292, fluctuate between an "in" and "out" conformation as a possible mechanism to facilitate RNA template strand translocation (Sholders and Peersen, 2014). This swinging motion likely not only creates a driving force for translocation but also would serve to create additional space within the polymerase core to allow for RNA movement. Upon completion of template strand translocation, the polymerase must reset in order for another round of nucleotide incorporation to occur. It has previously been suggested that residue 114 would act as a sort of catch for the template strand RNA to prevent mispositioning (Shu and Gong, 2017). If residue114 does indeed serve as a molecular RNA catch, then by varying the side chain size of residue 114 interactions between the polymerase and the template strand RNA would have been affected. An increase in size for the residue 114 side chain would result in not only a larger barrier for the template RNA strand to traverse, but increasing sized of either the residue 114 or 115 side change would result in template strand positioning that would ultimately shift the incoming NTP's position in the active site towards the nucleotide entry channel, likely resulting

in the observed increase in K_M seen for single CMP incorporation and the changes in K_D for sequential CMP incorporation due to a now repositioned nucleotide binding site. The findings here support this idea as changes in K_M were noticeable in the single nucleotide incorporation assay, however rate changes were subtle and not detectable until multiple rounds of nucleotide incorporation had occurred. Further kinetic modeling pointed to the need for a repositioning step before a second round of catalysis could occur. As for reducing the residue 114 side chain size, that had only effects on rate and none on K_M. This rate change, while not as drastic as when side chain size was increased, suggest that only the rate of RNA strand capture post-translocation was affected due to the lack of a barrier to translocation being present.

Unlike the T114 mutants, the S115 mutants have consistently higher values for each CMP incorporation event in comparison to WT. These mutations also have consistently faster rates for each catalytic cycle than WT. Likely this combination of faster rate and higher K_D gives the overall observation of similar single NTP incorporation rates seen in the single CTP incorporation assay. It is very clear that residue 115 seems to play a larger role in the elongation activation rate, as reduction from a serine to alanine resulted in an almost 3-fold increase in rate. As with the T114 mutants, the mutations to S115 that increase side chain size decrease repositioning rate, and follow the trend observed for processive elongation. It should also be noted that mutations to either T114 or S115 all seemed to decrease the amount of EC in the active state, which supports the idea that T114 and S115 help establish template-RNA positioning in the active site.

2.4.3 Sensor residues affect nucleotide capture and chemistry rates

The sensor mutant residues 161 and 174 form a salt bridge that directly positions residue 174 above the incoming NTP's α-β phosphate linkage. This may allow residue 174 to act as the

final checkpoint for the correct nucleotide, and a possible indicator that chemistry has occurred, especially if residue 174 is the general acid and proton donor for chemistry, which based on other classes of polymerases, it structurally could be (Gong and Peersen, 2010; Shen et al., 2012a). By mutating residue 174 from an arginine to a lysine, the delocalized positive charge that would normally interact with the delocalized negative charge of residue 161 is now a single point charge. The lysine residue would likely become trapped in an electrostatic interaction with residue 161, and no longer be able to readily donate a proton, which would account for the reduced rate observed for all steps within the catalytic cycle. The inherent lysine side chain flexibility would also have a secondary effect, in that now the ability of residue 174 to act as a checkpoint for nucleotide incorporation would be compromised and is supported by the large decrease in nucleotide discrimination factor. Interestingly, it appears that both effects are observed in the crystal structure, with a loss of clear electron density for the ε-ammonium group, thus implying that the lysine side chain is not able to form a consistent salt bridge with E161, even when there is no incoming NTP. By no longer having the ridged optimal geometry present in the active site, incoming NTP positioning is altered, accounting for the slower nucleotide capture rates that were observed. In addition, the slow rate for nucleotide capture, which shifted the catalytic cycle rate limiting step away from active site closure, points to the importance of residue 174 as this decreased rate would be indicative of a polymerase no longer being able to signal that the incoming NTP is correctly placed and ready for active site closure and chemistry to occur.

Mutations to residue 161 produced different results from those made to residue 174, even though they are two sides of a single salt bridge. By shortening the side chain of residue 161 from a glutamate to an aspartate, very few changes were observed with the exception of a

reduction of processive elongation rate and an increase in nucleotide capture rate. While these two results may appear counter to each other, it is possible that these results shine light on two aspects of the catalytic cycle residue 161 plays a role in. The first is nucleotide capture. As previously discussed, residue 161 plays a role in positioning residue 174 as an incoming nucleotide checkpoint. Shortening the side chain of residue 161 would move residue 174 further away from the incoming NTP and place it in a non-ideal geometry for hydrogen bonding to the incoming nucleotide, increasing the distance between the active site and without changing its relative positioning of residue 174 as a checkpoint. Thus, steric constraints associated with nucleotide capture would be reduced without sacrificing nucleotide discrimination. This relief of steric constraints would come at a cost, which would be the ability for residue 174 to efficiently donate a proton during chemistry, resulting in the decrease in processive elongation rate observed for this mutation. In contrast, a mutation to a glutamine from a glutamate had no effect on side chain length for residue 161 and instead disrupted the salt bridge electrostatics between residues 161 and 174 by removing the negative charge on residue 161. This disruption would place less constraints on the residue 174, ultimately changing its ability to act as a nucleotide checkpoint, accounting for the decrease in nucleotide discrimination factor for this mutation. The crystal structures for the E161Q mutation showed very little difference in active site structure, which could mean that even though the electrostatic interaction has been removed, hydrogen bonding between R174 and E161Q is still able to maintain a minimal amount of positioning needed for those two residues. This relaxation on the positioning of residue 174 could also decrease the efficiency of proton transfer, ultimately leading to the decreases in active site closure and chemistry, single nucleotide incorporation, and processive elongation rates that were observed.

By mutating both residues 161 and 174 to a glutamine and lysine respectively, the cooperative roles these two residues play in the catalytic cycle were able to be determined. The reduction in nucleotide capture rate, while not as severe as the mutation of residue 174 alone, points to a clear interplay between these residues, and the necessity of an arginine at residue 174 for effective NTP capture. Where the R174K mutation alone likely prevented residue 174 from interacting with the incoming nucleotide efficiently due to inconsistent positioning of the lysine side chain, the glutamine mutation counteracts this effect by releasing the electrostatic constraints placed on the lysine mutation. This interaction would likely now form a hydrogen bond, which would aid in residue 174 positioning and account for the minimal change in nucleotide discrimination factor that was observed. However, it is probable that the efficiency and in turn the rate for nucleotide capture is reduced due to a hydrogen bond being a much weaker interaction than an electrostatic one. As was seen in the crystal structures for the R174K alone, the E161Q R174K double mutation structure also had a loss of electron density on for the ε-ammonium group, and a clear deviation in lysine side chain positioning. Looking at the structure in conjunction with the rapid chemical quench data it seems that the ability to have both an electrostatic interaction with the incoming NTP as well as active site positioning via residue 161 is what drives the rate for nucleotide capture. Unlike with the R174K mutation alone, nucleotide capture does not become the rate limiting step, due to no negative charge hindering proton donation from residue 174 to the incoming NTP, however like the residue 161 mutations, the rate for active site closure and chemistry is reduced when compared to WT. This reduction would account for the observed changes in single nucleotide incorporation and processive elongation rates, and likely result for inefficient proton transfer from the lysine residue. As both polymerases with the R174K mutation do not have a clear positioning for residue 174 in the

crystal structure, it is likely that an optimal active site geometry plays an influential role in determining the reaction rate.

2.5 Conclusions

While picornaviral polymerases perform the same functions as other single subunit polymerases, mainly to create a new nucleic acid strand from an existing template strand, the structural constraints present for this family of polymerases presents a unique challenge. The restricted fingers domain movement has not only lead to a distinct active site closure mechanism, but also has prevented additional functions, such as translocation, to occur in the same manner as other single subunit polymerases. Our work sheds light on three structural elements located in the fingers domain and elucidates how these regions assist in the polymerase's nucleic acid addition cycle. Each set of residues preformed a unique function, with the kink residues assisting in RNA binding, the gateway residues serving as a molecular catch for the template strand post-translocation, and the sensor residues acting as a nucleotide checkpoint and proton donor for chemistry. Unlike the O-helix found in other single subunit polymerases, none of these regions served multiple functions during the catalytic cycle and instead work as part of a larger intra-domain network to carry out each round of nucleic acid addition. Thus, not only have these picornaviral polymerases had to evolve a unique active site closure mechanism, the other processes surrounding it have also been solved in a unique fashion.

CHAPTER 3

PICORNAVIRUS POLYMERASE DOMAIN EXCHANGES REVEAL A MODULAR BASIS FOR DISTINCT BIOCHEMICAL ACTIVITIES OF VIRAL RNA-DEPENDENT RNA POLYMERASES [1]

Author contributions to this chapter: authors Kempf and Beaucourt conducted the virology work in poliovirus and coxsackievirus, respectively. The chimeric polymerase creation, purification, and kinetic characterization as well as figures and tables in this chapter were done by me.

3.1 Introduction

Picornaviruses are a family of positive-strand RNA viruses responsible for a multitude of human and animal diseases. Upon infection, their 7.5 kb RNA genomes are directly translated via an internal ribosome entry site, yielding a ~250 kDa polyprotein that is cleaved into about a dozen individual proteins and multiple functional intermediates. RNA genome replication is carried out by the $3D^{pol}$ protein, a RNA-dependent RNA polymerase (RdRP) that first uses the infecting RNA as a template for making negative-strand RNA and then uses these as templates for generating new positive-strand RNA for translation or for packaging into progeny virus particles. The ≈460-residue picornaviral $3D^{pol}$ enzymes are the among smallest viral RdRPs, yet their core structure and mechanism are retained in larger enzymes from other positive-strand RNA viruses (Ferrero et al., 2018; Peersen, 2019a, 2017), making them a useful model for studying RNA virus polymerase biochemistry.

[1] This work has been published in The Journal of Biological Chemistry as: Watkins CL, Kempf BJ, Beaucourt S, Barton DJ, Peersen OB. Picornaviral polymerase domain exchanges reveal a modular basis for distinct biochemical activities of viral RNA-dependent RNA polymerases. *J Biol Chem*. 2020; https://doi.org/10.1074/jbc.RA120.013906.

The RdRP core structure has the classic polymerase resemblance to a cupped right hand with palm, fingers, and thumb domains. The palm domain contains conserved motifs A, B, and C that form the active site using the two-metal mechanism commonly found in replicative DNA polymerases (Steitz, 1998), the fingers domain binds template RNA, and the thumb domain interacts with dsRNA product. A notable and unique feature of positive-strand RNA virus RdRPs is that the fingers make a direct contact with the top of the thumb domain, effectively encircling the active site and creating a short tunnel whereby NTPs enter the active site. Mutation of this interaction severely destabilizes poliovirus $3D^{pol}$ in thermal denaturation studies (Thompson et al., 2007), leading to the theory that the viral RdRPs evolved the fingers–thumb contact to stabilize the fold of a small polymerase encoded within the space limitations of a small viral genome. This stabilization in turn has major implications for both nucleotide addition and translocation mechanisms.

Replicative polymerases often use a swinging motion of their fingers domain to capture and then position NTPs in the active site for catalysis, and these interactions play important roles in nucleotide recognition and replication fidelity (Steitz, 2006; Temiakov et al., 2004). The reverse motions to open the active site after catalysis facilitate translocation by directly contacting the RNA template and pushing it by one base-pair register in a power-stroke driven by pyrophosphate release (Yin and Steitz, 2004). In contrast, positive-strand RNA virus RdRPs with their thumb-tethered fingers domain cannot undergo such motions and they instead use a subtle movement of motifs A and D within the palm domain to fully structure the active site for chemistry only upon NTP binding (Gong and Peersen, 2010; Wu and Gong, 2018). Their fingers domains remain essentially stationary during the nucleotide addition cycle, and their mechanism

for translocation is not yet well understood beyond an initial half base-step product strand movement seen in enterovirus 71 polymerase structures (Shu and Gong, 2016).

This simple active site closure mechanism gives rise to the relatively low replication fidelity that is a common feature of positive-strand RNA virus RdRPs, and it results in progeny virus populations composed of a large pool of sequence variants that are sometimes referred to as quasispecies (Andino and Domingo, 2015; Korboukh et al., 2014; Lauring and Andino, 2010; McCrone and Lauring, 2018). The population consensus sequence defines the virus, but close inspection of individual genomes show one or more random mutations relative to that consensus. This genetic diversity is essential for virus growth and spread, particularly in hosts where the virus must replicate in a variety of cell types, and it has been shown that either increasing or decreasing polymerase fidelity will adversely affect virus growth (Gnadig et al., 2012; Korboukh et al., 2014; Peck and Lauring, 2018; Pfeiffer and Kirkegaard, 2003). To guard against hypermutation, RNA viruses also undergo recombination events whereby entire genome sections are swapped with counterparts from other viruses, allowing for the large-scale purging of deleterious mutations (Kempf et al., 2016, 2019; Woodman et al., 2018a; Xiao et al., 2016).

In a prior study of $3D^{pol}$ fidelity variants we made the observation that despite having essentially identical structures, the poliovirus (PV) and coxsackievirus B3 (CV) enzymes differ drastically in their biochemical characteristics (Campagnola et al., 2015). PV $3D^{pol}$ elongates RNA 4-fold faster than CV $3D^{pol}$ with processive elongation rates of 88 versus 22 nt/s at 37 °C, and it does so with much lower nucleotide discrimination than the CV enzyme. Interestingly, mutations in motif A tend to reduce CV $3D^{pol}$ fidelity, while analogous mutations in PV polymerase tend to increase fidelity. This led to the hypothesis that the picornaviral polymerases represent a minimal RdRP core structure whose biochemical characteristics can be easily fine-

tuned by mutations and evolution to optimize the growth of any given virus. The strongest fidelity variants identified to date have been in the palm domain, near the active site or within the structurally coupled motifs A and D, but some have also been identified in the fingers domain (Peersen, 2017; Yang et al., 2012). Both NMR data and molecular dynamics calculations implicate long range protein dynamics networks as modulators of fidelity (Cameron et al., 2009; Moustafa et al., 2014; Shi et al., 2019; Yang et al., 2010). This led us to ask if the different $3D^{pol}$ structural domains serve distinct and separable functions during the catalytic cycle, and the extent to which they can modulate each other's activity with respect to RNA replication functions. The project builds on prior work by Cornell *et al.* (Cornell and Semler, 2002) examining polyprotein processing and RNA binding activities of chimeric poliovirus–coxsackievirus B3 3CD proteins, but with a focus on polymerase functions and a more detailed protein structure driven approach.

In the work presented here, we use the high structural similarity of picornaviral polymerases to generate a set of chimeric enzymes that directly examine the role each polymerase structural domain plays in initiation, single cycle and processive elongation, translocation, elongation complex stability, and nucleotide discrimination. We also tested the ability of chimeric coxsackie-polio polymerases to support virus replication in cell culture.

3.2 Methods

3.2.1 Chimeric polymerase construction

The pinky finger exchanges, i.e. CPPP and PCCC proteins, are a straightforward swap of a single contiguous segment from residues 66-150, which includes all of polymerase motif G. In contrast, the full fingers domain exchange is more complicated because it has five crossover

points with the palm domain and it forms a structural contact with the top of the thumb. The palm-fingers junction points chosen were based on the palm residues resolved and modeled in the original wildtype PV $3D^{pol}$ structure (PDB: 1RDR) where a crystal packing interaction led to the complete unfolding of the fingers domain (Hansen et al., 1997; Thompson and Peersen, 2004). Note, however, that residues 27-37 at the tip of the index finger were always matched with the origin of the thumb domain so as to preserve the native fingers-thumb contact (Figure 3.1B). For the thumb domain exchanges, polymerases were made with two different junction points located on either side of motif E, often referred to as the "primer-grip" element in the reverse transcriptase literature. These polymerases are designated with a subscripted "l" for a long thumb exchange that includes motif E ($PPPC_l$), or a subscripted "s" for a short thumb exchange beginning after motif E ($PPPC_s$). These two thumb exchange sites generally provided equivalent results in biochemistry experiments and data are reported for the long thumb junction point that begins with residue 369_{PV} (370_{CV}) unless explicitly stated.

The chimeric polymerase genes were assembled into a T7-based expression vector with the In-Fusion cloning kit (Clontech) using a combination of PCR products, oligonucleotides, and GeneBloc fragments (IDT Inc., Coralville, IA) as needed, and the final constructs were verified by sequencing. Polymerase expression and purification were carried out as previously described (Hobdey et al., 2010) with the final purification step being size exclusion chromatography into a buffer composed of 5 mM Tris, pH 7, 300 mM NaCl, 2 mM Tris-(2-carboxyethyl)-phosphine (TCEP), 20% glycerol.

3.2.2 Initiation and stability

Initiation rates were assayed using 5 uM polymerase, 0.5 uM "10+1_12" RNA (Hobdey et al., 2010) and 40 uM GTP at room temperature in 75 mM NaCl, 4 mM MgCl2, 25 mM

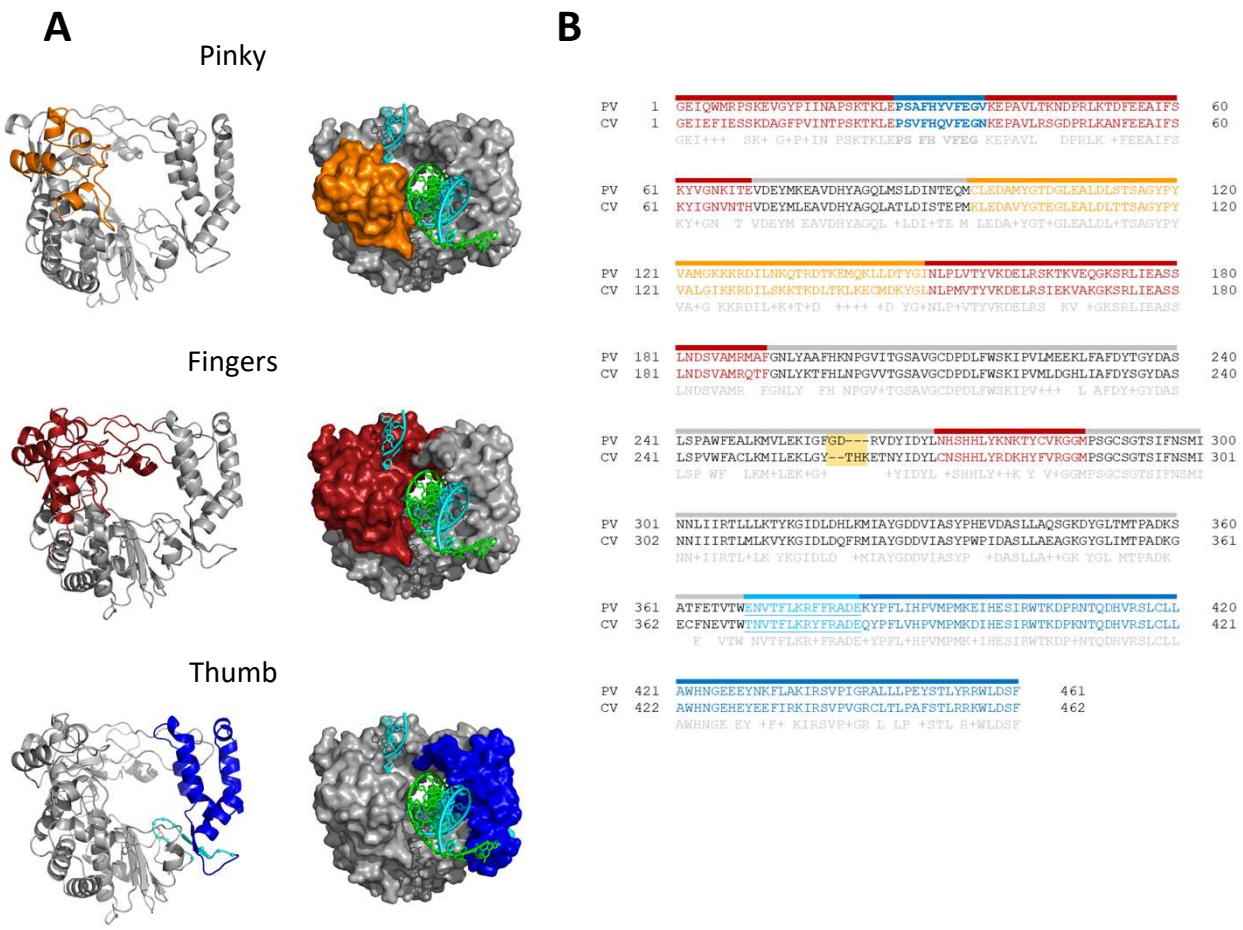

Figure 3.1: Chimeric polymerase construction.

(A) Domain exchanges illustrated with the structure of 3D^{pol} elongation complex (PDB entry 3OL6). Pinky finger is shown in orange, the entire fingers domain including the pinky finger is shown in red, and the thumb domain is shown in blue with the primer grip at the beginning of the thumb domain in cyan with alpha carbon spheres. RNA is colored cyan for template and green for product strand. (B) Structure based sequence alignment between poliovirus and coxsackievirus B3 polymerases to show the junction points used in the chimeric polymerase construction. Pinky, fingers, palm, and thumb domain colors correspond to panel A, with palm domain sequence shown in gray. Residues 27-37 were considered as part of the thumb domain to preserve the fingers-thumb contact. There is 75% amino acid identity and 86% similarity between the two polymerases, and their structures are essentially identical except for a single residue insertion in a loop at residue 260 (yellow box).

HEPES pH 6.5, 2 mM Tris-(2-carboxyethyl)-phosphine (TCEP). 1 uL samples were removed at various time points up to 15 minutes and added to 19 uL of quench buffer containing 50 mM EDTA, 400 mM NaCl, 50 mM HEPES pH 6.5, and 2 mM TCEP. After the initiation reaction had progressed for 15 minutes, the assay for the temporal stability of the resulting elongation complexes was begun by adding 10 uL of the initiation reaction to 90 uL of 300 mM NaCl, 4 mM MgCl2, 50 mM HEPES pH 6.5, and 2 mM TCEP, wherein the high salt concentration limited further initiation and re-initiation by preventing RNA binding. The amount of competent elongation complex still present in this solution was tested at various points over a four-hour period by removing 5 uL aliquots and mixing them with 5 uL buffer containing 160 uM each of ATP, GTP, and UTP, followed by a two-minute elongation reaction to produce full-length products before quenching with EDTA. Samples were analyzed by denaturing gel electrophoresis using 20% 19:1 polyacrylamide, 7M urea, 1X TBE gels. RNA bands were detected using a Li-Cor Odyssey 9120 infrared imager system to visualize the IRdye 800RS (Li-COR Biosciences) label on the RNA tetraloop, as previously described (Hobdey et al., 2010). Data were mathematically fit to single exponential equations using Kaleidagraph (Synergy Software) to determine time constants for the increased formation of +1 product in the initiation assay or the decreased formation of +7 chase product in the stability assay.

3.2.3 Enzyme kinetics assays

Stopped-flow fluorescence methods were used to measure kinetics of processive elongation (Gong et al., 2009) and single nucleotide incorporation (McDonald et al., 2016). Stalled elongation complexes were first generated by incubating 15 uM polymerase with 10 uM RNA in 75 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM TCEP, and 60 uM each of ATP and GTP for 15 minutes. The resulting complexes were diluted 200-fold to a final RNA

concentration of 50 nM with 50 mM HEPES pH 7, 75 mM NaCl, and 4 mM $MgCl_2$, generating the sample that was loaded into the stopped-flow instrument and then further diluted 2-fold (i.e. 25 nM final) when mixed with either only CTP or all four NTPs in the reaction cell. $MgCl_2$ was always maintained at 4 mM excess over the total NTP concentration, and for the processive elongation assays the four NTPs were present at equimolar concentrations. Kinetics data were collected at 30°C using a Bio-Logic SFM-4000 titrating stopped-flow instrument with a MOS-500 spectrometer. Fluorescence data from the single CMP incorporation and processive elongation reactions were fit using Kaleidagraph (Synergy Software, Reading, PA), and data from double CMP incorporation reactions were analyzed and modeled using KinTek Explorer (Kin-Tek Corp, Austin, TX) (Johnson et al., 2009b, 2009a).

Rapid chemical quench experiments were done by first forming stalled elongation complexes with 15 uM polymerase and 10 uM 5′-fluorescein labeled RNA for 15 minutes at room temperature in buffer containing 75 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM TCEP, 200 uM ATP, and 60 uM GTP. The resulting complexes were diluted 20-fold to a final concentration of 500 nM RNA with reaction buffer consisting of 50 mM HEPES pH 7, 75 mM NaCl, and 4 mM $MgCl_2$. Polymerase complexes were further diluted 2-fold in the Bio-Logic QFM-4000 instrument when mixed with various concentrations of CTP, the single nucleotide substrate, made up in this same buffer. The QFM-4000 instrument controls reaction time by varying flow rate through a 3 uL reaction chamber that ends at mixing with the quench solution. Stalled elongation complex and CTP were mixed and allowed to react for times ranging from 5 to 2000 ms before quenching with either 300 mM EDTA (100 mM final) or 3 M HCl (1M final). A total of 15 ul of each component was used in the reaction, yielding 45 uL final sample volumes. For the acid quench, 30 uL of the final sample was manually added to 30 ul 1 M

NaOH, 300 mM MES so as to neutralize the samples prior to gel analysis. Reaction products were analyzed by denaturing gel electrophoresis on 20% polyacrylamide (19:1), 7M urea, 1X TBE gels, and imaged using the 488 nm channel of an Amersham Typhoon Imager. Product formation rates were measured at eight different CTP concentrations with 8-12 timepoints collected for each concentration. Gel bands were quantified using the program PeakFit (Systat Software) to analyze lane scans by fitting band profiles to gaussian peaks, and closely spaced peaks in a given lane used a common (fitted) peak width to improve quantitation of weak bands. We verified that the 5′-fluorescein modification (/56-FAM/ from IDT Technologies, Inc) was chemically resistant to the acid quench conditions.

3.2.4 Virus replication studies

The chimeric polymerases were cloned into the full-length infectious poliovirus cDNA using the In-Fusion method (Clontech) with PCR products and proper insertion was verified by sequencing. RNA was transcribed from MluI-linearized plasmids using T7 RNA polymerase (Ampliscribe T7 high-yield transcription kit; CellScript Inc.) and 1 ug was transfected in triplicate into ~10^6 HeLa cells in 35 mm 6-well plates using Transmessenger transfection reagent (Qiagen) according to manufacturer's instructions. Transfected cells were fed with 2 mL of cell culture media (DMEM containing 10% fetal bovine serum, 100 U per ml penicillin, and 100 ug per ml streptomycin) and incubated at 37ºC. Cells were examined for cytopathic effect (CPE) at 24, 48, 72, and 96 hours post-transfection. Infectious virus (P_0, Passage zero) was harvested following three rounds of freeze-thaw at the time of CPE (or 96 hpt when CPE was absent) and quantified by plaque assay. P_0 virus was then used to infect ~10^7 HeLa cells in T_{75} flasks at a MOI of 10 PFU per cell. Infected cells were incubated with 12 ml of cell culture media. P_1 virus

was harvested by freeze-thawing cells at 24-72 hours post-infection (hpi), based on the timing of cytopathic effects, and quantified by plaque assay.

The P_1 virus RNA genomes were analyzed to verify the presence and integrity of the chimeric polymerase constructs. P_1 virus (8 ml) was layered onto 3 ml of 30% (w/v) sucrose in phosphate-buffered saline followed by centrifugation at 36,000 rpm for 4 h at 4 °C using a Beckman SW41 rotor. Pelleted material containing virus was resuspended in 400 µl 0.5% SDS buffer [0.5% SDS, 10 mM Tris–HCl (pH 7.5), and 100 mM NaCl]. Virion RNA was isolated by phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation. Virion RNA was solubilized in 10 ul water. cDNA was synthesized using 2 ul of virion RNA in 20 ul reactions using Superscript III (Invitrogen) reverse transcriptase and a primer complementary to poliovirus nucleotides 7415–7440 in the 3′ NTR (5′ $_{7440}$CTCCGAATTA AAGAAAAATT TACCCC$_{7415}$ 3′). Reactions were incubated at 65ºC for 5 min, followed by the addition of RT, followed by further incubation at 55ºC for 1 hr and 70ºC for 15 min. cDNA corresponding to the 3CD region of poliovirus RNA was amplified using forward and reverse primers:

Forward Primer (nucleotides 5816-5838):

5′ CTC GGT GGG CGC CAA ACT GCT CG 3′

Reverse Primer (nucleotides 7415-7740):

5′ CTC CGA ATT AAA GAA AAA TTT ACC CC 3′

PCR reactions (50 ul) containing 2 ul of viral cDNA were incubated for 2 min at 95 ºC and subjected to 30 cycles of PCR (94 °C 30 sec, 60 °C 30 sec, 72 °C 2 min) and a final extension at 72 °C for 2 min. PCR products were analyzed by agarose gel electrophoresis and

ethidium bromide staining. Plasmid DNAs and viral cDNAs (~ 3 ul of PCR products – ExoSAP treated) were sequenced using the forward and reverse primers noted above in the University of Colorado Cancer Center DNA Sequencing Core Laboratory.

3.2.5 Serial passage of PPPCsm and PPPClm (small plaque) viruses

Viruses containing the two mutated thumb exchanged polymerases, $PPPC_{sm}$ and $PPPC_{lm}$, exhibited a small plaque phenotype, and serial passage in HeLa cells was used to see if both the genotypes and phenotypes of these viruses were stable. ~10^7 HeLa cells in T_{75} flasks were infected with 1 ml of P_1 progeny virus (a MOI of ~10 PFU per cell). Virus was harvested at 24 hpi by three rounds of freeze-thaw in 10 ml of culture media. After 10 cycles of infection, plaque assays were used to examine viral titers and plaque phenotypes and viral cDNA was prepared as described above. Viral cDNA was sequence both before and after TOPO-TA cloning. 30 TOPO-TA clones were sequenced in the 3CD region of the genome to look for individual variants within the population. $PPPC_{sm}$ and $PPPC_{lm}$ viruses retained small plaque phenotypes after 10 cycles of infection and were indistinguishable from P_1 viruses. No mutations were detected in $PPPC_{sm}$ and $PPPC_{lm}$ viruses after 10 cycles of infection beyond those engineered into each virus.

3.2.6 Serial passage of non-viable chimeras

Several chimeras were non-viable (NV) following initial transfections and infections, as no plaques were detected at P_0 and P_1. For these NV chimeras ($PPPC_s$, $PPPC_l$, CCCP, PPCC, PCCC and CCCC), we used serial passage in HeLa cells to see if viable virus would arise during passage. Undiluted P_0 was used to inoculate ~10^7 HeLa cells in T_{75} flasks. Cells were monitored by microscopy for CPE. At 96 hpi, cells were subjected to three rounds of freeze-thaw in 10 ml of culture media. After 10 cycles of serial passage, plaque assays were used to detect virus and

cDNA was prepared from media concentrated by centrifugation as described above. Throughout serial passage, neither CPE nor plaques were observed for these NV chimeras ($PPPC_s$, $PPPC_l$, CCCP, PPCC, PCCC and CCCC) and cDNA failed to yield PCR products, indicating that neither cytopathic nor non-cytopathic virus were present (Smithee et al., 2015).

3.3 Results

3.3.1 Chimeric polymerase design and purification

Polymerase domains for chimeric RdRPs were identified by inspection of superposed PV and CV $3D^{pol}$ structures (Peersen, 2019a, 2019b) and involve four distinct folding domains; the pinky finger subdomain, the fingers domain, the palm domain and the thumb domain (Figure 3.1A). We name these chimeric polymerases using a four-letter code in the order pinky-fingers-palm-thumb, with the letter C or P denoting which virus (CV or PV) each domain was derived from. Thus, PPCC is a chimeric polymerase with pinky and fingers domains from poliovirus, and palm and thumb domains from coxsackievirus. The full set of chimeras generated are described in Experimental Procedures with borders shown in Figure 3.1B. The two wildtype proteins have 74% amino acid sequence identity and 86% similarity, and the CV enzyme is one residue longer due to a loop insertion at residue ~260 (Figure 3.1B).

The chimeric polymerases generally expressed well in *E. coli* and could be purified at high yield using the normal protocol of Ni-affinity, anion exchange, and size exclusion chromatography followed by concentration to ~200 uM in 300 mM NaCl buffer. The one exception was the PPCC polymerase that had a tendency to aggregate upon concentration in high salt buffer and during room temperature initiation assays at micromolar concentrations in 75 mM salt, suggesting its folding stability was compromised. This is not altogether unexpected

considering the full fingers exchanges involve five different junction points inside an essentially contiguously folded domain boundary (Figure 3.1B).

3.3.2 Initiation rates and elongation complex stability

Basic functionality was tested with RNA initiation assays that measure the time needed to form +1 RNA products after mixing polymerase with RNA and GTP, and with stability assays that provides a measure of stalled elongation complex lifetime. The initiation data were well fit by single exponential curves (Figure 3.2C and 3.2D), resulting in time constants that are dominated by a combination of RNA binding and a reorientation step that places the RNA into a catalytically competent conformation (Table 3.1). The wildtype PV (PPPP) and CV (CCCC) proteins have very different initiation times of 12±2 and 1.4±0.3 minutes, respectively. The data show this difference can almost entirely be attributed to the pinky finger structure; placing the CV pinky onto PV $3D^{pol}$ (CPPP) reduces initiation time from 12 to 2 minutes, and conversely placing the PV pinky on CV (PCCC) increases initiation time from 1.4 to 10 minutes. In contrast, swapping the entire fingers domains (CCPP and PPCC) leads to intermediate initiation times of 4 and 7 minutes, indicating that grafting the more extensive interdomain contacts between the palm and fingers domains disrupts but does not eliminate fingers domain function. Exchanging the thumb domains (PPPC and CCCP) had only minor effects on initiation rates.

The elongation complex stability assays test how much of the initiated RNA can be extended to full length product after incubation in a high salt buffer that prevents rebinding of RNA (Figure 3.2); polymerase-bound RNA will be rapidly elongated to a +7 product, while dissociated RNA will remain as the +1 product. The wild type elongation complexes are long lived and inactivate with time constants of ~100 and ~150 minutes for PV and CV, respectively. All the chimeric polymerases form less stable complexes than their wildtype counterparts, with

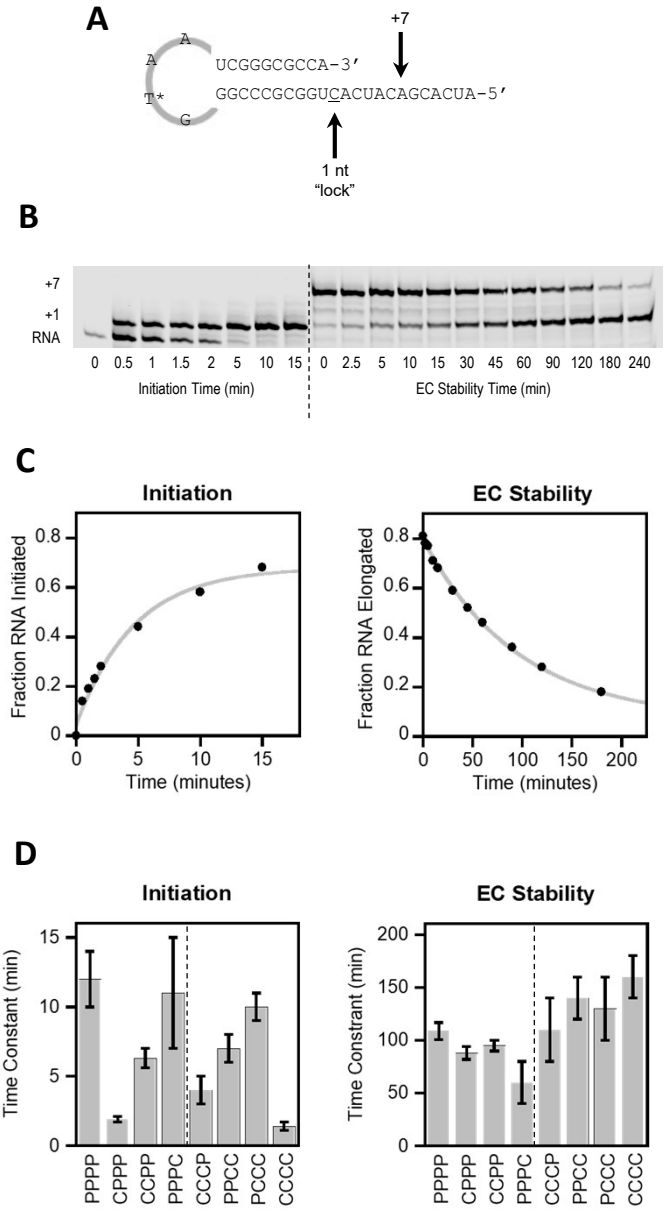

Figure 3.2: Polymerase initiation and stability.

(A) RNA construct used in initiation and stability assays. (B) Example gel used to quantitate RNA population changes with respect to time during the initiation and stability assays. Starting RNA, the +1 RNA product from a single GMP incorporation, and the +7 RNA product are labeled. Dashed vertical line denotes transition to the stability experiment where +1 elongation complexes are assayed for their ability to chase to a longer +7 product upon addition of NTPs. Times are labeled with respect to the assay being conducted. (C) Example single exponential fittings used to determine time constants associated with initiation and elongation complex stability. (D) Results of the initiation and stability assays for the two wildtype (PPPP and CCCC) and six chimeric polymerases, grouped by polymerase palm domain with poliovirus on the left and coxsackievirus on the right of the dashed line.

Table 3.1:Polymerase initial characterization kinetics data.

Wild type poliovirus (PPPP) and wild type coxsackievirus (CCCC) $3D^{pol}$s are denoted at the top for comparison with chimeric polymerases. Chimeric polymerases are grouped according to the domain exchanged.

$3D^{pol}$	Initiation (min)	EC Stability (min)	Single Nucleotide k_{pol} (sec^{-1})	Single Nucleotide K_M (μM)	Processive Elongation Rate (nt/sec)	Processive Elongation K_M (μM)	Discrimination Factor
PPPP (PV WT)	12 ± 2	109 ± 8	26 ± 1	30 ± 3	24 ± 1	105 ± 5	120 ± 20
CCCC (CV WT)	1.4 ± 0.3	160 ± 20	22 ± 1	35 ± 2	17 ± 1	153 ± 6	300 ± 20
Pinky Swaps							
CPPP	1.9 ± 0.2	88 ± 6	30 ± 1	42 ± 5	19 ± 1	76 ± 8	120 ± 20
PCCC	10 ± 1	130 ± 30	18 ± 1	45 ± 6	14 ± 1	120 ± 20	270 ± 20
Fingers Swaps							
CCPP	4 ± 1	95 ± 5	28 ± 1	32 ± 4	26 ± 1	83 ± 4	120 ± 20
PPCC	7 ± 1	140 ± 20	13 ± 1	60 ± 9	5 ± 1	101 ± 10	180 ± 20
Thumb Swaps							
PPPC	11 ± 4	60 ± 20	29 ± 2	37 ± 3	21 ± 1	148 ± 9	150 ± 20
CCCP	4 ± 1	110 ± 30	28 ± 1	52 ± 7	19 ± 1	147 ± 5	160 ± 20

the smallest effects due to the pinky swaps and the largest arising from the thumb swaps, but even the largest reductions of inactivation rate are less than 2-fold (Table 3.1). Notably, these experiments show that all the chimeric polymerases can form stable elongation complexes, enabling further biochemical analysis of nucleotide addition by rapid kinetics methods.

3.3.3 Rapid kinetics studies of elongation

Nucleotide incorporation kinetics were measured by stopped-flow fluorescence methods for processive elongation over a 20-nucleotide long single-stranded RNA template (Figure 3.3A) (Gong et al., 2009) and for a single nucleotide addition using CTP and 2′-deoxy-CTP (dCTP) templated by a guanosine (McDonald et al., 2016) (Figure 3.3B). The single cycle addition data were used to calculate a nucleotide discrimination factor as the ratio of the catalytic efficiencies for CTP and dCTP incorporation, i.e. $(k_{pol}/K_M)_{CTP}$ / $(k_{pol}/K_M)_{dCTP}$. In this context we use k_{pol} to indicate a full nucleotide addition cycle (NAC) going from NTP binding through catalysis and translocation, at which point we observed changes in 2-aminopurine fluorescence. We have previously shown that the biochemical discrimination factor, which is rooted in the importance of the NTP 2′-OH group in stabilizing the closed conformation active site (Gong and Peersen, 2010), serves as a proxy of changes to polymerase replication fidelity that correlated with mutation frequencies observed by genome sequencing methods (Campagnola et al., 2015; Gnadig et al., 2012).

The data from the full set of chimeric polymerases is summarized in Figure 3.4 as plots of nucleotide discrimination factors versus the maximal processive elongation rates. The data from the pinky finger and complete fingers domain exchanges show a clear delineation of function where the chimeric polymerases cluster by the origin of their palm and thumb domains (Figure 3.4A). Exchanging the coxsackievirus pinky finger (CPPP) or full fingers domain

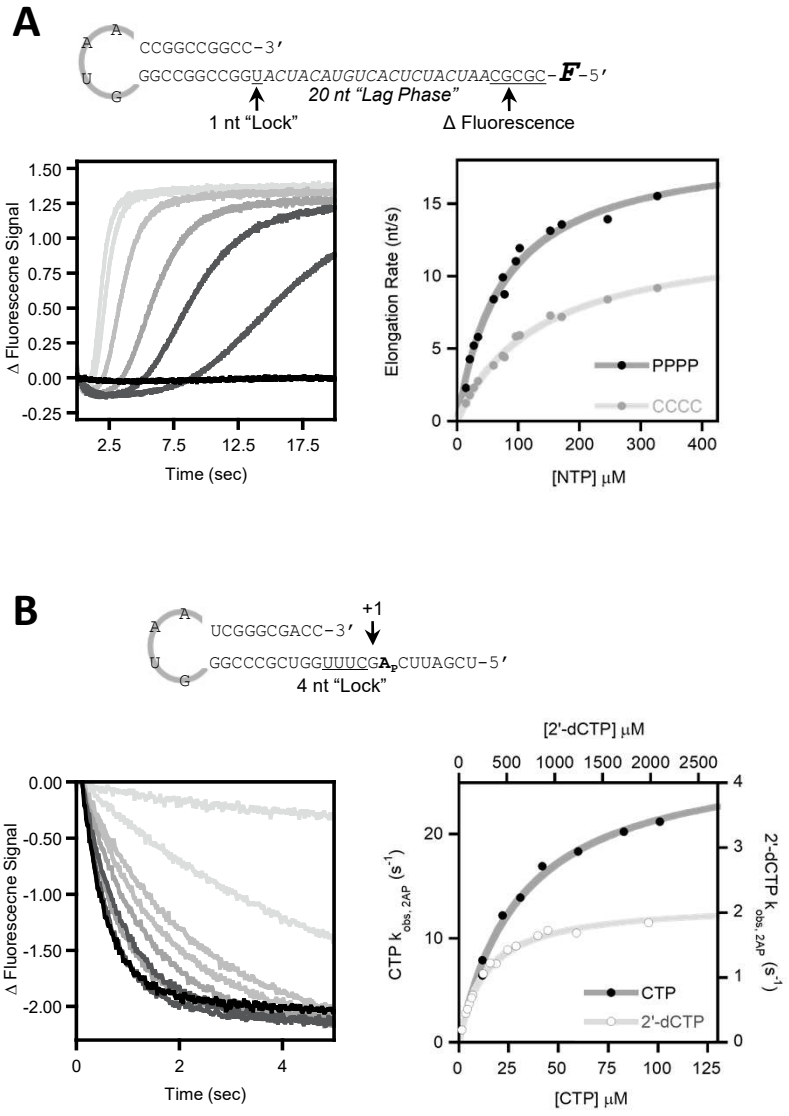

Figure 3.3: Elongation rate and nucleotide discrimination determination.

(A) RNA template used for processive elongation assay, where a single UMP incorporation was used to "lock" stalled elongation complexes prior to stopped flow assays of processive elongation rates. Sample data at left show an NTP titration series where the initial lag phase before the polymerase reaches and affects the 5′-fluorescein group shortens as NTP concentrations are increased, and the concentration dependence plot at right highlighting the difference in processive elongation rates between poliovirus and coxsackievirus polymerases. (B) RNA template designed for single CMP incorporation and nucleotide discrimination assay wherein a four-nucleotide "lock" for EC formation is done prior to using rapid stopped-flow fluorescence to monitor the translocation of the internal 2-aminopurine (2AP) base following incorporation of a single cytosine base at different concentrations. Right panel is a double-XY plot showing concentration dependence of rates for both CTP (dark gray) and 2′-dCTP (light gray) substrates.

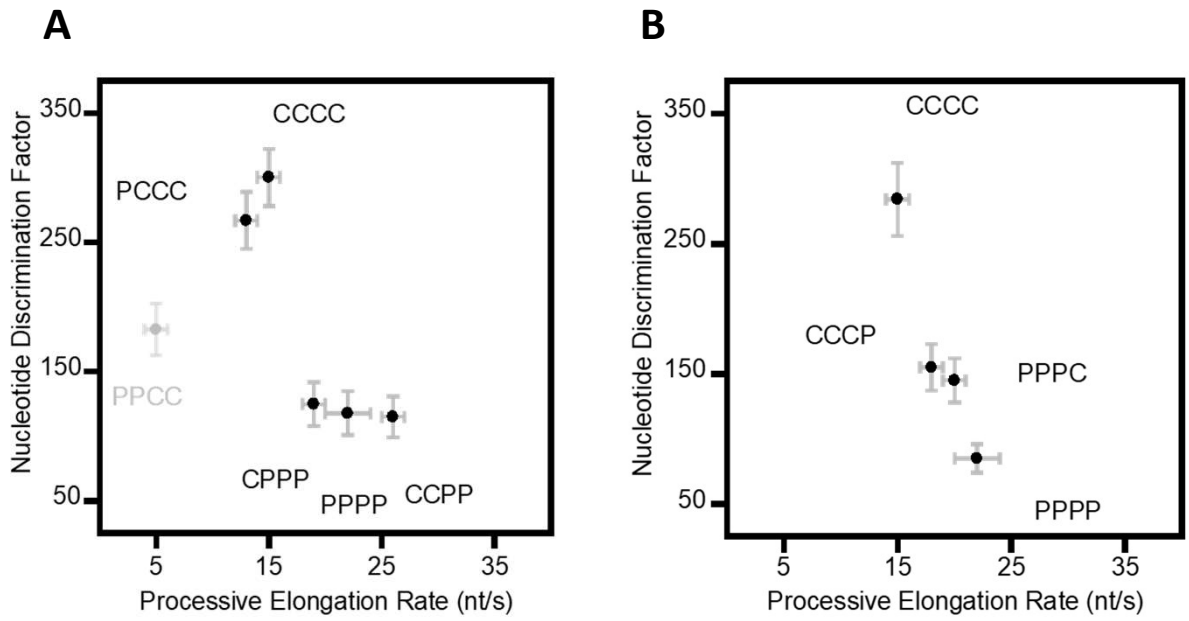

Figure 3.4: Nucleotide discrimination factor versus processive elongation rate plots.

(A) Pinky and full fingers exchanges showing that the chimeric polymerases cluster according to their palm domains. The PPCC construct is shown in gray to reflect it is an unstable protein that is prone to aggregation. (B) Data from the two thumb exchange polymerases showing elongation rates and nucleotide discrimination factors that are intermediate between the two wildtype enzymes. Note this experiment was done at a slightly lower pH (6.8 versus 7.0) than those in panel A, and this slightly reduces the observed discrimination factors for both wildtype and chimeric polymerases.

(CCPP) onto PV $3D^{pol}$ has only minor effects on discrimination compared to the wildtype enzyme (PPPP), and this is also true for the converse PCCC versus CCCC pinky exchange. The PPCC protein was prone to aggregation, suggesting its folding is compromised, and it is identified with light grey text in the figures.

The thumb domain exchanges, on the other hand, resulted in processive elongation rates and discrimination that were intermediates between those of the two wildtype enzymes (Figure 3.4B). Placing the coxsackievirus thumb onto poliovirus polymerase (PPPC) decreased the processive elongation rate, and conversely placing the poliovirus thumb onto coxsackievirus polymerase (CCCP) increased the processive elongation rate. The single cycle elongation data using 2-aminopurine translocation as a reporter showed similar k_{pol} values of ~28 s^{-1} for the two thumb exchanged enzymes, slightly elevated nucleotide K_M values compared to their parental palm and fingers domains, and nucleotide discrimination factors that were intermediate between those of either wild type polymerase (Table 3.1). Note the assays for the thumb exchanges were inadvertently done at pH 6.8 instead of 7.0 and this slightly reduces the discrimination factors for both wildtype and chimeric enzymes.

3.3.4 Rapid quench kinetics

The intermediate rates and discrimination factors observed for the thumb exchange polymerases led us to further dissect their mechanism by rapid chemical quench methods to directly monitor RNA product formation. A single RNA substrate (Figure 3.5A) was used for *i*) chemical quench with EDTA to assay the NTP capture step, *ii*) acid quench to assay the active site closure and chemistry steps via product formation rates, and *iii*) stopped-flow fluorescence to assay the overall NAC by reporting on the final translocation step (Figure 3.5B-D). The experiments were done at 30°C with Bio-Logic QFM-4000 and SFM-4000 instruments that use

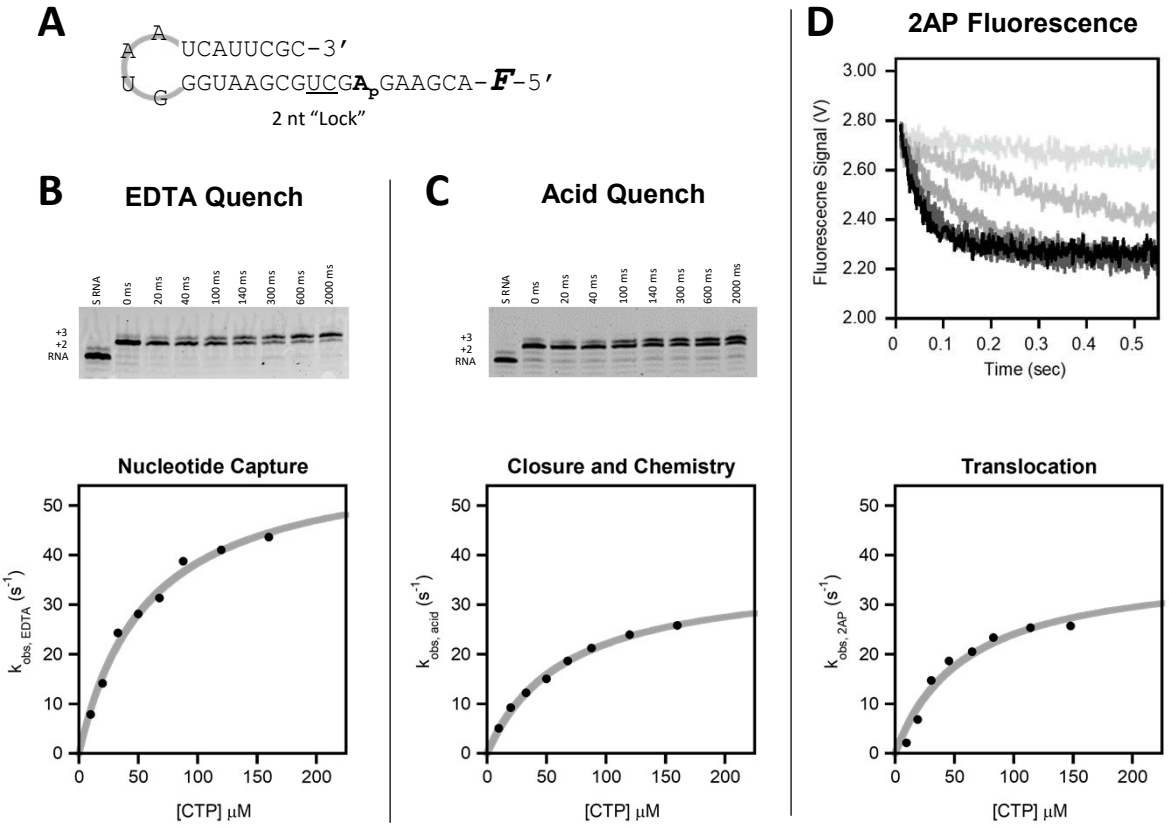

Figure 3.5: Rate and catalytic efficiency determination for individual steps of nucleotide addition cycle.

(A) RNA hairpin construct used for rapid chemical quench and stopped flow experiments. Briefly, stable EC was formed using a two-nucleotide "lock", the 5′-end fluorescein was used for quantitation of gel electrophoresis bands, and the internal 2-aminopurine was used for stopped-flow analysis of single CMP incorporation. Nucleotide capture parameters were determined by EDTA-quench, active site closure and chemistry were determined by acid quench, and overall catalytic cycle via 2AP translocation was determined using stopped-flow. (B) Example EDTA rapid quench data showing starting RNA, +2 "locked" EC, and formation of +3 species over time at a single CTP concentration. Corresponding concentration dependence plot for all CTP concentrations tested is shown below. (C) Example acid-quench data at a single CTP concentration with respect to time with concentration dependence plot shown below. For both panels B and C, note there is a minor impurity band running above the starting and +2 RNA species bands and this was corrected for in the data analysis by subtracting a value corresponding to 4% of the observed +2 band intensity. (D) Fluorescence trace for an example CTP titration series and the corresponding concentration dependence plot.

Table 3.2: Single NTP incorporation kinetics.

Values for rates and catalytic efficiency were determined directly, K_Ms were derived by dividing the rate by catalytic efficiency.

$3D^{pol}$	Nucleotide Capture			Active Site Closure and Chemistry			Complete Catalytic Cycle		
	Catalytic Efficiency (μM^{-1} sec^{-1})	k_{EDTA} (sec^{-1})	K_M (μM)	Catalytic Efficiency (μM^{-1} sec^{-1})	k_{acid} (sec^{-1})	K_M (μM)	Catalytic Efficiency (μM^{-1} sec^{-1})	k_{2Ap} (sec^{-1})	K_M (μM)
PPPP	1.8 ± 0.2	56 ± 3	31 ± 4	0.94 ± 0.08	42 ± 2	44 ± 5	0.8 ± 0.1	41 ± 4	50 ± 8
PPPC	1.5 ± 0.1	56 ± 2	38 ± 3	0.67 ± 0.07	37 ± 3	55 ± 7	0.59 ± 0.07	31 ± 3	53 ± 8
CCCP	1.06± 0.09	60 ± 4	57 ± 6	0.56 ± 0.02	37 ± 1	14 ± 1	0.6 ± 0.1	38 ± 5	59 ± 13
CCCC	1.1 ± 0.1	59 ± 5	56 ± 8	0.57 ± 0.07	35 ± 3	51 ± 4	0.56 ± 0.09	33 ± 4	59 ± 10

the same syringe systems for reagent delivery. The resulting k_{obs} values were plotted against NTP concentration and fit to the equation $k_{obs} = k\,[NTP] / ((k/(CE) + [NTP])$ to extract the catalytic efficiency, $CE \pm \sigma_{CE}$, and rate constant, $k \pm \sigma_k$, and their standard errors directly from the data (Table 3.2). Using the standard definition of CE as k/K_M then allows K_M values to be calculated as k/CE, with its error (σ_{K_M}) obtained by standard root mean square error propagation.

Fitting these data generally required a double exponential function, of which only the faster phase showed NTP concentration dependence reflecting 70-80% of the observed amplitude. In a prior coxsackievirus $3D^{pol}$ study we also observed biphasic 2-aminopurine fluorescence data and attributed the constant phase to a slow ($k \approx 1\ s^{-1}$) final relaxation event within the post-translocation complex that slightly altered the 2-aminopurine environment (Karr and Peersen, 2016). Here we now observed the same rate as a slow trailing phase of product formation in both the EDTA- and acid-quench experiments (Figure 3.6). Based on this, we conclude that this slower rate in fact represents a minor population of stalled elongation complexes that must undergo an initial slow transition event before becoming competent for NTP binding and catalysis, at which point they presumably progress through the NAC normally. The molecular nature of this event is not known, but it may be related to a conformational change step that occurs after RNA binding and before catalytic competence (Arnold and Cameron, 2000).

EDTA-quench derived rates that reflect nucleotide capture are very similar at 56±3 and 59±5 s^{-1} for the wildtype PV and CV enzymes (Table 3.2), and these are not significantly altered by the thumb domain exchanges that yield rates of 56±2 and 60±4 for PPPC and CCCP, respectively. In contrast, the acid quench data show that elongation product formation occurs more slowly at rates of 42±2 and 35±3 s^{-1} for the wildtype PV and CV enzymes, respectively,

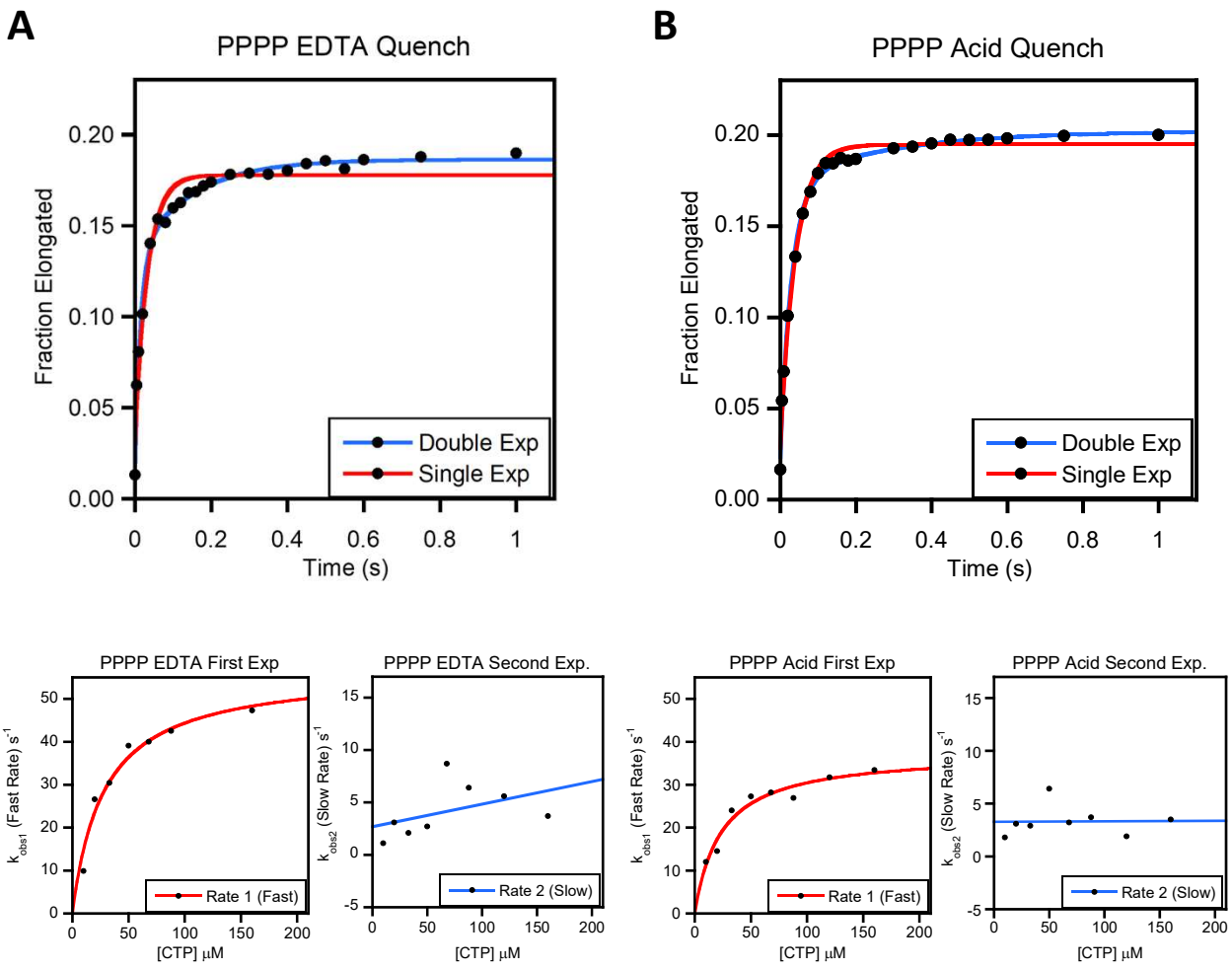

Figure 3.6: Identification of a second population in both EDTA and acid quench assays.

(A) Comparison of EDTA quench for poliovirus wild type polymerase with 160 uM CTP fit to a single exponential, red, or double exponential, blue. Below, concentration dependence for each rate, note the clear concentration dependence for the faster determined rate, and slight concentration dependence for the second rate (fit with a linear equation). (B) Acid quench data for poliovirus wild type for 160 uM CTP, and subsequent concentration dependence plots.

and in this case both PPPC and CCCP thumb exchanged enzymes have an intermediate rate of ~37 s^{-1}. Thus, the CV thumb slows product formation by PV polymerase, and conversely the PV thumb accelerates product formation by CV polymerase. These observations are echoed and exacerbated in the translocation dependent k_{2AP} values that reflect completion of the NAC; the PV polymerase rate is further slowed from 41±4 to 31±3 s^{-1} and the CV polymerase rate increases from 33±4 to 38±5 s^{-1}.

The wildtype PV polymerase has higher catalytic efficiency than the CV enzyme in all three assays, which is primarily a reflection of higher affinity nucleotide binding. This is most apparent in EDTA-quench data that is sensitive to the initial NTP capture event, where the two polymerase efficiencies differ 1.6-fold, but the trend of higher PV efficiency is also observed in the acid-quench and 2-aminopurine fluorescence assays where the thumb exchange reduces the PPPC efficiencies to be comparable to those of the wildtype CV. These data indicate the biochemical functions of the palm and thumb domains are tightly coordinated to maximize the speed of the nucleotide addition cycle after the rate limiting chemistry step.

3.3.5 Two-cycle elongation assays

To further investigate the different effects of palm and thumb domains on NAC rates, we carried out stopped-flow fluorescence studies with an RNA that was sensitive to two nucleotide incorporation steps (Figure 3.7A). The $3D^{pol}$ elongation complex structures show that the downstream +3 base on the template strand is partially stacked on the further downstream +4 base, and as a result its fluorescence will be quenched compared to the fully unstacked +2 position (Gong et al., 2013). This allowed us to design an RNA wherein a 2-aminopurine would start in the low-fluorescence +3 site, transition through the high-fluorescence +2 site, and finally enter the low-fluorescence +1 site where elongation is effectively terminated by a lack of rapid

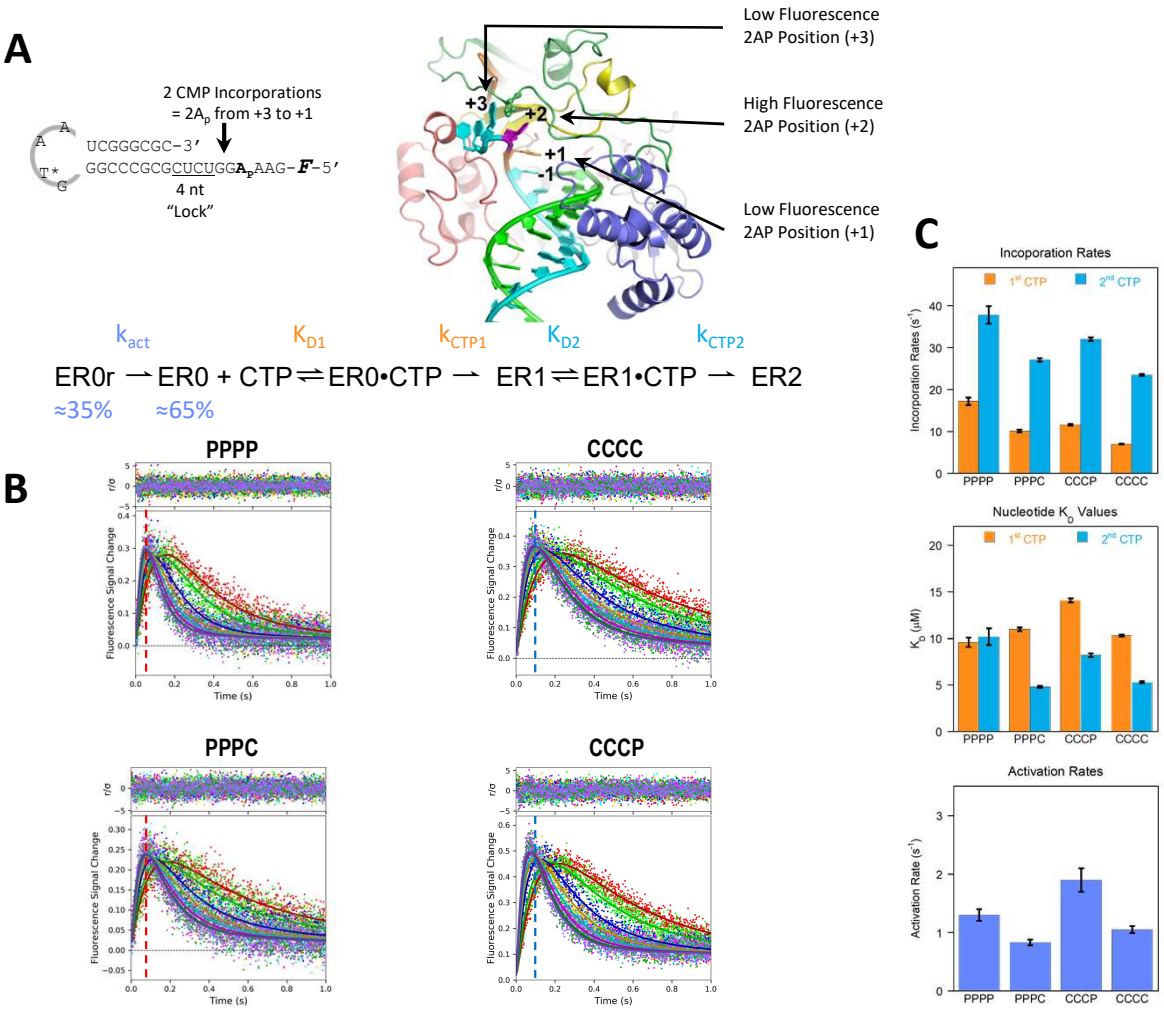

Figure 3.7: Kinetic modeling of double CMP incorporation reaction.

(A) RNA construct used for two CMP incorporation steps studied by stopped-flow 2-aminopurine fluorescence. The elongation complex structure shows the different environments for the 2-aminopurine as it is translocated from the +3 to the +2 site and then to the +1 site after the two CMP incorporation cycles. (B) Kinetic model used for analysis by global fitting in the program Kintek Explorer. Each nucleotide addition cycle incorporation step is fitted to a CTP dissociation constant (K_{D1}, K_{D2}) and an elongation rate constant (k_{CTP1}, k_{CTP2}), and a slow activation step (k_{act}) is used to account for ~35% of the 20 nM total locked elongation complexes being in a relaxed state (ER0r) at the beginning of the stopped-flow time course (see text). (B) Fluorescence traces showing raw data with simulated traces and associated residuals from the fitted kinetic model for the two wildtype (PPPP, CCCC) and thumb exchanged (PPPC, CCCP) polymerases. Data were collected at eight different CTP concentrations, the curves reflect KinTek Explorer model obtained by a global fit of the data, residuals are shown above each data plot, and vertical dashed lines are visual aids that reflect the time position of peak fluorescence for each wildtype polymerase (PV in red, CV in blue). (C) Bar graphs showing rate constants for the overall nucleotide addition cycle, CTP K_D values, and activation rates of the four polymerases. Error bars represent standard errors of the parameters obtained from the global fits.

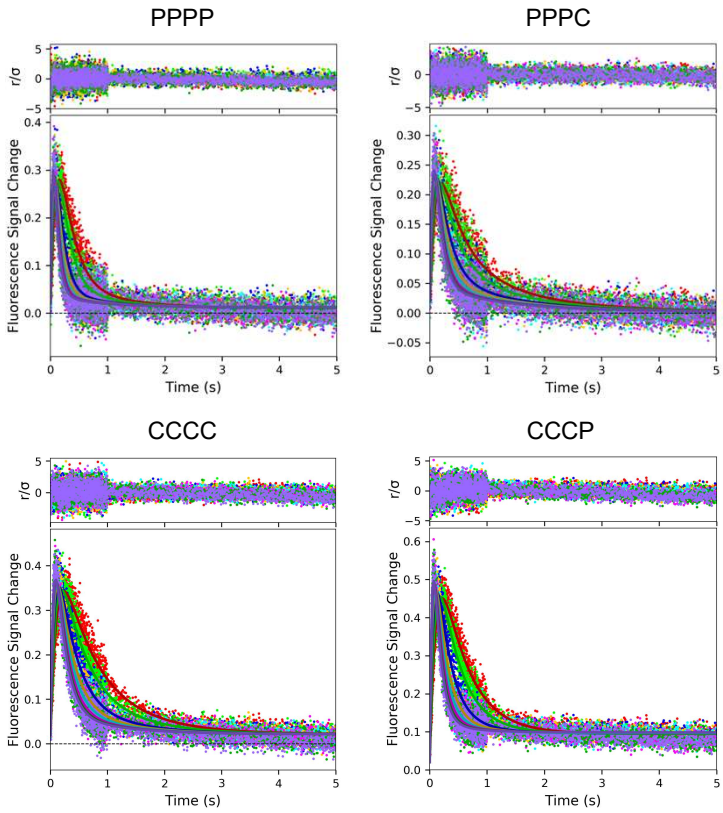

Figure 3.8: Extended time scale for two CMP incorporation.

Full data set collected for kinetic modeling. Extended time scale shows a clear slow tailing of the fluorescence signal as it returns to the base line.

nucleotide incorporation opposite the 2-aminopurine template. Initiation was done via the incorporation of a four nucleotide GAGA sequence to generate stalled elongation complexes and the stopped-flow fluorescence data reflect two sequential cytosine incorporation events. Data were analyzed using KinTek Explorer (KinTek Corp, Austin TX) (Johnson et al., 2009b, 2009a) with a double incorporation model where two CTP binding equilibria and two translocation rates were fit simultaneously. The explicit modeling of the full datasets also required inclusion of the slow activation step discussed above for a portion of the stalled elongation complexes, and it is this population that give rise to the slow trailing decrease observed in the fluorescence traces ($k \approx 1s^{-1}$). The fraction of total elongation complex in the initially "inactive" state represented 30-40% of the total 20 nM complex used in the reaction and was determined empirically for each dataset because it varied for each initiation reaction.

Figure 3.7B shows a portion of the data obtained from the two wildtype and two thumb exchanged polymerases as the CTP concentration was titrated from 6 uM (red trace) to 46 uM (purple trace), with the full data set shown in Figure 3.8. Each titration series dataset was collected using the variable volume mixing capabilities of the SFM-4000 instrument in ~10 minutes from initial instrument loading, which helps maintain consistent signal amplitudes by minimizing decay of the stalled elongation complex during the experiment. A visual comparison of the data traces from the two wildtype polymerases shows that PV $3D^{pol}$ is faster than CV $3D^{pol}$; at the highest CTP concentrations with near maximal rates, PV $3D^{pol}$ reaches peak fluorescence in ≈60 ms versus the slower ≈90 ms needed for CV $3D^{pol}$. The thumb exchanges affect this rate, slowing PPPC relative to the wildtype PPPP enzyme, and accelerating CCCP relative to CCCC (Figure 3.7B). Fitting these data to a double elongation model yielded K_M and k_{pol} values for both nucleotide incorporation cycles (Figure 3.7C). For all four polymerases, the

second nucleotide incorporation was more efficient than the first and this was primarily driven by 2-3-fold faster k_{pol} rates. Wildtype PV was faster than wildtype CV polymerase with rates for the second event of 38 s^{-1} versus 24 s^{-1} and exchanging the thumb domains essentially reversed this relationship, with PPPC slowing to 27 s^{-1} while CCCP became faster at 32 s^{-1}. The effects on CTP K_D values were relatively minor, ranging from 5–15 uM among all the constructs, and unlikely to have significant effects on elongation rates at millimolar physiological nucleotide concentrations. Last, the activation rates for the slow population of stalled elongation complexes ranged from 0.8 to 1.9 s^{-1}. These changed slightly with the thumb exchanges and appear to be somewhat faster with the PV thumb, although that is not a definitive conclusion from this limited dataset. We did also attempt to fit a more comprehensive kinetic model involving discrete nucleotide binding, chemistry, and translocation steps using global fitting of the EDTA, acid quench, and fluorescence datasets described in Figure 3.5, but could not get satisfactory convergence of all the kinetic parameters involved.

3.3.6 Infectious virus studies

We tested if the domain exchanged polymerases could support virus growth by inserting their coding sequence into both poliovirus and coxsackievirus genomes, generating viral RNA by T7 transcription, and transfecting that RNA into HeLa cells. We did not recover any virus using the coxsackievirus background, but in the poliovirus background we recovered infectious virus from both the pinky finger (CPPP) and full fingers domain (CCPP) exchanged polymerases (Figure 3.9A). The titers of virus from the initial RNA transfections were ~10 to 100-fold lower than those of the wildtype virus ($5x10^5$-10^6 versus 10^8 PFU/ml), but upon subsequent reinfection the domain exchanged viruses replicated as well as wildtype poliovirus. Sequencing showed the

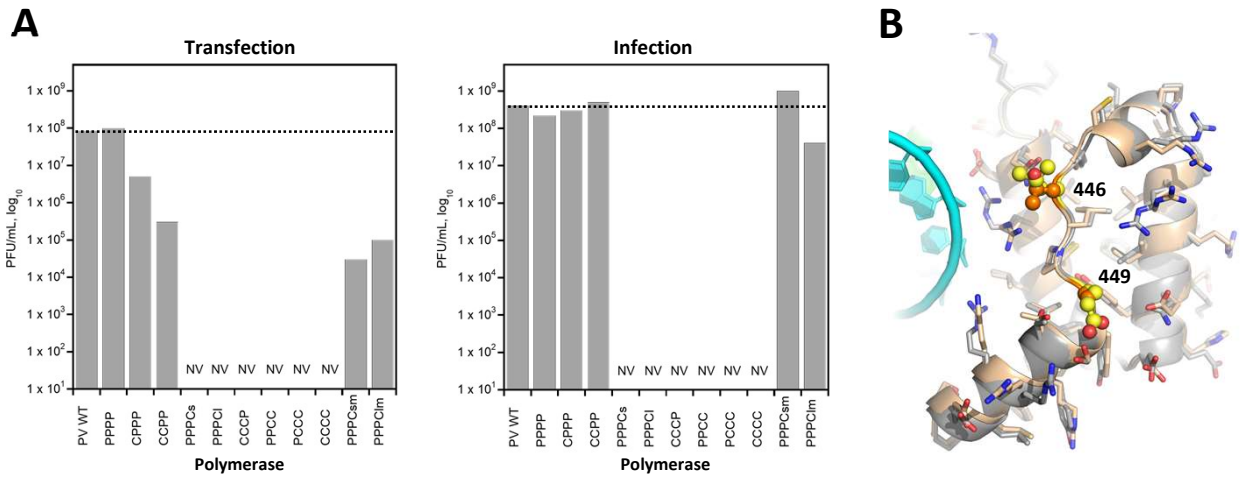

Figure 3.9: Chimeric virus growth analysis.

Viral titers obtained from poliovirus genomes carrying the chimeric polymerases upon (A) RNA transfection and (B) subsequent infection of cells with chimeric polymerase viruses (NV, nonviable). Polymerases containing both the palm and thumb domains from poliovirus were the only viable viruses from the original set of chimeras. Interestingly, virus was recovered from thumb exchange PPPC polymerase when its coxsackievirus thumb was mutated to make it more poliovirus-like (PPPCm), suggesting the thumb may be involved in enterovirus species-specific protein-protein interactions, where the (putative) polymerase protein-protein interaction is distinct within species B (CVB3) and species C (PV) enteroviruses. (C) Structural superposition of the thumb domains from poliovirus (gray) and coxsackievirus (wheat) with the two residues mutated in the PPPCm chimeric polymerases shown with atom spheres. These residue pairs are Leu446 / D449 in PV (yellow) and T447 / A450 in CV (orange).

chimeric polymerase constructs were genetically stable without any additional mutations arising within their 3CD polyprotein regions.

To address the possibility that the domain exchange with the coxsackievirus thumb prevented virus replication by disrupting a known interaction between the PV palm and thumb domains (see Discussion), we made two mutations in the PPPC construct to make its coxsackievirus thumb more polio-like. These $T447L_{CV}$ and $A450E_{CV}$ surface mutations effectively restore the poliovirus L446 and E449 residues involved in the $3D^{pol}$–$3D^{pol}$ interface, resulting in mutant chimeric polymerases we call $PPPC_m$ (Figure 3.9B). Interestingly, introducing these two mutations restored virus growth, resulting in small-plaque phenotype viruses that titered at near wildtype levels after one passage in HeLa cells. We recovered viruses with the thumb domain exchange starting before ($PPPC_{lm}$) and after ($PPPC_{sm}$) the $3D^{pol}$ motif E primer grip region, indicating that this element can also be exchanged between viruses. In an attempt to recover spontaneous mutations that could further enhance virus growth, we carried out ten serial blind passages of viruses containing mutated coxsackievirus thumb domains under conditions that would potentially select for more rapidly replicating viruses. However, resulting $PPPC_{sm}$ and $PPPC_{lm}$ viruses retained small plaque phenotypes and were indistinguishable from P_1 viruses. Sequencing of 30 TOPO-TA clones for each virus also failed to show any minor population variants or additional mutations in either $3D^{pol}$ or $3C^{pro}$ coding regions.

Several chimeras were non-viable (NV) following initial transfections and infections, as no plaques were detected at P_0 and P_1 (Figures 3.9A and 3.9B). After 10 cycles of passage, neither CPE nor plaques were observed for the following NV chimeras: $PPPC_s$, $PPPC_l$, CCCP, PPCC, PCCC and CCCC. cDNA also failed to yield PCR products, indicating that neither cytopathic nor non-cytopathic virus were present (Smithee et al., 2015). These data reinforce

other studies of interspecies incompatibility when swapping enterovirus B and C polymerases (Bell et al., 1999). Notably, $T447L_{CV}$ and $A450E_{CV}$ mutations in the CV thumb domain of PPPC chimeras overcome one aspect of this interspecies incompatibility (Figure 3.9, $PPPC_{sm}$ and $PPPC_{lm}$).

3.4 Discussion

The picornaviral $3D^{pol}$ enzymes are generally considered the smallest viral polymerases and their core structure composed of palm, fingers, and thumb domains is highly conserved at the structural level despite sequence divergence. Structures of polymerases alone and their complexes with RNA have provided insights into the molecular architecture and mechanisms of the elongation complex, including the RNA path through the enzyme and the existence of a unique palm-domain based active site closure mechanism that can be used to fine tune replication fidelity. In the study presented here we use poliovirus and coxsackievirus B3 polymerases, two enzymes with inherently different elongation rates and nucleotide discrimination factors, to examine the extent to which different biochemical functions could be assigned to modular $3D^{pol}$ structural domains.

3.4.1 Pinky finger controls RNA binding

Initiation, measured as a combination of slow RNA binding followed by the faster first nucleotide addition, is primarily controlled by the pinky finger as exchanging this structural element between the two polymerases almost perfectly reverses their initiation times. From a structural perspective, the pinky finger forms the outer wall of the template RNA channel and it contacts the RNA both before and after the active site. Notably, the pinky finger contains the polymerase motif G sequence within residues 109-118 that *i*) fold into the major groove of the

nascent duplex, *ii*) interacts with the single-stranded template backbone via residues 114-115 that are opposite the +2 nucleotide binding pocket, and *iii*) ends with Tyr118 located at the opening of a putative extended RNA entry channel (Kortus et al., 2012). Folded above motif G there is an extended conformation for residues 124-139 that includes Arg128 and Lys133, both of which contact product-strand phosphate groups, and Lys127 which interacts with a template-strand phosphate. This latter interaction is interesting because it is stabilized by an intricate protein fold wherein motif G residues wrap around the aliphatic portion of the Lys127 side chain to position its NH_3^+ group for an electrostatic contact with the template-strand phosphate of the priming base pair. As such, this is a direct link between proper folding of the pinky finger and structural interactions that recognize and/or position the RNA in the active site. The Lys127 residue is conserved in enteroviruses and there is rapid reversion of an introduced K127A mutation, indicating the lysine is required for poliovirus replication (Kortus et al., 2012). The pinky finger also contains a conserved *cis*-Pro119 flanked by conserved Gly117 and Gly124 residues in a folding motif that may be poised for conformational changes via proline *cis/trans* isomerization (Thompson and Peersen, 2004), and a G117A mutation results in complete loss of electron density for residues 112 to 129, suggesting cooperative folding within the pinky finger (Thompson, 2006).

Based on this, we propose that motions and folding transitions within the pinky finger open the polymerase for template loading, and then lock the RNA into place for processive elongation. The pinky finger folding does not require the presence of nucleic acid, as shown by essentially identical conformations in the absence and presence of RNA for multiple picornaviral polymerases (Peersen, 2019a, 2017). Initiation kinetics could thus be governed by either the rate of pinky unfolding to open the template RNA channel, or by the rate of pinky refolding to lock

the RNA in place. We do not yet know which of these events explains the different rates of the PV and CV pinky fingers, but suspect the slower initiation rate for PV $3D^{pol}$ is due to a slow refolding event based on the higher B-factors observed in structures solved in the absence versus presence of RNA. Considering the pinky finger as a semi-independent functional unit also makes sense in the context of other RdRP structures: While picornaviral polymerase structures are almost identical, there is a distinct difference in the pinky finger whereby a surface helix, composed of residues 128-138 in PV, is rotated ~90° in the aptho-, cardio-, and kobuviral polymerases (Dubankova et al., 2019; Ferrer-Orta et al., 2004). This alternate helix rotation is also seen in noroviral and caliciviral polymerases, whose overall structures are quite similar to the picornaviral ones.

3.4.2 Thumb domain affects translocation

The pinky and full fingers exchanged chimeras both kept the palm and thumb domains together in a polymerase "core" and the elongation rates and nucleotide discrimination factors cluster according to this core, i.e. CPPP and CCPP are polio-like while PCCC and PPCC are coxsackie-like (Figure 3.4). It is only when we separate the palm and thumb with the PPPC and CCCP constructs that we get intermediate values for the elongation rates and discrimination factors. To investigate this further we used a combination of rapid chemical quench and stopped-flow fluorescence experiments (Figure 3.5, Table 3.2). Data from both the EDTA- and acid-quench experiments show that nucleotide incorporation rates and catalytic efficiencies follow the origin of the palm and fingers domains, while the translocation rates follow the origin of the thumb domain.

The EDTA quench experiments reflect the kinetics of NTP capture that ultimately results in product formation, even if subsequent steps needed to do chemistry are slow; once bound to

the polymerase, the NTP•Mg^{2+} complex is effectively immune to chelation of its Mg^{2+} ion by EDTA present in the bulk solution. Exchanging the thumb domains does not significantly alter the observed EDTA-quench rates, indicating that the thumb does not play a major role first step in the nucleotide addition cycle. Next, acid quench experiments were used to investigate the rate at which active site closure and chemistry for +1 product formation occurs. Unlike the EDTA-quench, this assay immediately stops the reaction through protonation of active site groups and protein denaturation. The acid-quench rates are slower than the capture step rates because they include the additional molecular motions needed to close the active site and perform chemistry. The PPPC thumb exchange reduced both the rate and catalytic efficiency compared to PPPP, indicating the thumb domain can exert an indirect effect on the chemistry step. This is likely due to altering the exact positioning and/or motions of the priming ribose in the active site. Last, the 2-aminopurine fluorescence data reports on template translocation, the last step in the NAC. The wildtype polymerases have k_{2AP} values that are only slightly slower than their k_{acid} rates, indicating that translocation is normally a rapid step following catalysis. In contrast, the thumb exchanges resulted in an almost perfect reversal of the observed rates. The double CMP incorporation data similarly showed that placing the PV thumb onto CV polymerase increased NAC rates by ≈40%, and conversely the CV thumb slowed PV polymerase by ≈40% (Figure 3.7). This two-cycle modeling also showed that for any polymerase, the second nucleotide addition was more efficient than the first, primarily due to 2-3-fold faster incorporation rates. This likely reflect subtle biochemical and dynamics changes in the $3D^{pol}$•RNA complex as it transitions to the more processive state of a true elongation complex.

Altogether, these kinetic data are consistent with previous studies showing that active site closure is the rate limiting step in the $3D^{pol}$ catalytic cycle (Arnold and Cameron, 2004).

Structures of $3D^{pol}$ elongation complexes show that this step involves a subtle movement within the palm domain whereby the C-terminal end of motif A moves to fully form the 3-stranded β-sheet with motif C and position the active site aspartates for catalysis (Gong and Peersen, 2010). This is associated with movements of motif D, which forms the outer edge of the NTP entry tunnel and interacts tightly with motif A via backbone hydrogen bonding and co-folding (McDonald et al., 2016; Shen et al., 2012a; Yang et al., 2012). All these interactions are within the palm domain, consistent with our data showing that this domain is the primary effector of both the nucleotide binding and chemistry steps. The new revelation from our thumb exchanged polymerases is that the thumb domain plays a role in controlling translocation, the final step in the nucleotide addition cycle.

3.4.3 Poliovirus replication studies

In a PV background we recovered viruses with near wildtype titers for both the CPPP and CCPP polymerases, showing that both the pinky and the full fingers could be exchanged without significant impact on viral replication. We initially did not recover any virus from the thumb exchanged polymerases, and due to the biochemical coupling between the palm and thumb domains we tested this with constructs where the palm–thumb junction point was located either before or after the motif E primer-grip element. However, an alternative explanation for this result can be found in the observation that the poliovirus thumb domain has been implicated in the formation of two different protein-protein interfaces. The first is with the palm of another $3D^{pol}$ to form long polymerase fibers (Hobson et al., 2001; Spagnolo et al., 2010) and sheet like structures (Lyle et al., 2002; Tellez et al., 2011; Wang et al., 2013), and the second is a biochemically identified interaction with the 3C domain of the viral $3CD^{pro}$ protease to stabilize the viral uridylylation complex (Shen et al., 2008). To further explore the significance of these

interfaces for virus growth, we made the coxsackie thumb of PPPC more polio-like by restoring two PV surface residues, L446 and E449 (Figure 3.9B). Interestingly, these two mutations resulted in near wildtype titers of chimeric $PPPC_{sm}$ and $PPPC_{lm}$ viruses, clearly indicating this region of the thumb domain surface is essential for poliovirus replication. We further attempted to obtain viruses with improved growth characteristics by serial passaging of the $PPPC_{sm}$ and $PPPC_{lm}$ viruses under conditions that would favor the emergence of variants carrying spontaneous mutations. The locations of such mutations could have validated the importance of either the $3D^{pol}$–$3D^{pol}$ or the $3D^{pol}$–$3CD^{pro}$ interface, but unfortunately none were obtained.

3.5 Conclusions

In this study we used the high similarity of poliovirus and coxsackievirus polymerases to generate chimeric $3D^{pol}$ enzymes and determine the extent to which different biochemical functions could be assigned to RdRP structural domains. The results showed that the pinky finger plays a key role in initiation while nucleotide discrimination and catalysis are primarily associated with the palm domain, and the thumb domain subtly impacts the translocation step and elongation complex stability. We propose a general model for positive-strand RNA virus polymerases wherein the pinky forms a semi-independent subdomain atop the fingers domain whose folding transitions regulate template strand access to the active site.

CHAPTER 4

A KINETIC ANALYSIS TO DETERMINE THE ROLE RESIDUE L420 PLAYS IN RECOMBINATION AND ERROR CATASTROPHE

Author contributions: All experiments in this chapter were done by me, as well as figure creation. Authors Kemfp and Barton originally provided us with the scientific need for this project, thus their inclusion in this work.

4.1 Introduction

A long-standing question in virology has been the mechanism by which viruses exchange genetic information through recombination and what is the biological purpose for this recombination (Simon-Loriere and Holmes, 2011). In particular with picornaviruses, a positive sense ssRNA virus, the prevailing theory has become recombination is used to combat error catastrophe (Aguado et al., 2018; Crotty et al., 2002; Eigen, 2002). Picornaviruses have relatively small genomes, generally under 10 kb in length and accumulate 2 to 4 mutations per replication cycle (Drake and Holland, 1999; Jenkins et al., 2002; Kautz and Forrester, 2018; Sanjuán et al., 2010). This mutation rate is due to the low fidelity associated with the virally encoded RNA-dependent RNA polymerase, $3D^{pol}$. $3D^{pol}$ is responsible for both negative- and positive-strand synthesis during viral replication, and as such can incorporate mutations equally during synthesis of either RNA-strand. This variance in genome sequence leads to quasispecies formation during a viral infection (Dolan et al., 2018; Korboukh et al., 2014). While mutation is necessary for viral infection progression, in particular by aiding in viral infection of different cell types, many mutations can also be deleterious to the virus (Acevedo et al., 2018; Xiao et al., 2016). In the absence of genome recombination, viral populations become susceptible to the

accumulation of these detrimental mutations, ultimately driving the virus to error catastrophe by Muller's ratchet (Chao, 1990). Presently it is thought that recombination serves to counteract error catastrophe by facilitating the exchange of a mutated genome template with a different RNA, thus generating a new RNA product that has been purged of negative mutations and resetting the virus's molecular clock (Chao, 1997). While it has easy to justify the biological purpose of recombination, the mechanism of action is less straight forward. Presently, the prevailing theory how recombination is accomplished is by a copy-choice mechanism (Simon-Loriere and Holmes, 2011). During negative strand replication it is thought the polymerase exchanges RNA-templates, producing a product RNA derived from more than one parental strand, although how this template strand swap occurs has not come to a consensus. Studies have also shown that this template exchange does not always produce a clean junction genome, which requires resolution of duplicated regions as viral replication continues (Kirkegaard and Baltimore, 1986; Lowry et al., 2014). Despite the incomplete picture currently of recombination there is one key feature for each mechanism, the dependence on $3D^{pol}$ for this process to occur.

Multiple laboratories have set out to determine residues in $3D^{pol}$ from poliovirus and other picornavirus family members that play a role in genome recombination, and thus far several residues have been identified to causing changes in recombination frequency (Kempf et al., 2016; Li et al., 2019; Woodman et al., 2018b; Xiao et al., 2016). Our laboratory in collaboration with the Barton laboratory has identified one such mutation, L420A (Kempf et al., 2016). This residue is located on the thumb of $3D^{pol}$ and directly interacts with the product strand RNA backbone at the -3 position. Our data have shown that viral recombination is reduced ~50-fold in the presence of this mutation, rendering the virus more susceptible to error catastrophe induced by ribavirin incorporation. Outside of recombination, picornaviruses have developed

other strategies to combat error catastrophe, in particular by changing polymerase fidelity. G64S is a well-studied mutation within $3D^{pol}$ that has been shown to decrease ribavirin incorporation into the viral genome during replication (Pfeiffer and Kirkegaard, 2003). Interestingly, our previous work showed that despite the G64S mutation to $3D^{pol}$, the L420A mutation overpowers it and renders the G64S L420A double mutant susceptible to error catastrophe (Kempf et al., 2019).

To investigate this finding, we expanded our investigation to determine the role L420A plays in the polymerase catalytic cycle in both the presence and absence of G64S (Figure 4.1). By dissecting the catalytic cycle into individual steps, we determined that L420A and G64S affect not only different steps within the cycle but that they are orthogonal to each other when it comes to polymerase fidelity during genome replication. By looking at each of these components of genome replication individually and in conjunction with each other, we have tried to elucidate the mechanism as to how recombination effecting residues and fidelity effecting residues are both intimately linked and opposed to each other during the process of genome replication.

4.2 Materials and Methods

4.2.1 Nucleotide discrimination assay

$3D^{pol}$ mutants were purified as described previously. Nucleotide discrimination was determined using stopped-flow fluorescence readout detailed in Kempf 2019 (Kempf et al., 2019).

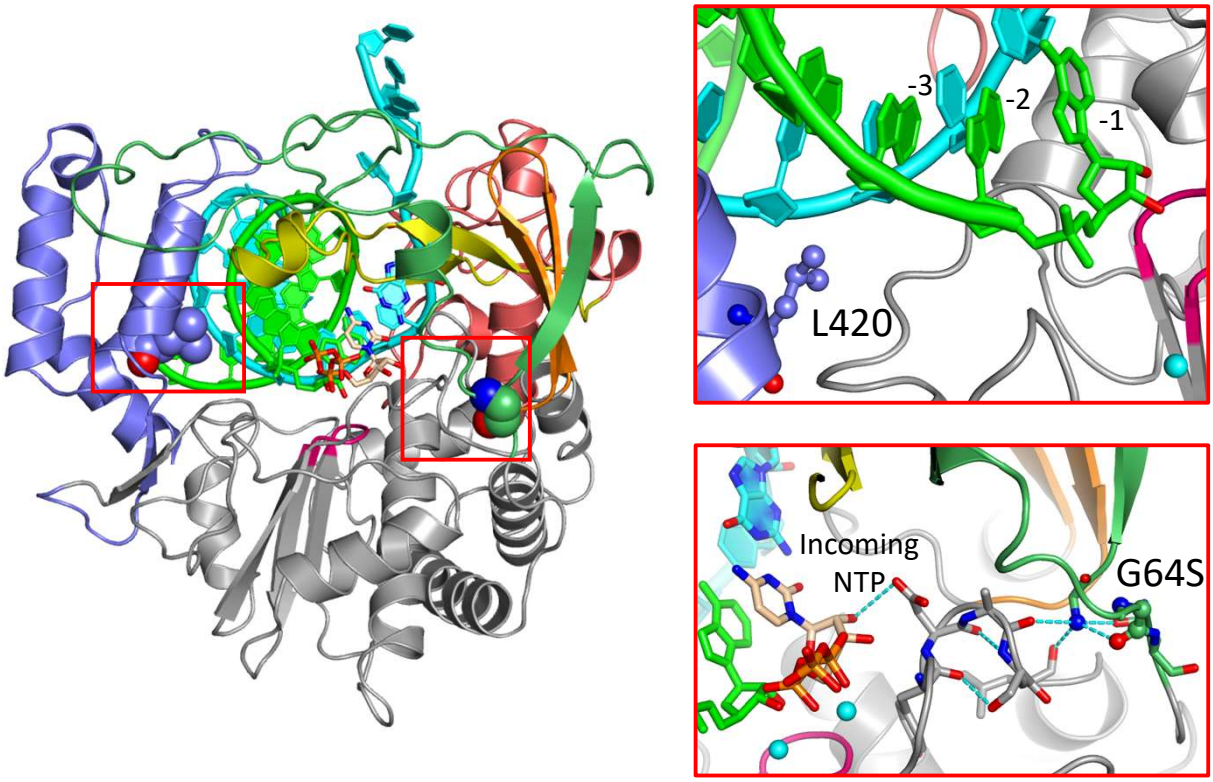

Figure 4.1: Location of residues G64 and L420 within the 3D^{pol} elongation complex.

Poliovirus elongation complex structure showing the fingers domain, right, containing residue G64 located on the index finger, and thumb domain residue L420. Note the positioning of L420 to the product strand backbone at base position -3, upper, and the relative position of G64S to the incoming nucleotide, lower.

4.2.2 Rapid chemical quench assays

Rapid EDTA quench experiments were done by mixing 15 uM polymerase and 10 uM "8+2_3_5 RNA" in buffer containing 50 mM NaCl, 4 mM MgCl2, 25 mM HEPES pH 6.5, 2 mM TCEP, 200 uM ATP, and 60 uM GTP. The "8+2_3_5 RNA" is based off previously used RNA hairpin constructs (Hobdey et al., 2010) and contains a 8 nucleotide downstream base paired portion and two nucleotide lock to form the elongation complex which is needed for further experiments. An internal 2-aminopurine was included for stopped-flow experiments and 5′-fluorescein for gel visualization of rapid quench reactions. Reactions were incubated for 15 minutes at room temperature to generate a stalled polymerase-RNA elongation complex. Reactions were diluted 20-fold in a buffer with 50 mM HEPES pH 7, 75 mM NaCl, and 4 mM MgCl2 to a final concentration of 250 nM. Kinetics data was collected using a Bio-Logic QFM-400 rapid quench instrument. Each reaction was aged for 5-2000 msec before EDTA-quench to a final concentration of 50 mM. Samples were visualized by denaturing gel electrophoresis using 20% polyacrylamide 19:1, 7 M urea and 1X TBE, and imaged using a GE Typhoon FLA 9500. Rapid acid-quench reactions were set up as described above before quench after aging to a final HCl concentration of 1 M. Reactions were neutralized by addition of 1 M NaOH to achieve a neutral pH before samples were visualized by denaturing gel electrophoresis, described above.

4.2.3 Nucleotide misincorporation assay

Nucleotide misincorporation was modified from a previously described GMP incorporation assay (Campagnola et al., 2015). Briefly, 30 uM polymerase was mixed with 10 uM "8+4_3_3" RNA in buffer containing 50 mM NaCl, 4 mM MgCl2, 25 mM HEPES pH 6.5, 2 mM TCEP and 40 uM GTP. Reactions were incubated at room temperature for 15 minutes to allow for single GMP incorporation and stalled elongation complex formation. Samples were

next diluted 10-fold into dilution buffer containing 300 mM NaCl 4 mM MgCl2, 25 mM HEPES pH 6.5, 2 mM TCEP, and 100 uM to 1 mM GTP. Reactions were incubated at room temperature for 60 minutes with time points taken throughout the reaction. At each timepoint 5 uL of polymerase reaction was quenched in 15 uL quench buffer containing 50 uM EDTA, 400 mM NaCl, 50 uM HEPES pH 6.5, and 2 mM TCEP. An elongation complex competency measurement was also made for each timepoint. 5 uL of polymerase reaction was mixed with 5 uL of chase buffer containing 100 uM ATP in addition to the dilution buffer components. Reactions were incubated for 2 minutes at room temperature before 10 uL of quench buffer was added. Samples were quantitated by denaturing gel electrophoresis as described for the rapid chemical quench assays.

4.3 Results

4.3.1 Nucleotide discrimination analysis

As part of our investigation in to characterizing how L420A affects the nucleotide incorporation cycle and the interplay between that and viral genome recombination, we developed two additional nucleotide discrimination assays to determine if the pattern we saw in CTP vs 2′-dCTP discrimination factor held true for purine addition, presented in Kempf 2019 (Kempf et al., 2019), as a way to mimic ribavirin (RTP) incorporation. We observed that the changes in discrimination factor did in fact hold for purine addition, suggesting that our results were a viable proxy for ribavirin incorporation. As was previously reported, L420A had the lowest nucleotide discrimination, followed by WT, and G64S and G64S L420A both had the highest discrimination factors that were comparable to each other.

Further analysis of the concentration dependence plots for each NTP or 2′-dNTP addition showed an interesting result. The changes in determined discrimination factor value are derived from k_{cat} and not K_M, thus making nucleotide discrimination a product of polymerase rate and not nucleotide affinity. As a whole NTP K_M did not significantly vary between polymerases, Figure 4.2A. 2′-dNTP K_Ms also did not show much variation with the exception of 2′-dCTP, with G64S and G64S L420A both having decreased K_Ms when compared to WT. An unexpected finding with this analysis was that there is very little K_M difference between GTP and 2′-dGTP as well as between ATP and 2′-dATP. We also determined that the K_M for GTP is between 1 and 3 uM, almost 10X lower than that of CTP determined under the same conditions. The determined K_M values for ATP did not significantly vary from the values determined for CTP, however the K_M values for 2′-dATP incorporation were almost 10-fold lower than that of 2′-dCTP.

Nucleotide incorporation rates all showed the same trend regardless of which NTP or 2′-dNTP was being incorporated, Figure 4.2B. For NTP incorporation, L420A always had a higher k_{cat} than WT. G64S L420A had a higher incorporation rate than G64S alone, but was slower than WT. This incorporation rate change was also observed for 2′-dNTP incorporation, although the faster rate for L420A was accentuated. The G64S mutation, in either the absence or presence of L420A, exacerbated the rate decrease compared to WT for 2′-dNTP addition. Rates for GTP and 2′-dGTP incorporation were much slower than either CTP or 2′-dCTP, a 5-fold or 10-fold decrease, respectively. ATP and 2′-dATP addition rates also showed some changes incorporation rate. While the ATP rate relative to the CTP rate did not show more than a 1.5- to 2-fold change, 2′-dATP rates were closer to that of 2′-dGTP incorporation than 2′-dCTP.

The catalytic efficiency for each polymerase for each NTP or 2′-dNTP followed the trends observed for k_{cat}, Figure 4.2C. For all polymerases, catalytic efficiency is highest for GTP

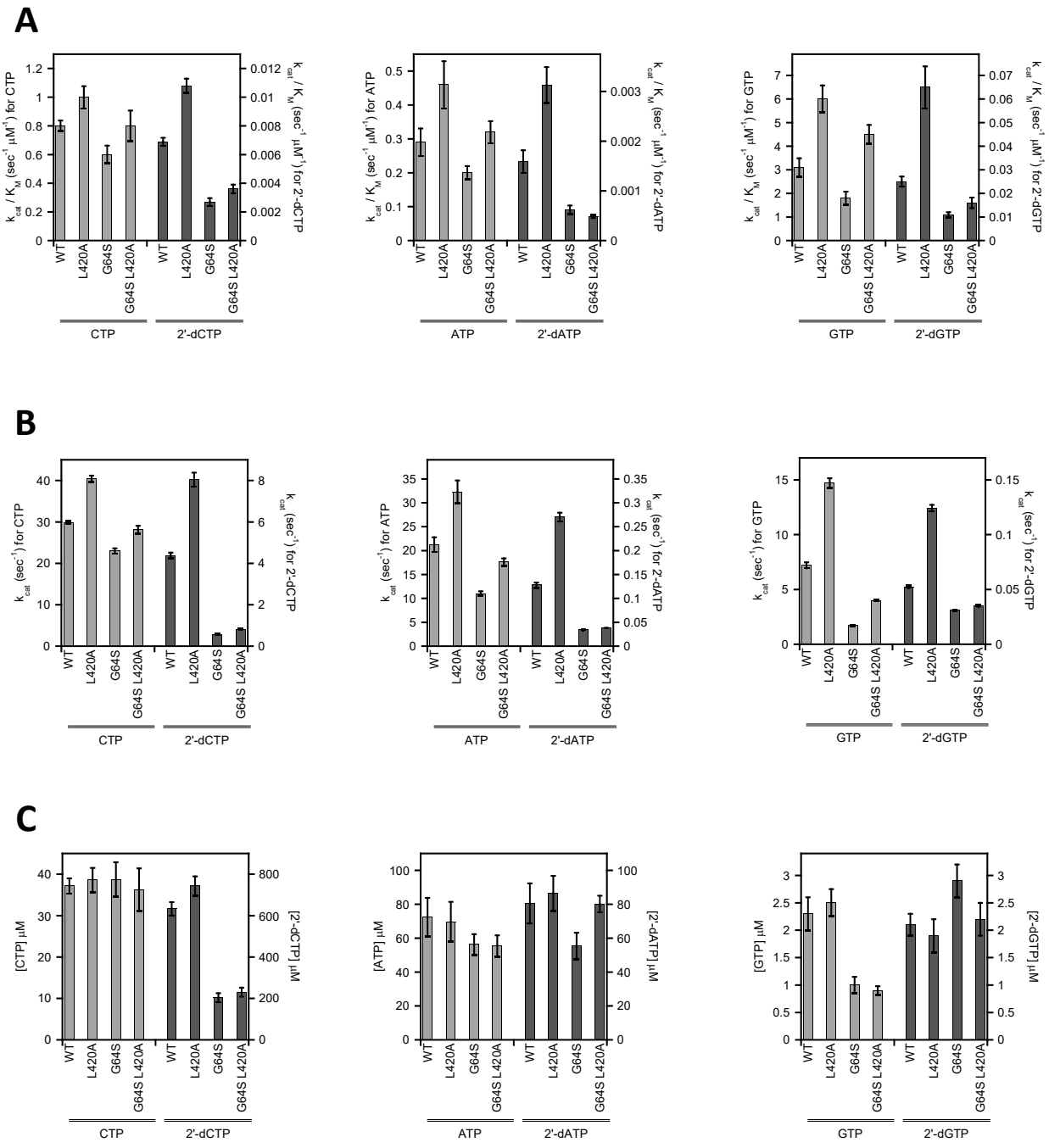

Figure 4.2: Kinetic parameters affecting NTP verses 2′-dNTP discrimination factors.

(A). Catalytic efficiency for nucleotide pairs Light gray denotes NTP, dark gray for 2′-dNTP. (B) Maximal complete catalytic cycle rate, k_{pol}, for respective nucleotide pairs. Note the decrease in rate for G64S containing mutants relative to WT, and rate increase for L420A. Variations in catalytic cycle rate follow the same pattern as catalytic efficiency. (C) Nucleotide K_M for respective NTP and 2′-dNTP pairs.

or 2′-dGTP, by approximately 7-fold for either substrate relative to CTP or 2′-dCTP. ATP or 2′-dATP incorporation had a decrease in catalytic efficiency of approximately 2-fold for either reaction relative to the respective CTP or 2′-dCTP reaction.

4.3.2 Rapid quench analysis

Rapid EDTA quench was conducted to assess NTP capture for each polymerase by preventing additional NTPs from binding once the reaction had been quenched, but not preventing bound nucleotides from completing the nucleic acid addition cycle. WT is fastest at nucleotide capture and has the highest catalytic efficiency for the reaction, Figure 4.3D. L420A had a nucleotide capture rate of 49 s^{-1}, lower than the rate determined for WT. L420A also had a catalytic efficiency approximately 2-fold lower than WT, Figure 4.3E. The G64S L420A mutant had a nucleotide capture rate of 44 s^{-1}, within error of the L420A mutation alone. However, unlike the L420A mutation alone, the catalytic efficiency for G64S L420A was within error of WT. G64S alone for nucleotide capture had both the lowest rate and catalytic efficiency for NTP capture.

Rapid acid quench interrogates the rates and catalytic efficiency for active site closure and chemistry. WT has a rate for these reactions of 32 s^{-1} and a catalytic efficiency of almost 1 uM^{-1} s^{-1}. Unlike the EDTA quench reactions, L420A had a faster active site closure rate than WT with a rate of 40 s^{-1} and a comparable catalytic efficiency, Figures 4.3D and 4.3E. G64S L420A had a slower rate for active site closure of 28 s^{-1} and a catalytic efficiency almost 2-fold lower than WT or the L420A mutation alone, differing significantly from the pattern seen with the EDTA quench. Once again G64S had the lowest rate and catalytic efficiency. Both of these values are much closer to those determined for G64S L420A.

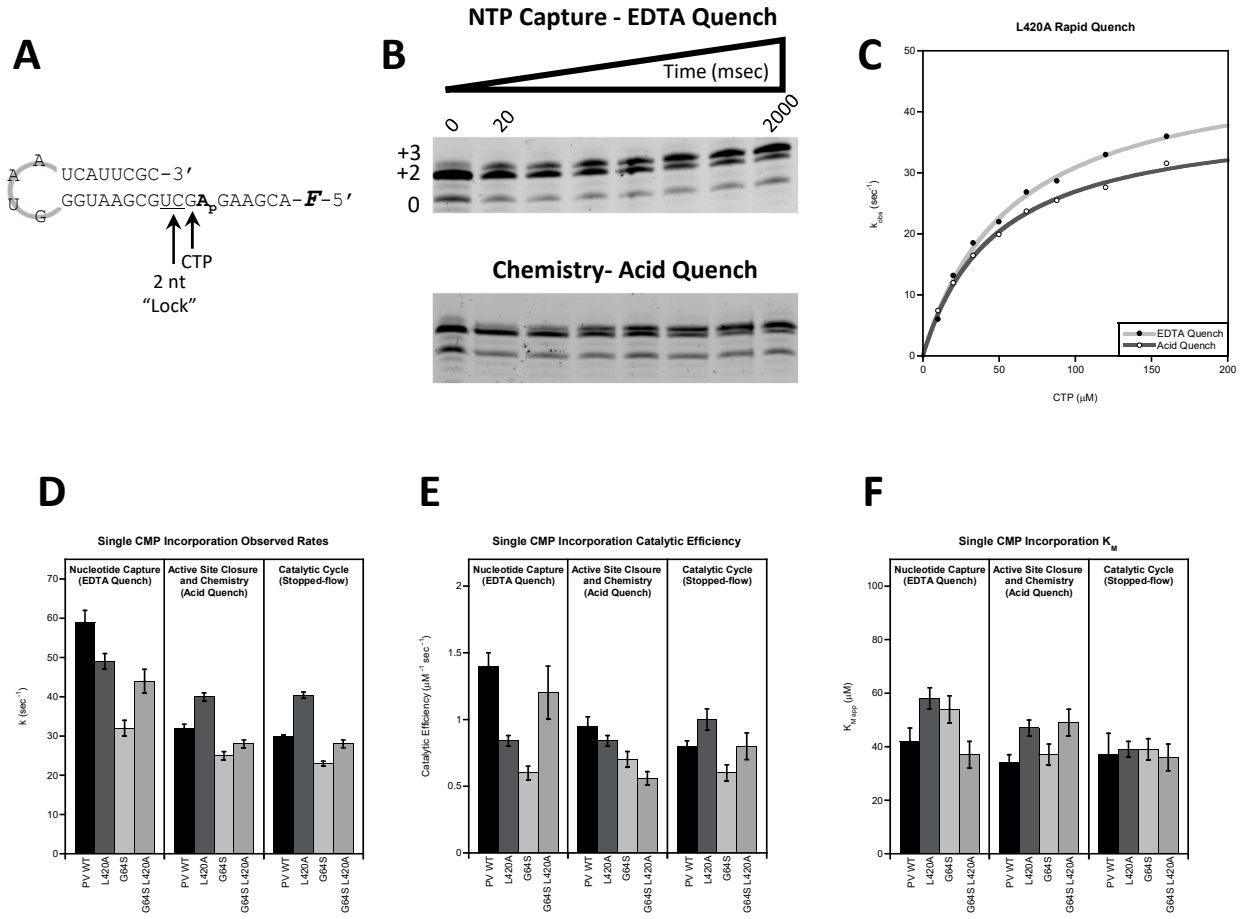

Figure 4.3: Mutation effects on single CMP incorporation for defined steps within the catalytic cycle.

(A) RNA hairpin construct. Elongation complex was achieved using a 2-nucleotide lock before CTP incorporation site. 5′-end fluorescein was used for gel visualization and internal 2-Ap was used to track catalytic cycle completion. (B) Representative quantitation gels for EDTA and acid rapid quench. (C) EDTA and acid quench concentration dependence plots. (D) CMP incorporation rates grouped by reaction step. (E) CMP incorporation catalytic efficiency for reaction steps. (F) CMP K_M for individual catalytic cycle steps.

To determine the rate and catalytic efficiency for a single nucleic acid addition cycle, 2-aminopurine, 2AP, translocation upon catalytic cycle completion was assessed using stopped-flow to measure fluorescence quenching as 2AP was translocated from the +2 to +1 position (McDonald et al., 2016). WT had a rate of 30 s^{-1}, a small decrease from the rate determined for active site closure using the acid quench, Figure 4.3D. L420A alone had a rate of 40 s^{-1}, almost identical to the rate determined by acid quench. G64S L420A also had a rate almost identical to the acid quench rate at 28 s^{-1}. Finally, G64S alone showed a small decrease in overall rate as compared to the acid quench rate at 23 s^{-1}. Thus, with the native L420 residue there is an additional small rate effect that is not observed when the L420A mutation is present. As was also seen in the acid quench reactions, L420A and WT have the highest catalytic efficiencies followed by G64S L420A, and G64S only had the lowest catalytic efficiency, Figure 4.3E.

4.3.3 GTP misincorporation

To further support our pervious claims that G64S L420A is indeed a highly discriminatory polymerase, we conducted a GTP misincorporation assay in which GTP is templated by UMP instead of the canonical CMP. These reactions are slow in comparison to either a native GTP incorporation or a 2′-dGTP incorporation. Because of this, we needed to consider the elongation complex stability, previously reported in Kempf 2019, and determined the amount of competent elongation complexes in the reaction at each time point as well as the complexes that had misincorporated a GMP into the product strand. This value was used to determine the corrected fraction of competent complexes at each time point that had incorporated GMP and is further detailed in the materials and methods, Figure 4.4A-4.4D. In this context the G64S L420A polymerase behaved like the L420A only mutant and had a high percentage of GMP misincorporation relative to WT. The G64S only polymerase had a lower percentage of

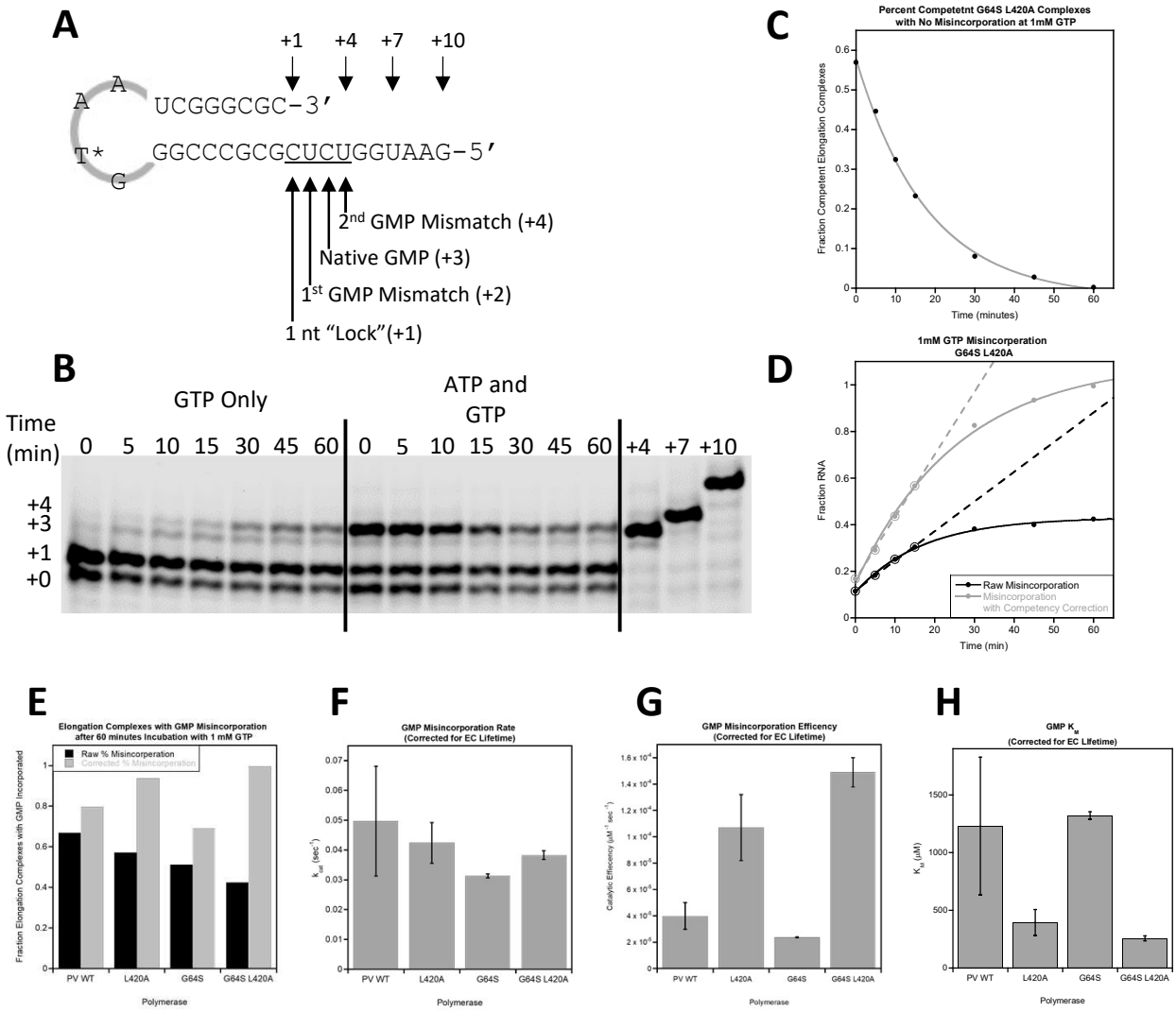

Figure 4.4: Different mutation effects on GMP misincorporation.

(A) RNA hairpin used to assess GMP misincorporation events. Elongation complex was formed at a low concentration of GTP before dilution into higher GTP concentration. Gel band notation and the correlated incorporation events are noted. (B) Gel quantification of GMP incorporations when reactions were incubated with GTP only or transferred to a solution containing both ATP and GTP. Native NTP incorporation products were run for band size verification. (C) Quantitation of decrease in elongation complex competency, determined by the difference in RNA total greater than +1 when incubated with ATP and GTP as compared to GTP only. (D) Gel species greater than +1 quantitation without elongation complex competency, black and corrected, gray. Early reaction timepoints were used to determine linear rate, open circles and dashed line. (E) Percentage RNA species with GMP misincorporations at the experiment's end. Black bars denote raw data with gray bars being corrected for elongation complex lifetime. (F) Maximal incorporation rate determined by GTP concentration dependence. (G) Catalytic efficiency for GMP incorporation relative to polymerase mutations. (H) K_M for GMP incorporation calculated from maximal rate and catalytic efficiency for each polymerase.

GMP misincorporation when compared to WT with and without the elongation competency correction. These findings are in direct contrast to the 2′-dNTP incorporation results, pointing to the idea that residues L420 and G64 affect nucleotide fidelity differently by altering different steps within the catalytic cycle.

To determine where this unexpected result is derived from, we ran the misincorporation assay at multiple GTP concentrations to determine the concentration dependence kinetic parameters. As was expected based on the length of time needed for a GMP misincorporation to occur, the rates for all polymerases were slow, Figure 4.4F. WT had the fastest rate, followed by L420A and G64S L420A, with G64S L420A slightly slower. However, none of these rates varied much relative to each other. G64S only had the slowest rate, and all of generally followed the pattern seen for a correct incorporation event.

The catalytic efficiencies showed a different pattern than was previously seen for a native incorporation event, Figure 4.4G. Where G64S L420A had previously giving a catalytic efficiency comparable to WT, for a misincorporation the catalytic efficiency was now significantly higher than WT by almost 2.5-fold. This was also the case for the L420A mutant alone, which previously had only showed a marginally higher catalytic efficiency. The catalytic efficiency for the G64S only polymerase was lower than wildtype, but this followed the observed pattern for a correct NTP incorporation.

Also, of interest was the K_M values determined for each polymerase. In the case of GMP misincorporation, K_M values differed dramatically between polymerases, where it had remained relatively constant for a correct incorporation event, Figure 4.4H. Under these conditions, the L420A containing polymerases had a drop in K_M between 3- and 4-fold relative to WT, resulting in K_M values closer to 400 uM than the approximately 1100 uM seen for WT and 1200 uM

determined for G64S only. This K_M drop resulted in the catalytic efficiency increase seen for the L420A containing polymerases, making NTP affinity the driving factor in the reaction.

4.4 Discussion

4.4.1 Nucleotide discrimination is derived from incorporation rate, not nucleotide affinity

Previous findings from ours and other labs have suggested that active site closure is the rate limiting step in the $3D^{pol}$'s nucleic acid addition cycle. We have shown here that is the efficiency of a nucleic acid addition, and in turn nucleotide discrimination factor, is determined by the reaction rate and not nucleotide binding affinity, for both native NTP and non-native 2′-dNTP incorporations. As the nucleotide discrimination assay is an end-point readout requiring catalytic cycle completion for a detectable fluorescence signal change, it cannot be assumed that the rates for individual steps within the catalytic cycle are indeed the same for each polymerase. Because the reaction rates for acid quench, which looks at active site closure and chemistry, are comparable to the rates determined for catalytic cycle completion and are lower than those for EDTA quench looking at nucleotide capture, it can therefore be assumed that the dominant rate detected in this assay is the rate limiting step of active site closure. Thus, the rate for active site closure in all of our nucleotide discrimination experiments is the determining the preference for cognate versus noncognate nucleotide. It is also clear from these assays that even in the presence of an L420A mutation, the rate decrease from the G64S mutation is dominant, suggesting that G64S and L420A have different effects on active site closure. Differences in purine versus pyrimidine incorporation K_Ms likely come from energetic costs associated with the binding of a larger nucleotide by the polymerase.

4.4.2 G64S and L420A affect different steps in the catalytic cycle and are independent from each other.

The differences in rates observed for G64S L420A when the reaction was quenched with EDTA versus acid gives insight into the steps in catalysis that are affected by each mutation independently. EDTA quench interrogates the nucleotide capture step in the catalytic cycle. The relatively close reaction rates for both L420A and G64S L420A suggest that in this context L420A is the dominant mutation over G64S. With both of these mutant polymerases having slower nucleotide capture rates than WT, it also suggests that multiple mechanisms have an effect on nucleotide capture. Because G64S alone has an even slower rate for nucleotide capture than the G64S L420A polymerase it is likely that L420A is in some way compensating for the G64S mutation. It has previously been shown that the G64S mutation alters the active site geometry, which in turn slows the rate of active site closure to occur after nucleotide capture (Marcotte et al., 2007). These results suggest that not only is active site closure affected by G64S, but the initial nucleotide capture as well. The L420A mutation may alter the product strand 3′-end positioning, and in turn add back the needed flexibility within the active site. However, with either or both mutations present, the active site has been altered away from its optimal geometry.

A recent paper by Boehr and colleagues showed that 2′-OH altered ribonucleotides hinder genome replication by $3D^{pol}$ through preventing active site closure for the incoming nucleotide after the modified nucleotide has been incorporated, but not incoming nucleotide binding (Boehr et al., 2019). Perhaps this observation works in reverse as well, in that removing steric constraints on the product-RNA strand would result in faster active site closure without much impact on incoming nucleotide binding. Mutation of residue 420 from a Leu to Ala reduces the

amino acid side chain size significantly and likely adds flexibility to the interaction between the polymerase and product strand RNA, which could allow for faster active site closure. This possibility is supported by the increase in active site closure and catalysis rate we observed during the acid quench experiments in which L420A had a faster rate than WT. It should also be noted that during these reactions G64S and G64S L420A had very similar reaction rates, pointing to G64S being the dominant mutation over L420A during active site closure. Despite the increased active site closure rate observed for L420A alone, in the context of G64S the possible added active site flexibility seems to be rendered moot.

4.4.3 Nucleobase and nucleosugar fidelity are modulated differently by G64S and L420A mutations.

Another previous paper from the Boehr and Cameron groups suggested that nucleobase and nucleosugar fidelity while comparable may not be identical for a given polymerase (Liu et al., 2015). The findings from our GMP misincorporation study do indeed support this theory. In the context of our discrimination assay G64S L420A produced a high discrimination factor similar to G64S only, while L420A only had a lower than WT discrimination factor. For the misincorporation assay, when elongation complex lifetime was accounted for, G64S L420A had a higher percentage of GMP misincorporation similar to that of L420A only, with G64S having a low amount of GMP misincorporated. While these results may seem contradictory, they may in fact be due to the disparity between nucleobase and nucleosugar fidelity. In the nucleotide discrimination assay, base pairing geometry is maintained while the sugar geometry has been perturbed by the 2′-OH loss on the DNA base. In the case of this assay the G64S mutation is dominant and controls active site closure through a hydrogen bond interaction network with the 2′-OH. Thus, this assay is ultimately looking at nucleosugar fidelity, which is increased by the

G64S mutation with or without the L420A mutation. In contrast, the GMP misincorporation assay interrogates the context of incorrect base pair geometry and maintains the correct nucleosugar. Under these conditions the added flexibility in the product strand's 3′-end positioning is able to compensate for the misaligned base pairs allowing for a faster GMP incorporation with or without the G64S mutation because the needed 2′-OH hydrogen bond network necessary for active site closure has been maintained. When taken together these results and the differences in rate observed for each step in the catalytic cycle point to the idea that L420A and G64S affect different steps in the catalytic cycle and act orthogonally to each other.

4.4.4 Product strand positioning may play a role in RNA translocation upon completion of catalysis in addition to active site closure rate.

Previous work from our laboratories has shown that not only does the L420A mutation reduce viral genome recombination frequency, it also increases the poly-A tail length at the end of the viral genome. One possibility to explain this finding is that the change in product strand 3′-end positioning implied by the mismatch assay has also resulted in a weakened grip on the product strand by the polymerase, causing the polymerase to slip and increase poly-A tail length. This could likely also lead to an overall faster translocation rate, explaining the rate decrease between the active site closure and chemistry steps and translocation being observed for WT and G64S only that is not observed in the L420A containing polymerases. In addition, L420 is located within the polymerase thumb domain, which we have established plays a role in translocation rate. It is highly likely that this residue not only plays an established role in recombination, but due to its RNA contacts also is involved in product strand translocation.

4.4.5 Elongation complex formation could play a role in recombination efficiency.

When the kinetic findings for the L420A mutation are analyzed in the context of an extended primer grip region it is clear these residues play a clear role in virus genome recombination through positioning of the product strand 3′-end. Genome recombination occurs at a relatively low frequency and requires both a polymerase dissociation and re-association event with the RNA genome. Mutations to the residues in this extended primer grip region have not only translocation effects, but also result in changes to both elongation complex formation rates and stability. With L420A in particular, an increase in elongation complex formation time could reduce recombination efficiency by hindering polymerase rebinding or RNA strand switching during recombination, thus extending the time this event takes to complete and possibly even causing recombination to become so slow it almost looks infrequent when compared to the rate of genome replication. It should also be noted that the extended primer grip region interacts with the product strand RNA and not the template strand. In the copy-choice mechanism that is the current prevailing theory as to how recombination occurs, it is thought the product strand remains bound to the polymerase and the template strand is exchanged. Because the mutations that effect recombination and elongation complex stability are interacting with the product strand it may warrant further investigation as to if it is actually the product strand that is dissociated from the polymerase during recombination and not the template strand.

4.5 Conclusions

The findings in this work show that modifying the polymerase's interaction with the product strand RNA changes the polymerase function beyond that of only viral genome recombination by altering multiple steps in the catalytic cycle. It is also conceivable that additional mutations altering the interactions between the polymerase and RNA in this region

closer to the -1 base position would also have effects on the polymerase's catalytic cycle. In particular, mutations in this region may not be accurately represented by a nucleosugar discrimination assay due to accommodation that may be made to either incoming nucleotide binding or 3′-OH positioning. This extended primer grip region may not only serve in primer positioning and genome recombination, but also as a checkpoint for correct nucleosugar incorporation during genome replication.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Introduction

The picornavirus family of viruses includes poliovirus, the causative agent of poliomyelitis, the endocarditis from coxsackievirus, and rhinovirus, which is responsible for the common cold. There has been a large push for an eradication effort of poliovirus starting in 1988 and continuing today with the use of both an inactivated poliovirus vaccine (IPV), and a live-attenuated poliovirus vaccine (OPV) (De Jesus, 2007). While this eradication has largely been successful using predominantly the OPV, there still remain countries in which poliovirus is endemic (World Health Assembly, 2019). Compounded with that issue is the emergence of vaccine derived poliovirus cases, caused by either reversion of attenuating mutations or recombination of circulating enteroviruses C family members with OPV (Burns et al., 2013; Combelas et al., 2011; Famulare et al., 2016; Jegouic et al., 2009; Korotkova et al., 2017). Whichever is the case, a mechanistic understanding is needed to prevent these events, and possibly continue the eradication effort to completion. The goal of this work has been to provide insight into the mechanism of action by the poliovirus polymerase, $3D^{pol}$, how these findings affect these two scenarios, and how $3D^{pol}$ can be altered as a form of mutation prevention by regulating replication speed independently of polymerase fidelity, and recombination disabling to further increase vaccine efficacy.

A trait common among viral RdRPs is a high mutation rate, which is essential for virus survival (Graci and Cameron, 2002; Graci et al., 2012; Vignuzzi et al., 2006). High mutation frequency allows the virus to adapt and replicate in a variety of host cells due to the diverse

genetic pool produced (Korboukh et al., 2014; Xiao et al., 2017). This mutation frequency is correlated to polymerase replication rate, with slower replicating RdRPs having a higher fidelity and lower fidelity RdRPs replicating much faster. A previous study in coxsackievirus showed mutations within the RdRP palm domain resulted in changes to both the fidelity and the nucleotide incorporation rate, while mutations in the fingers domain primarily change the nucleotide incorporation rate with less of an effect on fidelity (Campagnola et al., 2015). Even though the mutations in the palm domain of coxsackievirus $3D^{pol}$ resulted in faster active site closure, the elongation rate remained slower than that of poliovirus $3D^{pol}$. This finding led to the possibility that regulation of speed and fidelity are controlled by different regions within the RdRP, with the rate of translocation being controlled by the fingers domain. The first section of this work is aimed at addressing if distinct regions within the fingers domain are involved in translocation. The three poliovirus $3D^{pol}$ regions in the fingers domain were targeted for single residue mutations. The first region, the index finger "kink" is a stabilizing structure mediated by a large hydrogen bonding and electrostatic network. The second region targeted, found in the pinky finger, was the "gateway", a possible barrier to templating RNA translocation. The final region targeted was the "sensor", located in the ring finger, which may function as a sensor for catalysis, and ultimately translocation. The second section of this work investigates if there is a possibility that individual residues are not responsible for translocation, but instead the fingers domain functions as an entire unit. To address this, a set of chimeric polymerases based on poliovirus and coxsackievirus $3D^{pol}$s will be constructed by exchanging the fingers domains between the two viral polymerases. Finally, the third section of this work looks at how mutations to the polymerase that affect genome recombination also affect replication kinetics.

5.2 Fingers Domain Mutations: Summary and Future Directions

In chapter 2, I originally set out to determine if three distinct structures in the fingers domain were involved in the final step in the catalytic cycle, translocation. While we found that none of these regions directly affected translocation rates, we did determine the individual roles each play in the complete catalytic cycle. The kink region plays a role in RNA binding and likely positioning, the gateway residues act as a catch for the template strand RNA after translocation, and the sensor residues assist in verification that the correct nucleotide has bound prior to chemistry as well as potentially serving as the proton donor in the chemistry reaction. While much of the work in chapter 2 has clear experimental conclusions, the largest hole I see involves the kink residues. It is clear from the data that these residues affect both elongation complex formation and stability, but how this is achieved remains unknown. Work from the Cameron group has suggested a slow step, on the order of 1 sec^{-1}, after RNA binding by the polymerase during which time the complex repositions into its catalytically active form (Arnold and Cameron, 2000). If the EC is not actively elongating then it is possible the polymerase could revert back to the inactive form, which our rapid chemical quench data and data from a previous study in the laboratory by Karr and Peersen have also supported (Karr and Peersen, 2016).

To address these questions, an experiment needs to be conducted looking at RNA binding for these kink mutants. However, the readout would need to look at both local and global RNA binding. To accomplish this an RNA construct could be designed with two fluorophores, one on the 5′-end to look at global RNA binding via fluorescence polarization, and a second internal fluorophore to dissect the local RNA binding within the polymerase. The goal with using two fluorophores would be to discover a delay in binding rate between the global and local events, and therefore be able to determine if the kink residues have any involvement in RNA positioning

post initial binding. The issues I see with an experiment like this is the fluorescence change associated with local binding could be small and difficult to determine a rate from especially if there is very little local RNA re-positioning after global binding. Hopefully the concern is moot, and the local repositioning is the dominant fluorescence change observed, or the local binding fluorescence change is biphasic to represent both RNA binding aspects and could simply then be fit by a double exponential equation, with one rate being similar to the global binding rate. One possible work around if the local binding signal does get lost in the global binding signal would be to use two adjacent 2-aminopurine molecules as the internal probe and monitor the binding using Circular Dichroism (CD). When in solution the 2AP molecules would be stacked against each other and the confirmation can be detected around a wavelength of 320 nm. When the RNA become bound by the polymerase the stacking interaction would be lost only when one 2AP is bound into the +2 pocket, so only when the RNA is repositioned would the signal change be detected, circumventing the fluorescence signal issue above, much like the work done by the von Hippel group with the T7 RNA polymerase (Datta et al., 2006).

5.3 Chimeric Polymerase Domain Swaps Conclusions and Future Directions

Chapter 3 in this document continued to investigate how RdRPs translocated, but instead of taking a residue focused approach as in chapter 2, this section looked at whole domains as interaction networks controlling translocation (Yang et al., 2010). While conducting this research the domain exchanges uncovered multiple interaction networks that control different polymerase functions. These findings suggest that the pinky finger predominately acts as its own intra-domain network and controls RNA binding, but there is some involvement by the entire fingers domain. The palm domain controls polymerase fidelity and influences the elongation rate through nucleotide capture, active site closure, and chemistry. The elongation rates are also

affected by the thumb domain that controls translocation rate, unique to RdRPs when compared to other single subunit polymerases. It was also determined throughout this study that an interaction between the palm and thumb domains was required for successful virus growth.

As with the work in chapter 2, the RNA binding kinetics were not experimentally addressed. Likely the experiment laid out in the previous section will also shed light on the outstanding questions surrounding the pinky finger and fingers domain. The work in this section was able to identify the thumb domain as the controller of translocation, however no individual residues or smaller interaction networks were identified. I did investigate if a change in a single residue, PV R456/ CV K457, from arginine to lysine (PV to CV) within the thumb domain that contracts the template RNA strand plays a role in the observed translocation rates, but that proved to be not to be the case. I believe a structure guided approach, such as looking at the outer two helices on the thumb domain, which includes the helix that contains residue 420, as an entire unit may be more fruitful than the single residue approach. The other missing component from this work is a rate for translocation was never directly determined. I was never able to reconcile the EDTA- and acid-quench data with that acquired by stopped-flow. The two CMP incorporation assay is clearly the best for determining the rate of translocation as having multiple fluorescence changes gives more information about reaction rates, however that RNA is not quench compatible since when the elongated species are run on a gel they are not able to be resolved from each other. This will require the design of a new RNA template that does only two identical incorporations with each species clearly visible on a gel. From there an expanded kinetic model can be determined and the rate for translocation derived.

5.4 Genome Recombination Conclusions and Future Directions

Chapter 4 in this work investigates how the lysine to alanine mutation of residue 420 that reduces recombination efficiency affects the polymerase replication kinetics (Kempf et al., 2016, 2019). The major findings from this analysis were that the L420A mutation likely changes the product strand 3′-hydroxyl positioning allowing for better accommodation of a mismatched base pair yielding a slightly lower polymerase fidelity and more frequent misincorporation events. In turn the polymerase has a weaker association with the product strand, generating the faster elongation rates that were observed most likely by facilitating a faster product strand translocation. We were also able to discern in this work that the previously identified G46S mutation that results in higher polymerase fidelity (Arnold et al., 2005; Pfeiffer and Kirkegaard, 2003) likely does so by monitoring changings in nucleosugar positioning, which is altered as a consequence of a misincorporation. These two residues, while both altering polymerase fidelity, function within different long-range interaction networks and as a result alter different steps in the $3D^{pol}$ catalytic cycle. These mutations also operate orthogonally to each other, as shown by a polymerase containing both mutations having multiple catalytic steps altered derived from each residue individually.

The interesting finding from this study is a possible increase in translocation rate derived from the L420A mutation. I would therefore say it is important to further characterize this mutation using a two CMP incorporation once rapid chemical quench change be reconciled with the stopped-flow work. As is noted in appendix 2, this residue is likely part of a much larger extended primer grip that is involved in recombination. I would propose this extended primer grip to be looked at both as an entire functional unit, which could be exchanged much like the work in chapter 3, and as individual residues to determine their role in translocation. To that end,

it may be also beneficial to determine if exchanging the primer grip alone would support virus growth, which would need to be done in conjunction with the Barton laboratory. I am curious if exchanging the extended primer grip would modify the kinetics enough in a poliovirus context to produce an attenuated virus with too many mutations to revert via genome replication and also yield the virus being unable to recombine with circulating enteroviruses. Ultimately this would produce a virus that would be an ideal vaccine candidate as it circumvents the two largest obstacles in live-attenuated vaccine design.

5.5 Conclusions

The works presented in these studies all revolve around resolving an incomplete mechanistic picture for the complete catalytic cycle and recombination, both of which are carried out by the picornavirus RdRP $3D^{pol}$. The two greatest hindrances to complete poliovirus eradication have been issues surrounding viral mutation during replication and recombination between circulating enteroviruses and the live attenuated virus within the OPV. Hopefully the work presented here begins to shed light on how these processes are carried out and ways to manipulate them in future vaccines. While poliovirus is no longer a threat to the vast majority of the world, the fundamental understanding of how viral RdRPs function will prove viral in the years to come as emerging viruses disrupt our lives and develop into global pandemics. It is my sincerest hope that the findings in this work assist not only with our knowledge and prevention of current known diseases, but also provide a foundation upon which we can face head on the unknown diseases before us.

REFERENCES

Acevedo, A., Woodman, A., Arnold, J.J., Yeh, M. Te, Evans, D., Cameron, C.E., and Andino, R. (2018). Genetic recombination of poliovirus facilitates subversion of host barriers to infection. BioRxiv 273060.

Agol, V.I., and Gmyl, A.P. (2018). Emergency Services of Viral RNAs: Repair and Remodeling. Microbiol. Mol. Biol. Rev. *82*, e00067-17.

Aguado, L.C., Jordan, T.X., Hsieh, E., Blanco-Melo, D., Heard, J., Panis, M., Vignuzzi, M., and TenOever, B.R. (2018). Homologous recombination is an intrinsic defense against antiviral RNA interference. Proc. Natl. Acad. Sci. *115*, E9211–E9219.

Andino, R., and Domingo, E. (2015). Viral quasispecies. Virology *479–480*, 46–51.

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.

Arnold, J.J., and Cameron, C.E. (2000). Poliovirus RNA-dependent RNA Polymerase (3Dpol). J. Biol. Chem. *275*, 5329–5336.

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.

Arnold, J.J., Gohara, D.W., and Cameron, C.E. (2004). Poliovirus RNA-Dependent RNA Polymerase (3Dpol): Pre-Steady-State Kinetic Analysis of Ribonucleotide Incorporation in the Presence of Mn2+. Biochemistry *43*, 5138–5148.

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.

Baicus, A. (2012). History of polio vaccination. World J. Virol. *1*, 108.

Barton, N.H. (1998). Why Sex and Recombination? Science. *281*, 1986–1990.

Bell, Y.C., Semler, B.L., and Ehrenfeld, E. (1999). Requirements for RNA replication of a poliovirus replicon by coxsackievirus B3 RNA polymerase. J. Virol. *73*, 9413–9421.

Belov, G.A., and van Kuppeveld, F.J. (2012). (+)RNA viruses rewire cellular pathways to build replication organelles. Curr. Opin. Virol. *2*, 740–747.

Benschop, K.S.M., Williams, Ç.H., Wolthers, K.C., Stanway, G., and Simmonds, P. (2008). Widespread recombination within human parechoviruses: analysis of temporal dynamics and constraints. J. Gen. Virol. *89*, 1030–1035.

Bessaud, M., Joffret, M.-L., Blondel, B., and Delpeyroux, F. (2016). Exchanges of genomic domains between poliovirus and other cocirculating species C enteroviruses reveal a high degree of plasticity. Sci. Rep. *6*, 38831.

Bodian, D., and Horstmann, D.H. (1965). Polioviruses. In Viral and Rickettsial Infections of Man, (Lippincott), pp. 430–473.

Boehr, A.K., Arnold, J.J., Oh, H.S., Cameron, C.E., and Boehr, D.D. (2019). 2′-C-methylated nucleotides terminate virus RNA synthesis by preventing active site closure of the viral RNA-dependent RNA polymerase. J. Biol. Chem. *294*, 16897–16907.

Brown, B., Oberste, M.S., Maher, K., and Pallansch, M.A. (2003). Complete Genomic Sequencing Shows that Polioviruses and Members of Human Enterovirus Species C Are Closely Related in the Noncapsid Coding Region. J. Virol. *77*, 8973–8984.

Bull, J.J., Sanjuán, R., and Wilke, C.O. (2007). Theory of Lethal Mutagenesis for Viruses. J. Virol.

Burns, C.C., Shaw, J., Jorba, J., Bukbuk, D., Adu, F., Gumede, N., Pate, M.A., Abanida, E.A., Gasasira, A., Iber, J., et al. (2013). Multiple Independent Emergences of Type 2 Vaccine-Derived Polioviruses during a Large Outbreak in Northern Nigeria. J. Virol. *87*, 4907–4922.

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. *19*, 768–774.

Campagnola, G., McDonald, S., Beaucourt, S., Vignuzzi, M., and Peersen, O.B. (2015). Structure-Function Relationships Underlying the Replication Fidelity of Viral RNA-Dependent RNA Polymerases. J. Virol. *89*, 275–286.

Castro, C., Smidansky, E., Maksimchuk, K.R., Arnold, J.J., Korneeva, V.S., Gotte, M., Konigsberg, W., and Cameron, C.E. (2007). Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. Proc. Natl. Acad. Sci. *104*, 4267–4272.

Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Götte, 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.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. Nature *348*, 454–455.

Chao, L. (1997). Evolution of sex and the molecular clock in RNA viruses. Gene *205*, 301–308.

Cheetham, G.M. (1999). Structure of a Transcribing T7 RNA Polymerase Initiation Complex. Science. *286*, 2305–2309.

Chen, C., Wang, Y., Shan, C., Sun, Y., Xu, P., Zhou, H., Yang, C., Shi, P.-Y., Rao, Z., Zhang, B., et al. (2013). Crystal Structure of Enterovirus 71 RNA-Dependent RNA Polymerase Complexed with Its Protein Primer VPg: Implication for a trans Mechanism of VPg Uridylylation. J. Virol. *87*, 5755–5768.

Collis, P.S., O'Donnell, B.J., Barton, D.J., Rogers, J.A., and Flanegan, J.B. (1992). Replication of poliovirus RNA and subgenomic RNA transcripts in transfected cells. J. Virol. *66*, 6480–6488.

Combelas, N., Holmblat, B., Joffret, M.-L., Colbère-Garapin, F., and Delpeyroux, F. (2011). Recombination between Poliovirus and Coxsackie A Viruses of Species C: A Model of Viral Genetic Plasticity and Emergence. Viruses *3*, 1460–1484.

Cornell, C.T., and Semler, B.L. (2002). Subdomain Specific Functions of the RNA Polymerase Region of Poliovirus 3CD Polypeptide. Virology *298*, 200–213.

Crotty, S., Maag, D., Arnold, J.J., Zhong, W., Lau, J.Y.N., Hong, Z., Andino, R., and Cameron, C.E. (2000). The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat. Med. *6*, 1375–1379.

Crotty, S., Cameron, C.E., and Andino, R. (2001). RNA virus error catastrophe: Direct molecular test by using ribavirin. Proc. Natl. Acad. Sci. *98*, 6895–6900.

Crotty, S., Cameron, C., and Andino, R. (2002). Ribavirin's antiviral mechanism of action: lethal mutagenesis? J. Mol. Med. *80*, 86–95.

Da, L.-T., E, C., Duan, B., Zhang, C., Zhou, X., and Yu, J. (2015). A Jump-from-Cavity Pyrophosphate Ion Release Assisted by a Key Lysine Residue in T7 RNA Polymerase Transcription Elongation. PLOS Comput. Biol. *11*, e1004624.

Da, L.-T., E, C., Shuai, Y., Wu, S., Su, X.-D., and Yu, J. (2017). T7 RNA polymerase translocation is facilitated by a helix opening on the fingers domain that may also prevent backtracking. Nucleic Acids Res. *45*, 7909–7921.

Van Damme, P., De Coster, I., Bandyopadhyay, A.S., Revets, H., Withanage, K., De Smedt, P., Suykens, L., Oberste, M.S., Weldon, W.C., Costa-Clemens, S.A., et al. (2019). The safety and immunogenicity of two novel live attenuated monovalent (serotype 2) oral poliovirus vaccines in healthy adults: a double-blind, single-centre phase 1 study. Lancet *394*, 148–158.

Datta, K., Johnson, N.P., and von Hippel, P.H. (2006). Mapping the Conformation of the Nucleic Acid Framework of the T7 RNA Polymerase Elongation Complex in Solution Using Low-energy CD and Fluorescence Spectroscopy. J. Mol. Biol. *360*, 800–813.

Deng, C.-L., Yeo, H., Ye, H.-Q., Liu, S.-Q., Shang, B.-D., Gong, P., Alonso, S., Shi, P.-Y., and Zhang, B. (2014). Inhibition of Enterovirus 71 by Adenosine Analog NITD008. J. Virol. *88*, 11915–11923.

Diamond, S.E., and Kirkegaard, K. (1994). Clustered charged-to-alanine mutagenesis of poliovirus RNA-dependent RNA polymerase yields multiple temperature-sensitive mutants defective in RNA synthesis. J. Virol. *68*, 863–876.

Dolan, P.T., Whitfield, Z.J., and Andino, R. (2018). Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. Annu. Rev. Virol. *5*, 69–92.

Domingo, E., and Holland, J.J. (1997). RNA Virus Mutations and Fitness for Survival. Annu. Rev. Microbiol. *51*, 151–178.

Drake, J.W., and Holland, J.J. (1999). Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. *96*, 13910–13913.

Duarte, E., Clarke, D., Moya, A., Domingo, E., and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. Proc. Natl. Acad. Sci. *89*, 6015–6019.

Dubankova, A., Horova, V., Klima, M., and Boura, E. (2019). Structures of kobuviral and siciniviral polymerases reveal conserved mechanism of picornaviral polymerase activation. J. Struct. Biol. *208*, 92–98.

Van Dung, N., Anh, P.H., Van Cuong, N., Hoa, N.T., Carrique-Mas, J., Hien, V.B., Campbell, J., Baker, S., Farrar, J., Woolhouse, M.E., et al. (2014). Prevalence, genetic diversity and recombination of species G enteroviruses infecting pigs in Vietnam. J. Gen. Virol. *95*, 549–556.

Van Dyke, T.A., and Flanegan, J.B. (1980). Identification of poliovirus polypeptide P63 as a soluble RNA-dependent RNA polymerase. J. Virol. *35*, 732–740.

Van Dyke, T.A., Rickles, R.J., and Flanegan, J.B. (1982). Genome-length copies of poliovirion RNA are synthesized in vitro by the poliovirus RNA-dependent RNA polymerase. J. Biol. Chem. *257*, 4610–4617.

Egger, D., and Bienz, K. (2002). Recombination of Poliovirus RNA Proceeds in Mixed Replication Complexes Originating from Distinct Replication Start Sites. J. Virol. *76*, 10960–10971.

Eigen, M. (2002). Error catastrophe and antiviral strategy. Proc. Natl. Acad. Sci. *99*, 13374–13376.

Erickson, A.K., Jesudhasan, P.R., Mayer, M.J., Narbad, A., Winter, S.E., and Pfeiffer, J.K. (2018). Bacteria Facilitate Enteric Virus Co-infection of Mammalian Cells and Promote Genetic Recombination. Cell Host Microbe *23*, 77-88.e5.

Escarmís, C., Lázaro, E., Arias, A., and Domingo, E. (2008). Repeated Bottleneck Transfers Can Lead to Non-cytocidal Forms of a Cytopathic Virus: Implications for Viral Extinction. J. Mol. Biol. *376*, 367–379.

Escarmís, C., Gómez-Mariano, G., Dávila, M., Lázaro, E., and Domingo, E. (2002). Resistance to extinction of low fitness virus subjected to plaque-to-plaque transfers: diversification by mutation clustering 1 1Edited by J. Karn. J. Mol. Biol. *315*, 647–661.

Famulare, M., Chang, S., Iber, J., Zhao, K., Adeniji, J.A., Bukbuk, D., Baba, M., Behrend, M., Burns, C.C., and Oberste, M.S. (2016). Sabin Vaccine Reversion in the Field: a Comprehensive Analysis of Sabin-Like Poliovirus Isolates in Nigeria. J. Virol. *90*, 317–331.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmís, 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.

Ferrer-Orta, C., Arias, A., Escarmis, C., and Verdaguer, N. (2006). A comparison of viral RNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. *16*, 27–34.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E., and Verdaguer, N. (2007). Sequential structures provide insights into the fidelity of RNA replication. Proc. Natl. Acad. Sci. *104*, 9463–9468.

Ferrero, D.S., Ferrer-Orta, C., and Verdaguer, N. (2018). Virus Protein and Nucleoprotein Complexes (Singapore: Springer Singapore).

Fitzsimmons, W.J., Woods, R.J., McCrone, J.T., Woodman, A., Arnold, J.J., Yennawar, M., Evans, R., Cameron, C.E., and Lauring, A.S. (2018). A speed–fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLOS Biol. *16*, e2006459.

Flanegan, J.B., and Baltimore, D. (1977). Poliovirus-specific primer-dependent RNA polymerase able to copy poly(A). Proc. Natl. Acad. Sci. *74*, 3677–3680.

Fricks, C.E., and Hogle, J.M. (1990). Cell-induced conformational change in poliovirus: externalization of the amino terminus of VP1 is responsible for liposome binding. J. Virol. *64*, 1934–1945.

Garcia-Ruiz, H., Diaz, A., and Ahlquist, P. (2018). Intermolecular RNA Recombination Occurs at Different Frequencies in Alternate Forms of Brome Mosaic Virus RNA Replication Compartments. Viruses *10*, 131.

Gnadig, N.F., Beaucourt, S., Campagnola, G., Borderia, A. V., Sanz-Ramos, M., Gong, P., Blanc, H., Peersen, O.B., and Vignuzzi, M. (2012). Coxsackievirus B3 mutator strains are attenuated in vivo. Proc. Natl. Acad. Sci. *109*, E2294–E2303.

Gohara, D.W., Arnold, J.J., and Cameron, C.E. (2004). Poliovirus RNA-Dependent RNA Polymerase (3Dpol): Kinetic, Thermodynamic, and Structural Analysis of Ribonucleotide Selection. Biochemistry *43*, 5149–5158.

Gong, P., and Peersen, O.B. (2010). Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. *107*, 22505–22510.

Gong, P., Campagnola, G., and Peersen, O.B. (2009). A quantitative stopped-flow fluorescence assay for measuring polymerase elongation rates. Anal. Biochem. *391*, 45–55.

Gong, P., Kortus, M.G., Nix, J.C., Davis, R.E., and Peersen, O.B. (2013). Structures of Coxsackievirus, Rhinovirus, and Poliovirus Polymerase Elongation Complexes Solved by Engineering RNA Mediated Crystal Contacts. PLoS One *8*, e60272.

Graci, J.D., and Cameron, C.E. (2002). Quasispecies, Error Catastrophe, and the Antiviral Activity of Ribavirin. Virology *298*, 175–180.

Graci, J.D., and Cameron, C.E. (2006). Mechanisms of action of ribavirin against distinct viruses. Rev. Med. Virol. *16*, 37–48.

Graci, J.D., Gnadig, N.F., Galarraga, J.E., Castro, C., Vignuzzi, M., and Cameron, C.E. (2012). Mutational Robustness of an RNA Virus Influences Sensitivity to Lethal Mutagenesis. J. Virol. *86*, 2869–2873.

Guo, Q., and Sousa, R. (2006). Translocation by T7 RNA Polymerase: A Sensitively Poised Brownian Ratchet. J. Mol. Biol. *358*, 241–254.

Hansen, J.L., Long, A.M., and Schultz, S.C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. Structure *5*, 1109–1122.

Hobdey, S.E., Kempf, B.J., Steil, B.P., Barton, D.J., and Peersen, O.B. (2010). Poliovirus Polymerase Residue 5 Plays a Critical Role in Elongation Complex Stability. J. Virol. *84*, 8072–8084.

Hobson, S.D., Rosenblum, E.S., Richards, O.C., Richmond, K., Kirkegaard, K., and Schultz, S.C. (2001). Oligomeric structures of poliovirus polymerase are important for function. EMBO J. *20*, 1153–1163.

Huang, H. (1998). Structure of a Covalently Trapped Catalytic Complex of HIV-1 Reverse Transcriptase: Implications for Drug Resistance. Science. *282*, 1669–1675.

Jarvis, T.C., and Kirkegaard, K. (1992). Poliovirus RNA recombination: mechanistic studies in the absence of selection. EMBO J. *11*, 3135–3145.

Jegouic, S., Joffret, M.-L., Blanchard, C., Riquet, F.B., Perret, C., Pelletier, I., Colbere-Garapin, F., Rakoto-Andrianarivelo, M., and Delpeyroux, F. (2009). Recombination between Polioviruses and Co-Circulating Coxsackie A Viruses: Role in the Emergence of Pathogenic Vaccine-Derived Polioviruses. PLoS Pathog. *5*, e1000412.

Jenkins, G.M., Rambaut, A., Pybus, O.G., and Holmes, E.C. (2002). Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. J. Mol. Evol. *54*, 156–165.

De Jesus, N.H. (2007). Epidemics to eradication: the modern history of poliomyelitis. Virol. J. *4*, 70.

Johnson, K.A. (2010). The kinetic and chemical mechanism of high-fidelity DNA polymerases. Biochim. Biophys. Acta - Proteins Proteomics *1804*, 1041–1048.

Johnson, K.A., Simpson, Z.B., and Blom, T. (2009a). Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. Anal. Biochem. *387*, 20–29.

Johnson, K.A., Simpson, Z.B., and Blom, T. (2009b). FitSpace Explorer: An algorithm to evaluate multidimensional parameter space in fitting kinetic data. Anal. Biochem. *387*, 30–41.

Kamer, G., and Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucleic Acids Res. *12*, 7269–7282.

Karr, J.P., and Peersen, O.B. (2016). ATP Is an Allosteric Inhibitor of Coxsackievirus B3 Polymerase. Biochemistry *55*, 3995–4002.

Kati, W.M., Johnson, K.A., Jerva, L.F., and Anderson, K.S. (1992). Mechanism and fidelity of HIV reverse transcriptase. J. Biol. Chem. *267*, 25988–25997.

Kautz, T., and Forrester, N. (2018). RNA Virus Fidelity Mutants: A Useful Tool for Evolutionary Biology or a Complex Challenge? Viruses *10*, 600.

Kempf, B.J., and Barton, D.J. (2015). Picornavirus RNA polyadenylation by 3Dpol, the viral RNA-dependent RNA polymerase. Virus Res. *206*, 3–11.

Kempf, B.J., Kelly, M.M., Springer, C.L., Peersen, O.B., and Barton, D.J. (2013). Structural Features of a Picornavirus Polymerase Involved in the Polyadenylation of Viral RNA. J. Virol. *87*, 5629–5644.

Kempf, B.J., Peersen, O.B., and Barton, D.J. (2016). Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. J. Virol. *90*, 8410–8421.

Kempf, B.J., Watkins, C.L., Peersen, O.B., and Barton, D.J. (2019). Picornavirus RNA Recombination Counteracts Error Catastrophe. J. Virol. *93*, e00652-19.

Kim, H., Ellis, V.D., Woodman, A., Zhao, Y., Arnold, J.J., and Cameron, C.E. (2019). RNA-Dependent RNA Polymerase Speed and Fidelity are not the Only Determinants of the Mechanism or Efficiency of Recombination. Genes (Basel). *10*, 968.

Kirkegaard, K., and Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. Cell *47*, 433–443.

Kohlstaedt, L., Wang, J., Friedman, J., Rice, P., and Steitz, T. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science. *256*, 1783–1790.

Koike, S., and Nomoto, A. (2010). Poliomyelitis. In The Picornaviruses, (asm Pub2Web), pp. 339–351.

Koonin, E. V., Wolf, Y.I., Nagasaki, K., and Dolja, V. V. (2008). The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. Nat. Rev. Microbiol. *6*, 925–939.

Korboukh, V.K., Lee, C.A., Acevedo, A., Vignuzzi, M., Xiao, Y., Arnold, J.J., Hemperly, S., Graci, J.D., August, A., Andino, R., et al. (2014). RNA Virus Population Diversity, an Optimum for Maximal Fitness and Virulence. J. Biol. Chem. *289*, 29531–29544.

Korotkova, E., Laassri, M., Zagorodnyaya, T., Petrovskaya, S., Rodionova, E., Cherkasova, E., Gmyl, A., Ivanova, O., Eremeeva, T., Lipskaya, G., et al. (2017). Pressure for Pattern-Specific Intertypic Recombination between Sabin Polioviruses: Evolutionary Implications. Viruses *9*, 353.

Kortus, M.G., Kempf, B.J., Haworth, K.G., Barton, D.J., and Peersen, O.B. (2012). A Template RNA Entry Channel in the Fingers Domain of the Poliovirus Polymerase. J. Mol. Biol. *417*, 263–278.

Krupovic, M., Dolja, V. V., and Koonin, E. V. (2019). Origin of viruses: primordial replicators recruiting capsids from hosts. Nat. Rev. Microbiol. *17*, 449–458.

Landsteiner, K., and Popper, E. (1908). Mikroscopische präparate von einem menschlichen und zwei affentückermarker. Wein Klin Wschr. *21*, 1930.

Lauring, A.S., and Andino, R. (2010). Quasispecies Theory and the Behavior of RNA Viruses. PLoS Pathog. *6*, e1001005.

Lauring, A.S., Acevedo, A., Cooper, S.B., and Andino, R. (2012). Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus. Cell Host Microbe *12*, 623–632.

Lázaro, E., Escarmís, C., Domingo, E., and Manrubia, S.C. (2002). Modeling Viral Genome Fitness Evolution Associated with Serial Bottleneck Events: Evidence of Stationary States of Fitness. J. Virol. *76*, 8675–8681.

Li, C., Wang, H., Shi, J., Yang, D., Zhou, G., Chang, J., Cameron, C.E., Woodman, A., and Yu, L. (2019). Senecavirus-Specific Recombination Assays Reveal the Intimate Link between Polymerase Fidelity and RNA Recombination. J. Virol. *93*, e00576-19.

Liu, X., Musser, D., Lee, C., Yang, X., Arnold, J., Cameron, C., and Boehr, D. (2015). Nucleobase but not Sugar Fidelity is Maintained in the Sabin I RNA-Dependent RNA Polymerase. Viruses *7*, 5571–5586.

Love, R.A., Maegley, K.A., Yu, X., Ferre, R.A., Lingardo, L.K., Diehl, W., Parge, H.E., Dragovich, P.S., and Fuhrman, S.A. (2004). The Crystal Structure of the RNA-Dependent RNA Polymerase from Human Rhinovirus. Structure *12*, 1533–1544.

Lowry, K., Woodman, A., Cook, J., and Evans, D.J. (2014). Recombination in Enteroviruses Is a Biphasic Replicative Process Involving the Generation of Greater-than Genome Length 'Imprecise' Intermediates. PLoS Pathog. *10*, e1004191.

Lyle, J.M., Bullitt, E., Bienz, K., and Kirkegaard, K. (2002). Visualization and Functional Analysis of RNA-Dependent RNA Polymerase Lattices. Science. *296*, 2218–2222.

Marcotte, L.L., Wass, A.B., Gohara, D.W., Pathak, H.B., Arnold, J.J., Filman, D.J., Cameron, C.E., and Hogle, J.M. (2007). Crystal Structure of Poliovirus 3CD Protein: Virally Encoded Protease and Precursor to the RNA-Dependent RNA Polymerase. J. Virol. *81*, 3583–3596.

Martin, E.T., Fairchok, M.P., Stednick, Z.J., Kuypers, J., and Englund, J.A. (2013). Epidemiology of Multiple Respiratory Viruses in Childcare Attendees. J. Infect. Dis. *207*, 982–989.

McCrone, J.T., and Lauring, A.S. (2018). Genetic bottlenecks in intraspecies virus transmission. Curr. Opin. Virol. *28*, 20–25.

McDonald, S., Block, A., Beaucourt, S., Moratorio, G., Vignuzzi, M., and Peersen, O.B. (2016). Design of a Genetically Stable High Fidelity Coxsackievirus B3 Polymerase That Attenuates Virus Growth in Vivo. J. Biol. Chem. *291*, 13999–14011.

McIntyre, C.L., McWilliam Leitch, E.C., Savolainen-Kopra, C., Hovi, T., and Simmonds, P. (2010). Analysis of Genetic Diversity and Sites of Recombination in Human Rhinovirus Species C. J. Virol. *84*, 10297–10310.

McIntyre, C.L., Savolainen-Kopra, C., Hovi, T., and Simmonds, P. (2013). Recombination in the evolution of human rhinovirus genomes. Arch. Virol. *158*, 1497–1515.

McWilliam Leitch, E.C., Cabrerizo, M., Cardosa, J., Harvala, H., Ivanova, O.E., Kroes, A.C.M., Lukashev, A., Muir, P., Odoom, J., Roivainen, M., et al. (2010). Evolutionary Dynamics and Temporal/Geographical Correlates of Recombination in the Human Enterovirus Echovirus Types 9, 11, and 30. J. Virol. *84*, 9292–9300.

McWilliam Leitch, E.C., Cabrerizo, M., Cardosa, J., Harvala, H., Ivanova, O.E., Koike, S., Kroes, A.C.M., Lukashev, A., Perera, D., Roivainen, M., et al. (2012). The Association of Recombination Events in the Founding and Emergence of Subgenogroup Evolutionary Lineages of Human Enterovirus 71. J. Virol. *86*, 2676–2685.

Mehndiratta, M.M., Mehndiratta, P., and Pande, R. (2014). Poliomyelitis. The Neurohospitalist *4*, 223–229.

Minor, P. (2014). The polio endgame. Hum. Vaccin. Immunother. *10*, i–iii.

Monto, A.S. (1999). Francis Field Trial of Inactivated Poliomyelitis Vaccine: Background and Lessons for Today. Epidemiol. Rev. *21*, 7–23.

Morasco, B.J., Sharma, N., Parilla, J., and Flanegan, J.B. (2003). Poliovirus cre(2C)-Dependent Synthesis of VPgpUpU Is Required for Positive- but Not Negative-Strand RNA Synthesis. J. Virol. *77*, 5136–5144.

Moratorio, G., Henningsson, R., Barbezange, C., Carrau, L., Bordería, A. V., Blanc, H., Beaucourt, S., Poirier, E.Z., Vallet, T., Boussier, J., et al. (2017). Attenuation of RNA viruses by redirecting their evolution in sequence space. Nat. Microbiol. *2*, 17088.

Moscato, B., Swain, M., and Loria, J.P. (2016). Induced Fit in the Selection of Correct versus Incorrect Nucleotides by DNA Polymerase β. Biochemistry *55*, 382–395.

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.

Moustafa, I.M., Korboukh, V.K., Arnold, J.J., Smidansky, E.D., Marcotte, L.L., Gohara, D.W., Yang, X., Sánchez-Farrán, M.A., Filman, D., Maranas, J.K., et al. (2014). Structural Dynamics as a Contributor to Error-prone Replication by an RNA-dependent RNA Polymerase. J. Biol. Chem. *289*, 36229–36248.

Murray, K.E., and Barton, D.J. (2003). Poliovirus CRE-dependent VPg uridylylation is required for positive-strand RNA synthesis but not for negative-strand RNA synthesis. J. Virol. *77*, 4739–4750.

Muslin, Kain, Bessaud, Blondel, and Delpeyroux (2019). Recombination in Enteroviruses, a Multi-Step Modular Evolutionary Process. Viruses *11*, 859.

Nagy, P.D., and Simon, A.E. (1997). New Insights into the Mechanisms of RNA Recombination. Virology *235*, 1–9.

Nathanson, N., and Langmuir, A.D. (1963). The Cutter Incident Poliomyelitis Following Formaldehyde-Inactivated Poliovirus Vaccination in the United States during the Spring of 1955. Am. J. Epidemiol. *78*, 61–81.

Nathanson, N., and Martin, J.R. (1979). The Epidemiology of Poliomyelitis: Enigmas Surrounding its Appearance, Epidemicity, and Disappearance. Am. J. Epidemiol. *110*, 672–692.

Oberste, M.S., Maher, K., and Pallansch, M.A. (2004a). Evidence for Frequent Recombination within Species Human Enterovirus B Based on Complete Genomic Sequences of All Thirty-Seven Serotypes. J. Virol. *78*, 855–867.

Oberste, M.S., Peñaranda, S., Maher, K., and Pallansch, M.A. (2004b). Complete genome sequences of all members of the species Human enterovirus A. J. Gen. Virol. *85*, 1597–1607.

Palmenberg, A., Neubauer, D., and Skern, T. (2010). Genome Organization and Encoded Proteins. In The Picornaviruses, (asm Pub2Web), pp. 3–17.

Paul, A. V., van Boom, J.H., Filippov, D., and Wimmer, E. (1998). Protein-primed RNA synthesis by purified poliovirus RNA polymerase. Nature *393*, 280–284.

Paul, A. V., Peters, J., Mugavero, J., Yin, J., van Boom, J.H., and Wimmer, E. (2003). Biochemical and Genetic Studies of the VPg Uridylylation Reaction Catalyzed by the RNA Polymerase of Poliovirus. J. Virol. *77*, 891–904.

Peck, K.M., and Lauring, A.S. (2018). Complexities of Viral Mutation Rates. J. Virol. *92*.

Peersen, O. (2019a). A Comprehensive Superposition of Viral Polymerase Structures. Viruses *11*, 745.

Peersen, O. (2019b). Polymerase Structure Alignments - v2a Update (October 2019).

Peersen, O.B. (2017). Picornaviral polymerase structure, function, and fidelity modulation. Virus Res. *234*, 4–20.

Pfeiffer, J.K., and Kirkegaard, K. (2003). A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. Proc. Natl. Acad. Sci. *100*, 7289–7294.

Racaniello, V.R. (2006). One hundred years of poliovirus pathogenesis. Virology *344*, 9–16.

Ren, R., and Racaniello, V.R. (1992). Human poliovirus receptor gene expression and poliovirus tissue tropism in transgenic mice. J. Virol. *66*, 296–304.

Runckel, C., Westesson, O., Andino, R., and DeRisi, J.L. (2013). Identification and Manipulation of the Molecular Determinants Influencing Poliovirus Recombination. PLoS Pathog. *9*, e1003164.

Sabin, A.B. (1960). Oral, Live Poliovirus Vaccine for Elimination of Poliomyelitis. Arch. Intern. Med. *106*, 5.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral Mutation Rates. J. Virol. *84*, 9733–9748.

Shen, H., Sun, H., and Li, G. (2012a). What Is the Role of Motif D in the Nucleotide Incorporation Catalyzed by the RNA-dependent RNA Polymerase from Poliovirus? PLoS Comput. Biol. *8*, e1002851.

Shen, H., Moustafa, I.M., Cameron, C.E., and Colina, C.M. (2012b). Exploring the Dynamics of Four RNA-Dependent RNA Polymerases by a Coarse-Grained Model. J. Phys. Chem. B *116*, 14515–14524.

Shen, M., Reitman, Z.J., Zhao, Y., Moustafa, I., Wang, Q., Arnold, J.J., Pathak, H.B., and Cameron, C.E. (2008). Picornavirus Genome Replication. Identification of the Surface of the Poliovirus (PV) 3C Dimer that Interacts with PV 3Dpol during VPg Uridylylation and Construction of a Structural Model for the PV 3C2-3Dpol Complex. J. Biol. Chem. *283*, 875–888.

Shi, J., Perryman, J.M., Yang, X., Liu, X., Musser, D.M., Boehr, A.K., Moustafa, I.M., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2019). Rational Control of Poliovirus RNA-Dependent RNA Polymerase Fidelity by Modulating Motif-D Loop Conformational Dynamics. Biochemistry *58*, 3735–3743.

Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen, J.-J., Liu, L., et al. (2018). The evolutionary history of vertebrate RNA viruses. Nature *556*, 197–202.

Shi, W., Ye, H.-Q., Deng, C.-L., Li, R., Zhang, B., and Gong, P. (2020). A nucleobase-binding pocket in a viral RNA-dependent RNA polymerase contributes to elongation complex stability. Nucleic Acids Res. *48*, 1392–1405.

Sholders, A.J., and Peersen, O.B. (2014). Distinct Conformations of a Putative Translocation Element in Poliovirus Polymerase. J. Mol. Biol. *426*, 1407–1419.

Shu, B., and Gong, P. (2016). Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. Proc. Natl. Acad. Sci. *113*, E4005–E4014.

Shu, B., and Gong, P. (2017). The uncoupling of catalysis and translocation in the viral RNA-dependent RNA polymerase. RNA Biol. *14*, 1314–1319.

Sierra, M., Airaksinen, A., González-López, C., Agudo, R., Arias, A., and Domingo, E. (2007). Foot-and-Mouth Disease Virus Mutant with Decreased Sensitivity to Ribavirin: Implications for Error Catastrophe. J. Virol. *81*, 2012–2024.

Simmonds, P. (2006). Recombination and Selection in the Evolution of Picornaviruses and Other Mammalian Positive-Stranded RNA Viruses. J. Virol. *80*, 11124–11140.

Simmonds, P., and Welch, J. (2006). Frequency and Dynamics of Recombination within Different Species of Human Enteroviruses. J. Virol. *80*, 483–493.

Simon-Loriere, E., and Holmes, E.C. (2011). Why do RNA viruses recombine? Nat. Rev. Microbiol. *9*, 617–626.

Smithee, S., Tracy, S., and Chapman, N.M. (2015). Mutational Disruption of cis -Acting Replication Element 2C in Coxsackievirus B3 Leads to 5′-Terminal Genomic Deletions. J. Virol. *89*, 11761–11772.

Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.-C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. Nature *364*, 593–599.

Spagnolo, J.F., Rossignol, E., Bullitt, E., and Kirkegaard, K. (2010). Enzymatic and nonenzymatic functions of viral RNA-dependent RNA polymerases within oligomeric arrays. RNA *16*, 382–393.

Steil, B.P., and Barton, D.J. (2008). Poliovirus cis-Acting Replication Element-Dependent VPg Uridylylation Lowers the Km of the Initiating Nucleoside Triphosphate for Viral RNA Replication. J. Virol. *82*, 9400–9408.

Steil, B.P., Kempf, B.J., and Barton, D.J. (2010). Poly(A) at the 3′ End of Positive-Strand RNA and VPg-Linked Poly(U) at the 5′ End of Negative-Strand RNA Are Reciprocal Templates during Replication of Poliovirus RNA. J. Virol. *84*, 2843–2858.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231–232.

Steitz, T.A. (2006). Visualizing polynucleotide polymerase machines at work. EMBO J. *25*, 3458–3468.

Sun, Y., Wang, Y., Shan, C., Chen, C., Xu, P., Song, M., Zhou, H., Yang, C., Xu, W., Shi, P.-Y., et al. (2012). Enterovirus 71 VPg Uridylation Uses a Two-Molecular Mechanism of 3D Polymerase. J. Virol. *86*, 13662–13671.

Svetlov, V., and Nudler, E. (2013). Basic mechanism of transcription by RNA polymerase II. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 20–28.

Tan, Y., Hassan, F., Schuster, J.E., Simenauer, A., Selvarangan, R., Halpin, R.A., Lin, X., Fedorova, N., Stockwell, T.B., Lam, T.T.-Y., et al. (2016). Molecular Evolution and Intraclade Recombination of Enterovirus D68 during the 2014 Outbreak in the United States. J. Virol. *90*, 1997–2007.

Tellez, A.B., Wang, J., Tanner, E.J., Spagnolo, J.F., Kirkegaard, K., and Bullitt, E. (2011). Interstitial Contacts in an RNA-Dependent RNA Polymerase Lattice. J. Mol. Biol. *412*, 737–750.

Temiakov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S., and Vassylyev, D.G. (2004). Structural Basis for Substrate Selection by T7 RNA Polymerase. Cell *116*, 381–391.

Thomen, P., Lopez, P.J., and Heslot, F. (2005). Unravelling the Mechanism of RNA-Polymerase Forward Motion by Using Mechanical Force. Phys. Rev. Lett. *94*, 128102.

Thompson, A.A. (2006). X-ray Crystal Structure and Biophysical Analysis of the Poliovirus RNA-dependent RNA Polymerase.

Thompson, A.A., and Peersen, O.B. (2004). Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. EMBO J. *23*, 3462–3471.

Thompson, A.A., Albertini, R.A., and Peersen, O.B. (2007). Stabilization of Poliovirus Polymerase by NTP Binding and Fingers–Thumb Interactions. J. Mol. Biol. *366*, 1459–1474.

Underwood, M. (1789). Debility of the lower extremities. In Treatise on the Diseases of Children, pp. 53–57.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature *439*, 344–348.

Virgen-Slane, R., Rozovics, J.M., Fitzgerald, K.D., Ngo, T., Chou, W., van der Heden van Noort, G.J., Filippov, D. V., Gershon, P.D., and Semler, B.L. (2012). An RNA virus hijacks an incognito function of a DNA repair enzyme. Proc. Natl. Acad. Sci. *109*, 14634–14639.

Vives-Adrian, L., Lujan, C., Oliva, B., van der Linden, L., Selisko, B., Coutard, B., Canard, B., van Kuppeveld, F.J.M., Ferrer-Orta, C., and Verdaguer, N. (2014). The Crystal Structure of a Cardiovirus RNA-Dependent RNA Polymerase Reveals an Unusual Conformation of the Polymerase Active Site. J. Virol. *88*, 5595–5607.

Vo, N. V., Young, K.-C., and Lai, M.M.C. (2003). Mutagenic and Inhibitory Effects of Ribavirin on Hepatitis C Virus RNA Polymerase †. Biochemistry *42*, 10462–10471.

Wang, C., Wang, C., Li, Q., Wang, Z., and Xie, W. (2017). Crystal Structure and Thermostability Characterization of Enterovirus D68 3Dpol. J. Virol. *91*.

Wang, J., Lyle, J.M., and Bullitt, E. (2013). Surface for Catalysis by Poliovirus RNA-Dependent RNA Polymerase. J. Mol. Biol. *425*, 2529–2540.

Ward, C.D., and Flanegan, J.B. (1992). Determination of the poliovirus RNA polymerase error frequency at eight sites in the viral genome. J. Virol. *66*, 3784–3793.

Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V. V., and Koonin, E. V. (2018). Origins and Evolution of the Global RNA Virome. MBio *9*.

Wong, I., Patel, S.S., and Johnson, K.A. (1991). An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. Biochemistry *30*, 526–537.

Woodman, A., Arnold, J.J., Cameron, C.E., and Evans, D.J. (2016). Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination. Nucleic Acids Res. *44*, 6883–6895.

Woodman, A., Lee, K.-M., Janissen, R., Gong, Y.-N., Dekker, N.H., Shih, S.-R., and Cameron, C.E. (2018a). Predicting Intraserotypic Recombination in Enterovirus 71. J. Virol. *93*, 445783.

Woodman, A., Lee, K.-M., Janissen, R., Gong, Y.-N., Dekker, N.H., Shih, S.-R., and Cameron, C.E. (2018b). Predicting Intra- and Intertypic Recombination in Enterovirus 71. BioRxiv 445783.

World Health Assembly (2019). Polio Eradication: Report by the Director-General.

Wu, J., and Gong, P. (2018). Visualizing the Nucleotide Addition Cycle of Viral RNA-Dependent RNA Polymerase. Viruses *10*, 24.

Xiao, Y., Rouzine, I.M., Bianco, S., Acevedo, A., Goldstein, E.F., Farkov, M., Brodsky, L., and Andino, R. (2016). RNA Recombination Enhances Adaptability and Is Required for Virus Spread and Virulence. Cell Host Microbe *19*, 493–503.

Xiao, Y., Dolan, P.T., Goldstein, E.F., Li, M., Farkov, M., Brodsky, L., and Andino, R. (2017). Poliovirus intrahost evolution is required to overcome tissue-specific innate immune responses. Nat. Commun. *8*, 375.

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.

Yang, X., Smidansky, E.D., Maksimchuk, K.R., Lum, D., Welch, J.L., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2012). Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. Structure *20*, 1519–1527.

Yang, Y., Rijnbrand, R., McKnight, K.L., Wimmer, E., Paul, A., Martin, A., and Lemon, S.M. (2002). Sequence Requirements for Viral RNA Replication and VPg Uridylylation Directed by the Internal cis-Acting Replication Element (cre) of Human Rhinovirus Type 14. J. Virol. *76*, 7485–7494.

Yin, Y.W. (2002). Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase. Science. *298*, 1387–1395.

Yin, Y.W., and Steitz, T.A. (2004). The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. Cell *116*, 393–404.

Yip, C., Lo, J., Sridhar, S., Lung, D., Luk, S., Chan, K.-H., Chan, J., Cheng, V., Woo, P., Yuen, K.-Y., et al. (2017). First Report of a Fatal Case Associated with EV-D68 Infection in Hong Kong and Emergence of an Interclade Recombinant in China Revealed by Genome Analysis. Int. J. Mol. Sci. *18*, 1065.

Yu, J., and Oster, G. (2012). A Small Post-Translocation Energy Bias Aids Nucleotide Selection in T7 RNA Polymerase Transcription. Biophys. J. *102*, 532–541.

Zhu, R., Salahub, D.R., Maroulis, G., and Simos, T.E. (2007). Mechanisms of Nucleotidyl Transfer Catalyzed by the Yeast RNA Polymerase II. In AIP Conference Proceedings, (AIP), pp. 104–110.

APPENDIX 1

PICORNAVIRUS RNA RECOMBINATION COUNTERACTS ERROR CATASTROPHE[2]

Author contributions to this appendix: The original project was thought up by Kempf and Barton, who also wrote the paper. Virology related experiments were carried out by Kempf. I contributed the in vitro kinetics data, summarized in figure A1.5 and its associated text.

A1.1 Introduction

RNA viruses are outstanding tools to study the molecular basis of mutation and selection due to their template-dependent replication mechanisms, their error prone polymerases, their ability to produce vast amounts of genetically diverse progeny in very short periods of time, and the strong selection for efficiently replicating progeny (Domingo and Holland, 1997; Drake and Holland, 1999). Consequences of template-dependent RNA replication mechanisms and error prone polymerases are a loss of fitness after repeated genetic bottlenecks (Duarte et al., 1992) and error catastrophe (Graci and Cameron, 2002), an overwhelming accumulation of mutations in viral RNA genomes incompatible with viability. Viral RNA recombination, a form of sexual replication (Chao, 1997) wherein progeny are produced from more than one parental genome (Kirkegaard and Baltimore, 1986; Nagy and Simon, 1997), likely arose to counteract the negative consequences of asexual template-dependent RNA replication mechanisms – namely error catastrophe (Barton, 1998; Chao, 1997; Simon-Loriere and Holmes, 2011; Xiao et al., 2016). Viral RNA recombination is advantageous because it can purge deleterious mutations

[2] This appendix is published in the Journal of Virology as: Kempf BJ, Watkins CL, Peersen OB, Barton DJ. 2019. Picornavirus RNA recombination counteracts error catastrophe. *J Virol* 93:e00652-19. https://doi.org/10.1128/JVI .00652-19.

while enriching beneficial mutations in viral genomes (Barton, 1998; Simon-Loriere and Holmes, 2011; Xiao et al., 2016). In this study, we explore the relationship between viral RNA recombination and error catastrophe.

Picornaviruses, along with other RNA viruses, arose before the radiation of eukaryotic supergroups (Koonin et al., 2008; Shi et al., 2018). Viral RNA recombination contributed to viral evolution in ancient times and remains important today, where it shapes and maintains picornavirus species groups (Brown et al., 2003; Oberste et al., 2004a; Simmonds, 2006). Because viral RNA recombination is a natural aspect of virus replication, we sought to define polymerase features that mediate viral RNA recombination. An L420A mutation in the poliovirus polymerase significantly reduces RNA recombination without any deleterious effects on template-dependent RNA replication or virus production (Kempf et al., 2016). Thus, poliovirus with an L420A mutation has normal asexual RNA replication mechanisms, but is defective for sexual RNA replication / recombination. In this study, we examine how this L420A polymerase mutation impacts error catastrophe.

Ribavirin, a mutagenic antiviral drug, restricts poliovirus replication via error catastrophe (Crotty et al., 2000, 2001). As a nucleoside analog pro-drug, ribavirin is converted into ribavirin triphosphate in cells and used as a substrate by the polymerase during viral RNA replication, where it can be incorporated opposite cytosine or uracil on the template strand (Ferrer-Orta et al., 2007). The incorporation of ribavirin into viral RNA induces G->A and C->U transition mutations during subsequent rounds of RNA replication (Graci and Cameron, 2002). Ribavirin-induced error catastrophe arises when the accumulation of mutations in viral RNA become incompatible with viability. A G64S mutation in poliovirus polymerase mediates ribavirin resistance by increasing viral polymerase fidelity, reducing the number of ribavirin-induced

mutations in viral RNA (Pfeiffer and Kirkegaard, 2003). In this study, we assess how a G64S polymerase mutation impacts viral RNA recombination and ribavirin-induced error catastrophe.

The ability of viral RNA recombination to counteract error catastrophe is theoretically sound (Barton, 1998; Simon-Loriere and Holmes, 2011; Xiao et al., 2016); however, there is conflicting evidence in the literature (Figure A1.1). Previously, we found that an L420A mutation in the poliovirus polymerase inhibits RNA recombination while exacerbating ribavirin-induced error catastrophe (Kempf et al., 2016). In contrast, the Andino lab reported that a D79H polymerase mutation inhibits RNA recombination, but without impacting ribavirin-induced error catastrophe (Xiao et al., 2016). Due to these conflicting reports, we further investigated the relationship between RNA recombination and ribavirin-induced error catastrophe. Our data indicate that RNA recombination counteracts error catastrophe, substantiating theories regarding the advantages and disadvantages of asexual and sexual RNA replication strategies (Barton, 1998; Chao, 1997; Drake and Holland, 1999).

A1.2 Materials and Methods

A1.2.1 Poliovirus RNAs and cDNA clones.

Poliovirus type 1 (Mahoney) and cDNA clones thereof were used for this study (Collis et al., 1992; Kempf et al., 2013, 2016; Steil et al., 2010). Viral RNAs were produced by T7 transcription of MluI-linearized cDNA clones (Ampliscribe T7 high-yield transcription kit, Cellscript Inc).

A.

3Dpol Mutation	Mutation Inhibits RNA Recombination	Mutation Exacerbates Ribavirin-Induced Error Catastrophe
D79H	Yes	No
L420A	Yes	Yes

B.

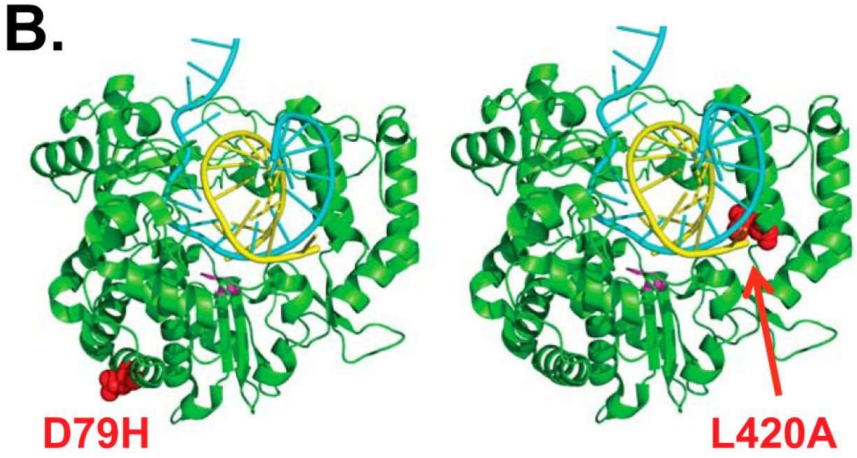

Figure A1.l: Conflicting evidence regarding viral RNA recombination and error catastrophe.

(A) Xiao et al. report that a D79H mutation in poliovirus polymerase inhibits viral RNA recombination but has no effect on ribavirin-induced error catastrophe, whereas Kempf et al. report that an L420A mutation in the poliovirus polymerase inhibits viral RNA recombination and exacerbates ribavirin-induced error catastrophe (Xiao, et. al., 2016, Kempf, et. al., 2016). (B) Locations of D79H and L420A mutations in poliovirus polymerase. Poliovirus polymerase PDB entry 4K4T rendered using the PyMOL molecular graphics system (Schrödinger, LLC). Polymerase is shown in green with D79H (red on left) and L420A (red on right) mutations; the active site GDD residues (fuchsia), template (blue), and product (yellow) RNAs indicated in color.

A1.2.2 Viral RNA recombination.

L929 cells were co-transfected with a mixture of two viral RNAs, each of which contained a lethal mutation (Figure A1.3A, ΔCapsid Donor and $3D^{pol}$ ΔGDD Recipient). ΔCapsid Donor is a subgenomic replicon containing an in-frame deletion of VP2 and VP3 capsid gene sequences (Δ nucleotides 1175-2956) (Collis et al., 1992; Kempf et al., 2016). $3D^{pol}$ ΔGDD Recipient is a full-length poliovirus RNA containing a 9 base deletion in $3D^{pol}$ (Δ ^{6965}GGU GAU GAU^{6973}) (Kempf et al., 2016). Deleting three catalytic residues from the viral polymerase (ΔGDD) results in a non-infectious, RNA replication incompetent derivative of poliovirus (Figure Al.3, $3D^{pol}$ ΔGDD Recipient) (Kempf et al., 2016). Wildtype and mutant derivatives of ΔCapsid Donor RNA were used – with the following $3D^{pol}$ substitution mutations: G64S, $G64^{Fix}$, $G64^{Fix}$ L420A, D79H, L419A and L420A. Mutations were engineered into ΔCapsid Donor cDNA clones as previously described (Kempf et al., 2013, 2016).

L929 cells were plated in 35mm 6-well dishes ~24 hours before transfection, with ~10^6 cells per well. ΔCapsid Donor and $3D^{pol}$ ΔGDD Recipient RNAs (1 ug of each) were transfected into each well, in triplicate (i.e., three independent samples for every experimental condition) (Transmessenger Transfection Reagent, Qiagen). Following transfection, 2 ml of culture medium (Dulbecco modified Eagle medium containing 100 units of penicillin and 100 ug per ml of streptomycin, 10% fetal bovine serum, and 10 mM $MgCl_2$) was added to each well and cells were incubated at 37°C in 5% CO_2. Virus was harvested at 72 hours post-transfection (hpt) and quantified by plaque assay in HeLa cells.

Mean titer from triplicates was plotted with standard deviation (error bars). Statistical significance was determined using pairwise t-test from GraphPad Prism software (La Jolla, CA).

A1.2.3 Replication controls

L929 cells were transfected with full-length infectious poliovirus RNAs (wt and mutant derivatives as noted in Figure A1.3C). Wildtype and mutant derivatives of poliovirus RNA were used – with the following $3D^{pol}$ substitution mutations: G64S, $G64^{Fix}$, $G64^{Fix}$ L420A, D79H, L419A and L420A. Mutations were engineered into infectious cDNA clones as previously described (Kempf et al., 2013, 2016).

L929 cells were plated in 35mm 6-well dishes ~24 hours before transfection, with ~10^6 cells per well. Poliovirus RNA (2 ug) was transfected into each well, in triplicate (i.e., three independent samples for every experimental condition) (Transmessenger Transfection Reagent, Qiagen). Following transfection, 2 ml of culture medium (Dulbecco modified Eagle medium containing 100 units of penicillin and 100 ug per ml of streptomycin, 10% fetal bovine serum, and 10 mM $MgCl_2$) was added to each well and cells were incubated at 37°C in 5% CO_2. Virus was harvested at 72 hours post-transfection (hpt) and quantified by plaque assay in HeLa cells. Mean titer from triplicates was plotted with standard deviation (error bars).

A1.2.4 Wildtype and mutant poliovirus stocks

Wildtype and mutant derivatives of poliovirus were grown in HeLa cells as previously described (Kempf et al., 2013). Poliovirus stocks with the following $3D^{pol}$ genotypes were prepared: WT, G64S, $G64^{Fix}$, $G64^{Fix}$ L420A, D79H, L419A and L420A.

HeLa cells were plated in 35mm 6-well dishes ~24 hours before transfection, with ~10^6 cells per well. Poliovirus RNA (2 ug) from cDNA clones was transfected into each well (Transmessenger Transfection Reagent, Qiagen). Following transfection, 2 ml of culture medium (Dulbecco modified Eagle medium containing 100 units of penicillin and 100 ug per ml of

streptomycin, 10% fetal bovine serum, and 10 mM $MgCl_2$) was added to each well and cells were incubated at 37°C in 5% CO_2. P_0 virus was harvested at 72 hours post-transfection (hpt), recovered after three rounds of freezing and thawing, cleared of cellular debris by centrifugation at 3,000 rpm, and quantified by plaque assay in HeLa cells. Larger P_1 virus stocks were obtained by infecting T150 flasks containing HeLa cell monolayers. Following infections, 20 ml of culture medium (Dulbecco modified Eagle medium containing 100 units of penicillin and 100 ug per ml of streptomycin, 10% fetal bovine serum, and 10 mM $MgCl_2$) was added to each flask and cells were incubated at 37°C in 5% CO_2 for 72 hrs at which time complete CPE was evident. P_1 virus was quantified by plaque assay in HeLa cells. cDNA synthesis and sequencing were used to confirm the stability of $3D^{pol}$ mutations in each virus (Kempf et al., 2013).

A1.2.5 Ribavirin-induced error catastrophe

HeLa cells were infected with wildtype or mutant derivatives of poliovirus and incubated in media with escalating concentrations of ribavirin (0-1000 uM) (Sigma-Aldrich), as previously described (Crotty et al., 2000, 2001; Xiao et al., 2016).

HeLa cells were plated in 35mm 6-well dishes ~24 hours before infection, with ~10^6 cells per well. Cells were infected with wt and mutant polioviruses at a multiplicity of infection (MOI) of 0.1 PFU per cell. Following 1 hour of virus adsorption, inoculums were removed and cells were incubated for 24 hours at 37°C in 2 ml of media with or without ribavirin (Sigma Aldrich). Viruses were titered by plaque assay.

Mean titer from triplicates was plotted with standard deviation (error bars). Statistical significance was determined using pairwise t-test from GraphPad Prism software (La Jolla, CA).

A1.2.6 3D^{pol} biochemistry

Biochemical characteristics of purified 3D^{pol} were examined as previously described (Hobdey et al., 2010; Kortus et al., 2012). Briefly, initiation experiments were done at room temperature by mixing 5 uM polymerase with 0.5 uM "10+1–12 RNA" (Hobdey et al., 2010) and 40 uM GTP in reaction buffer containing 50 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM Tris-(2-carboxyethyl)-phosphine (TCEP). 1 uL samples were removed from the reaction at various times and quenched with 19 uL of quench buffer consisting of 50 mM EDTA, 400 mM NaCl, 50 mM HEPES pH 6.5, and 2 mM TCEP. The amount of +1 product formed as a function of time was determined by gel electrophoresis and mathematically fit to a single exponential equation to obtain an initiation time constant. Elongation complex stability was determined by allowing the initiation reaction to proceed for 15 minutes and then diluting 10 uL of reaction into 90 uL of high salt buffer (300 mM NaCl, 4 mM $MgCl_2$, 50 mM HEPES pH 6.5, and 2 mM TCEP) to prevent further RNA binding. The amounts of elongation-competent complexes remaining at various time points up to four hours was assayed by mixing 5 uL aliquots of this diluted sample with 80 uM final concentrations of ATP, GTP, and UTP and allowing elongation to take place for 2 minutes before the addition of 10 uL quench buffer and 10 uL gel loading dye. Samples were separated by denaturing gel electrophoresis in 20% polyacrylamide 19:1, 7M urea and 1X TBE. RNA was labeled with an IRdye 800RS NHS ester (LI-COR Biosciences) and detected using a LI-COR Odyssey 9120 infrared imager system.

Kinetics assays were performed using stopped-flow methods to assess processive elongation with a fluorescein 5′ end-labeled RNA (Gong et al., 2009), and single nucleotide incorporation and rNTP-vs-2′-dNTP discrimination with an internally 2-aminopurine labeled RNA (McDonald et al., 2016). For these experiments, stalled polymerase-RNA elongation

complexes were formed at room temperature for 15 minutes with 15 uM polymerase, 10 uM RNA, 50 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES pH 6.5, 2 mM TCEP, and 60 uM each of ATP and GTP. Samples were then diluted 200-fold to a final RNA concentration of 50 nM in buffer containing 50 mM HEPES pH 7, 75 mM NaCl, and 4 mM $MgCl_2$. Elongation reactions were done using a Bio-Logic SFM-4000 titrating stopped-flow instrument with an MOS-500 spectrometer at 30°C. $MgCl_2$ was always present at 4 mM excess over the total NTP concentration, and NTP mixes in the processive elongation experiments contained equal concentrations of all four nucleotides.

A1.3 Results

A1.3.1 One-step growth of wildtype and mutant viruses

Before addressing the relationship between viral RNA recombination and error catastrophe, we investigated the one-step growth phenotypes of the wildtype and mutant viruses used in this investigation (Figure A1.2). Wildtype poliovirus (wt), along with seven mutant viruses, were used in the investigation: $G64^{Fix}$, G64S, D79H, L419A, L420A, $G64^{Fix}$ L420A and G64S L420A. The emergence of the G64S ribavirin resistance mutation typically requires a single transition nucleotide change (GGU^{Gly} -> AGU^{Ser}), and to prevent this we engineered the $G64^{Fix}$ mutation as a GGG codon that requires two or three point mutations for conversion to serine which is encoded by AGC, AGU, UCA, UCC, UCG, or UCU. Viruses containing G64S, D79H, L419A and L420A $3D^{pol}$ mutations have been reported in the literature (Kempf et al., 2013, 2016; Pfeiffer and Kirkegaard, 2003; Xiao et al., 2016). For this investigation, we engineered new viruses containing both $G64^{Fix}$ L420A and G64S L420A $3D^{pol}$ mutations. All of these mutations were genetically stable in the viruses grown in HeLa cells, as confirmed by

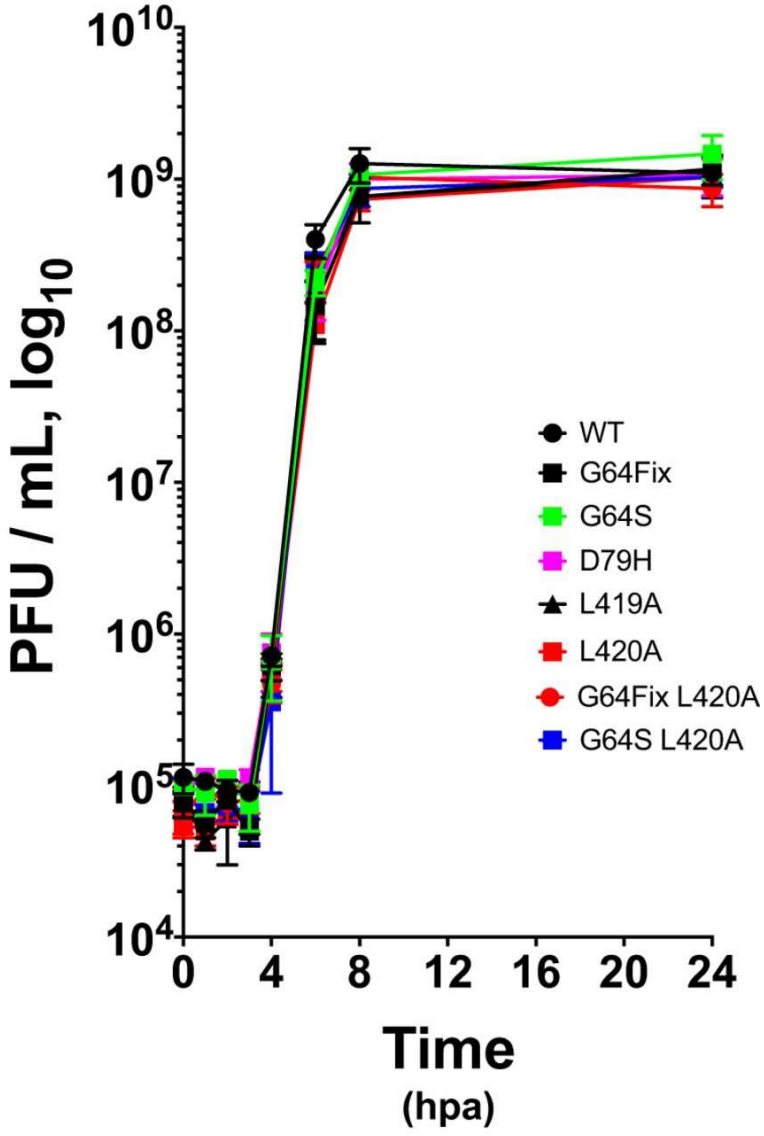

Figure A1.2: One-step growth of wild-type and mutant polioviruses.

HeLa cells were infected with wild-type or mutant poliovirus at an MOI of 10 PFU per cell. Virus was harvested at the indicated times by freeze-thawing cells. Titers were determined by plaque assay and plotted versus time. hpa, hours postadsorption.

cDNA sequencing. Under one-step growth conditions, each of the viruses used in this investigation grew by four orders of magnitude between 4 and 8 hpa, reaching titers ~ 10^9 PFU per ml (Figure A1.2). These data indicate that the wildtype and mutant viruses used in this investigation have similar one-step growth phenotypes. Other studies examined viral RNA synthesis by poliovirus containing G64S, L419A and L420A $3D^{pol}$ mutations (Kempf et al., 2013). The $3D^{pol}$ mutations used in this investigation are genetically stable within polioviruses and support viral RNA replication at rates and magnitudes comparable to wildtype polymerase.

A1.3.2 Impact of $3D^{pol}$ mutations on viral RNA recombination

To explore the relationship between viral RNA recombination and error catastrophe, we first examined the impact of $3D^{pol}$ mutations on the frequency of viral RNA recombination (Figure A1.3). The frequency of viral RNA recombination was measured using murine L929 cells co-transfected with viral RNAs containing lethal mutations in the capsid and $3D^{pol}$ genes (Figure A1.3A). The ΔCapsid Donor is an RNA replicon with a large in-frame deletion in the capsid genes. When transfected into cells, the ΔCapsid Donor RNA is translated, producing all of the viral proteins needed for viral RNA replication and recombination, including $3D^{pol}$. Furthermore, because the ΔCapsid Donor RNA has all of the cis-active elements needed for viral RNA replication, this viral RNA begins to replicate within cells. Nonetheless, the ΔCapsid Donor RNA is unable to form infectious virus on its own due to the lethal capsid gene deletion. The $3D^{pol}$ ΔGDD recipient RNA has a three amino acid deletion in the catalytic domain of the polymerase, preventing the expression of functional $3D^{pol}$. Otherwise, the $3D^{pol}$ ΔGDD recipient RNA is intact, containing a full complement of capsid genes and all of the cis-active elements needed for viral RNA replication. When co-transfected into murine cells, nascent RNA products from the ΔCapsid Donor RNA move from one template to another – resuming elongation on the

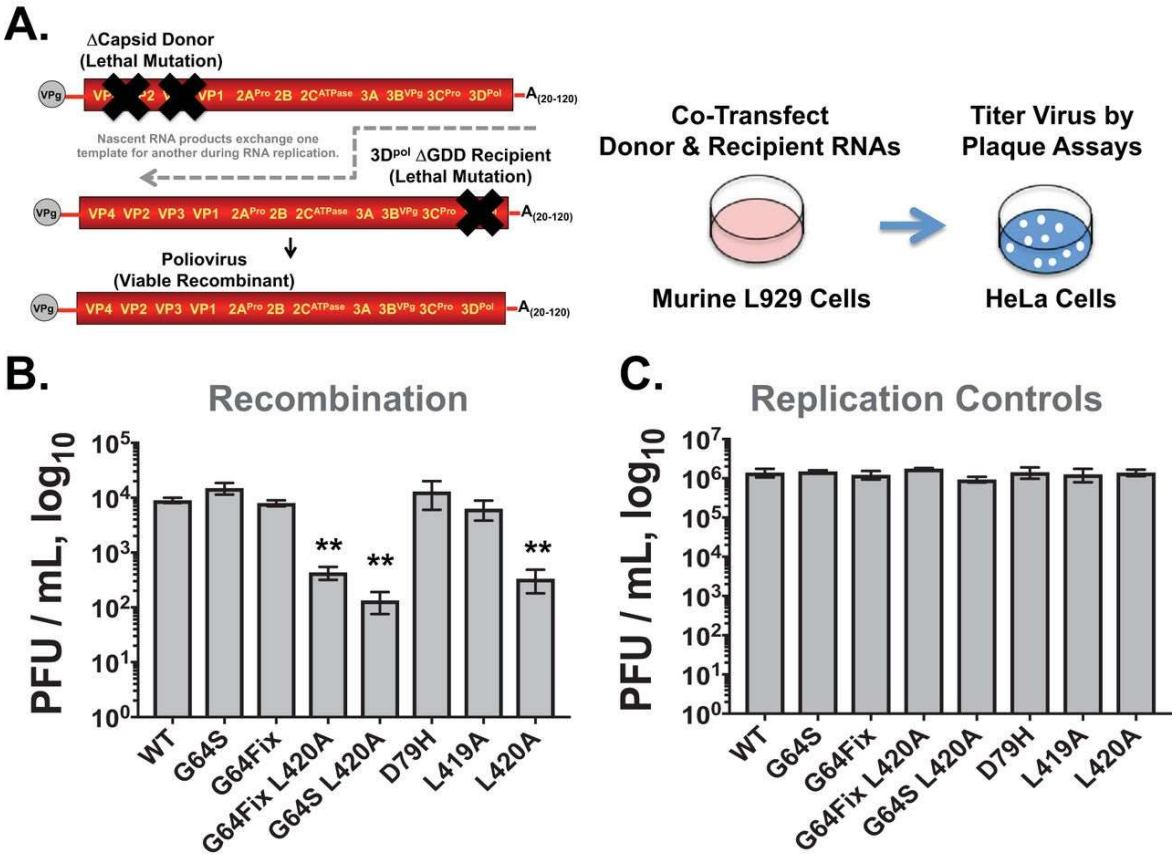

Figure A1.3: Impact of 3D^{pol} mutations on viral RNA recombination.

(A) Diagram of viral RNA recombination assays. The ΔCapsid "Donor" and ΔGDD "Recipient" RNAs have lethal mutations. When they are cotransfected into murine cells, poliovirus is produced via viral RNA recombination. The titer of virus produced is determined by plaque assays in HeLa cells. (B) Titers of poliovirus produced in cotransfected murine cells. ΔCapsid Donor contained wild-type or mutant polymerases as indicated on the *x* axis. (C) Replication controls. Titers of poliovirus produced in murine cells transfected with poliovirus RNA containing wild-type or mutant polymerases as indicated on the *x* axis. **, *P* values are <0.0001 compared with the wild type.

$3D^{pol}$ ΔGDD template, leading to the production of recombinant poliovirus (Figure A1.3A). Recombinant poliovirus genomes introduced by transfection can replicate in by not re-infect the murine cells, and the infectious virus produced is detected by subsequent plaque assays in HeLa cells (Figure A1.3A).

The impact of $3D^{pol}$ mutations on the frequency of viral RNA recombination was determined by engineering $3D^{pol}$ mutations into the ΔCapsid Donor RNA (Figure A1.3). When the ΔCapsid Donor RNA has a wildtype polymerase, ~ 1 x 10^4 PFU per ml of poliovirus was produced in co-transfected murine cells (Figure A1.3B, WT). In contrast, as previously shown (Kempf et al., 2016), an L420A mutation in the ΔCapsid Donor RNA led to substantially less poliovirus (Figure A1.3B, L420A). By comparison, mutations other than L420A, i.e. G64S, $G64^{Fix}$, D79H and L419A, did not significantly impact the recombination frequency (Figure A1.3B). Thus, in our assays, a D79H mutation in $3D^{pol}$ had no detectable impact on the frequency of viral RNA recombination, which is inconsistent with the data reported by Xiao et al. (Xiao et al., 2016).

We engineered additional mutations into $3D^{pol}$ to further investigate viral RNA recombination and error catastrophe, including a $G64^{Fix}$ codon mutation, a $G64^{Fix}$ codon mutation with L420A, and a G64S L420A double mutant (Figure A1.3). The G64S L420A double mutation was engineered to determine how a high-fidelity polymerase mutation (G64S) and a viral RNA recombination disabling mutation (L420A) function together. The L420A mutation inhibited the frequency of viral RNA recombination in both cases (Figure A1.3B, $G64^{Fix}$ L420A and G64S L420A). Thus, the L420A $3D^{pol}$ mutation inhibited viral RNA recombination in three contexts: by itself (L420A), when combined with a $G64^{Fix}$ codon

mutation ($G64^{Fix}$ L420A), and when combined with a high-fidelity polymerase mutation (G64S L420A).

Replication controls were performed to determine whether any of the $3D^{pol}$ mutations inhibit virus replication in murine cells (Figure A1.3C). Poliovirus RNAs encoding either the wildtype or mutant polymerase all produced ~ 1 X 10^6 PFU per ml of poliovirus, showing that the $3D^{pol}$ mutations under investigation did not affect the magnitude of poliovirus replication in murine cells (Figure A1.3C).

A1.3.3 Impact of $3D^{pol}$ mutations on ribavirin-induced error catastrophe

Ribavirin-induced error catastrophe was assayed in poliovirus-infected HeLa cells treated with ribavirin (Figure A1.4). The amounts of poliovirus produced within cells decreased as the dose of ribavirin increased from 0 to 1000 μM, consistent with the antiviral mechanism of ribavirin (Crotty et al., 2001). Poliovirus with a wildtype polymerase decreased ~ 50-fold from titers ~ 10^9 PFU per ml in the absence of ribavirin to titers of ~ $2x10^7$ PFU per ml in the presence of 1000 μM ribavirin (Figure A1.4, WT). Poliovirus with a G64S $3D^{pol}$ mutation resisted the antiviral activity of ribavirin (Figure A1.4, G64S), as previously established (Pfeiffer and Kirkegaard, 2003), and titers decreased ~ 10-fold. In contrast, poliovirus with an L420A $3D^{pol}$ mutation was substantially more sensitive to inhibition by ribavirin, showing a ~ 500-fold decrease in titers (Figure A1.4, L420A). Based on these data, we conclude that poliovirus with an L420A mutation is more susceptible than wildtype poliovirus to ribavirin-induced error catastrophe.

Poliovirus with $G64^{Fix}$, D79H and L419A $3D^{pol}$ mutations exhibited wildtype sensitivity to ribavirin (Figure A1.4, WT, $G64^{Fix}$, D79H and L419A). Furthermore, it is important to note

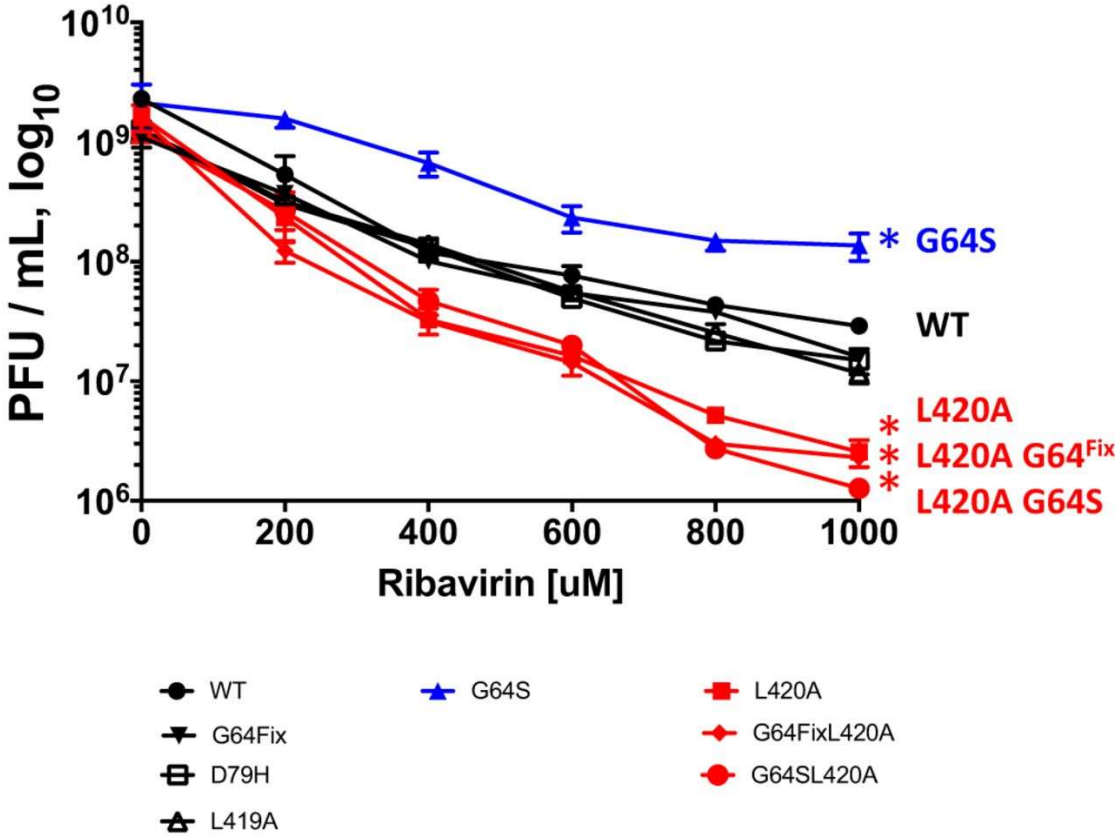

Figure A1.4: Impact of 3D^{pol} mutations on ribavirin-induced error catastrophe.

HeLa cells were infected with wild-type or mutant poliovirus at an MOI of 0.1 PFU per cell and incubated for 24 h with the indicated concentrations of ribavirin. Each condition was performed in triplicate. After 3 freeze-thaw cycles, viral titers were determined by plaque assay and plotted versus ribavirin concentration. *, *P* values are <0.01 compared with the wild type at all concentrations of ribavirin.

that adding a high fidelity G64S mutation to poliovirus with an L420A mutation did not make poliovirus more resistant to ribavirin (Figure A1.4, compare WT, G64S, L420A and G64S L420A; P value < 0.003 for G64S versus G64S L420A at all ribavirin concentrations). Based on these results, we conclude that a G64S mutation can make poliovirus resistant to ribavirin-induced error catastrophe, but only when viral RNA recombination is at near wildtype levels.

A1.3.4 Biochemical characteristics of mutant polymerases

We examined the biochemical characteristics of purified polymerase to determine the effects of individual mutations on RNA synthesis initiation, elongation complex stability, elongation rates and fidelity of nucleotide addition as measured by a discrimination factor reflecting the relative efficiency of rNTP versus 2′-dNTP incorporation (Campagnola et al., 2015). By comparison with wildtype polymerase, the D79H polymerase exhibited wildtype phenotypes in all assays (Figure A1.5 and Table A1.1, compare WT and D79H). In contrast, a G64S mutation decreased the processive elongation rate from 22 to 12 nt per sec (Table A1.1). By slowing the rate of elongation at a pre-catalytic step of RNA synthesis, the G64S polymerase has more time to discriminate between correct and incorrect nucleotides, leading to an increase in the fidelity of RNA synthesis (Arnold et al., 2005). Accordingly, a G64S polymerase exhibited an increased discrimination factor for both pyrimidines and purines (Figure A1.5C). Polymerase with an L420A mutation exhibited slightly slower initiation rates, decreased elongation complex stability and faster elongation rates, and a nucleotide discrimination factor only slightly lower than that of wildtype polymerase. These data suggest that the faster elongation rates of the L420A polymerase may slightly reduce fidelity, perhaps contributing to increased susceptibility to ribavirin. However, the G64S L420A double mutant exhibited a hybrid biochemical profile with faster elongation rates like L420 polymerase and a high discrimination factor like G64S

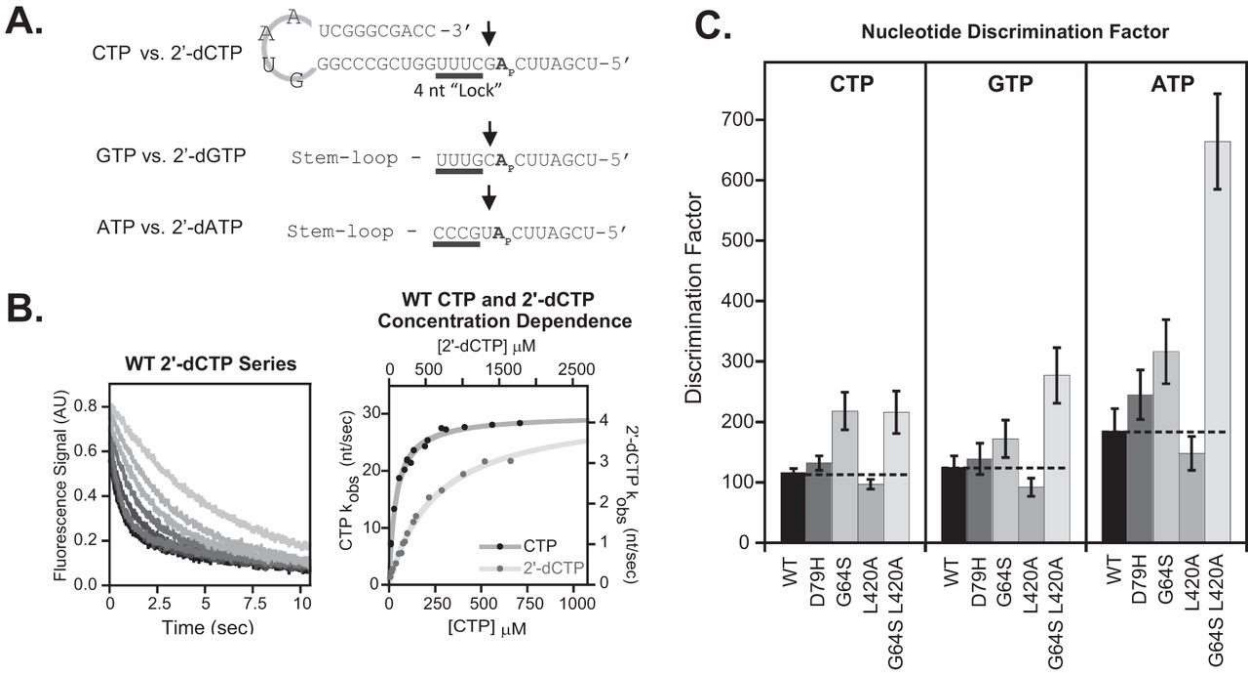

Figure A1.5: Biochemical assays of wild-type and mutant polymerases.

(A) RNA hairpin structures used for single nucleotide discrimination assays of CTP, GTP, and ATP. Each hairpin contained an identical stem-loop, shown for the first hairpin only, while four initiation nucleotides ("lock") and templating bases were varied for each construct. (B) Example 2-aminopurine fluorescence trace and concentration dependence plot from single-nucleotide incorporation stopped-flow assay. The signal change reflects the postcatalysis quenching of 2-aminopurine fluorescence when it is translocated from the +2 pocket into the +1 position in the polymerase active site. (C) NTP versus 2′-dNTP nucleotide discrimination factors for each polymerase, organized by NTP. Dashed lines denote value of wild-type $3D^{pol}$ discrimination for each NTP.

Table A1.1 Biochemical phenotypes of wild-type and mutant polymerases[a]

[a] Single-nucleotide incorporation and processive elongation were assayed in 75 mM NaCl (pH 7) at 30°C. Values are ± standard errors of parameters from curve fitting. EC, elongation complex.

$3D^{pol}$ mutation(s)	RNA initiation (min)	EC stability (min)	Processive elongation (nt/sec)	Processive K_m value (μM)
WT	6 ± 1	88 ± 9	22 ± 1	67 ± 2
D79H	5 ± 1	70 ± 30	21 ± 1	64 ± 3
G64S	8 ± 1	24 ± 7	12 ± 1	46 ± 4
L420A	8 ± 1	28 ± 4	30 ± 1	75 ± 3
G64S and L420A	6 ± 3	40 ± 8	26 ± 1	84 ± 9

polymerase (Figure A1.5 and Table A1.1). Thus, while L420A polymerase discrimination is slightly reduced as compared to WT, G64S L420A polymerase discrimination is comparable to that of the high fidelity G64S polymerase (Figure A1.5C). From these data, we conclude that the G64S L420A polymerase is a high-fidelity polymerase, like G64S. Notably, the virus containing this high fidelity G64S L420A polymerase exhibits the high ribavirin sensitivity like virus containing the L420A mutation alone (Figure A1.4). Taken together, these data support the conclusion that viral RNA recombination counteracts ribavirin-induced error catastrophe.

A1.4 Discussion

Our data reveal two distinct mechanisms by which picornaviral RdRPs influence error catastrophe: fidelity of RNA synthesis and RNA recombination. A G64S mutation increased viral polymerase fidelity and rendered the virus resistant to ribavirin-induced error catastrophe, but only when RNA recombination was present at wildtype levels. An L420A mutation in the polymerase inhibited RNA recombination and exacerbated ribavirin-induced error catastrophe. Furthermore, the high fidelity G64S polymerase failed to make virus resistant to ribavirin when RNA recombination was substantially reduced by an L420A mutation. Taken together, these data provide strong evidence to conclude that viral RNA recombination counteracts error catastrophe.

A1.4.1 Asexual versus sexual RNA replication mechanisms.

Picornaviruses have asexual and sexual RNA replication mechanisms, both of which are catalyzed by $3D^{pol}$, the viral RNA-dependent RNA polymerase (Figure A1.6A). Template-dependent RNA replication is an efficient form of replication while replicative RNA recombination is used more sparingly to purge mutations from viral RNA. We find that an L420A mutation in $3D^{pol}$ reduces viral RNA recombination without inhibiting template-

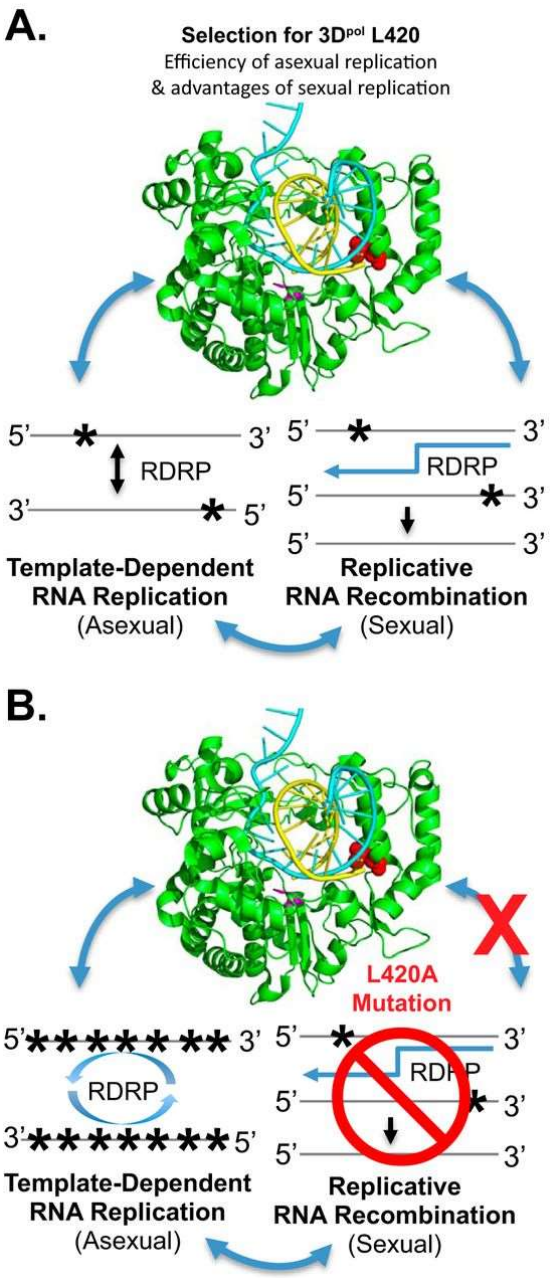

Figure A1.6: Reiterative cycles of template-dependent RNA replication lead to ribavirin-induced error catastrophe when RNA recombination is impaired.

(A) $3D^{pol}$ is required for both template-dependent RNA replication and replicative RNA recombination. $3D^{pol}$ L420 was selected during evolution to mediate replicative RNA recombination. (B) An L420A mutation in $3D^{pol}$ disables replicative RNA recombination without impacting template-dependent RNA replication. Reiterative cycles of template-dependent RNA replication, in the context of impaired RNA recombination, lead to ribavirin-induced error catastrophe. Mutations in viral RNA indicated with asterisks (*). Poliovirus polymerase PDB entry 4K4T is rendered as in Figure A1.1B.

dependent RNA replication. Reiterative cycles of template-dependent RNA replication, in the context of impaired RNA recombination, exacerbates ribavirin-induced error catastrophe (Figures A1.4 and A1.6B).

The frequency of viral RNA recombination is controlled in part by the size and shape of membrane-anchored replication complexes that facilitate the co-localization of two or more parental templates (Egger and Bienz, 2002; Garcia-Ruiz et al., 2018), and in part by molecular features of $3D^{pol}$. An L420A mutation in $3D^{pol}$ reduces RNA recombination-dependent viral titers ~50-fold without inhibiting template-dependent RNA replication in cells (Figure A1.3B,C). It is important to note that an L420A mutation does not completely inhibit viral RNA recombination; residual amounts of replicative RNA recombination remain and non-replicative RNA recombination is unaffected by the L420A mutation (Kempf et al., 2016). It is also noteworthy that an L420A mutation failed to render poliovirus susceptible to error catastrophe in the absence of ribavirin (Kempf et al., 2016). Serial genetic bottlenecks can result in error catastrophe (Duarte et al., 1992); however, limiting dilution serial passage of L420A poliovirus in HeLa cells, in the absence of ribavirin, did not lead to an increased accumulation of mutations in poliovirus genomes or to error catastrophe (Kempf et al., 2016). While an L420A mutation does not lead to the extinction of virus populations during limiting dilution serial passage, it does render poliovirus more susceptible to ribavirin-induced error catastrophe. This ability to avoid error catastrophe in the absence of ribavirin may be due to residual levels of replicative and non-replicative RNA recombination, or due to the highly permissive nature of HeLa cells. Error catastrophe is not expected when large virus populations replicate competitively within highly permissive cells.

Asexual template-dependent RNA replication can sustain wildtype magnitudes of virus replication in cells in the absence of mutagenic pressures (Figure A1.3C). Wildtype magnitudes of viral RNA recombination, on the other hand, were needed to avoid error catastrophe in the presence of ribavirin (Figure A1.4). These data indicate that some viral polymerase features evolved specifically to mediate RNA recombination. Based on the protein-RNA interactions found in the structures of $3D^{pol}$–RNA elongation complexes, we previously identified poliovirus $3D^{pol}$ leucine 420 as an important residue that mediates efficient viral RNA recombination. Leu420 is located on a long helix in the polymerase thumb domain that packs into the minor groove of the RNA duplex as it exits the polymerase, and the Leu420 sidechain forms a direct contact with the product-strand ribose that is three bases out from the active site (Figure A1.6). Based on these structural insights, we suspect that Leu420 helps to distinguish between homologous and non-homologous partners in recombination by playing an indirect role in favoring proper Watson-Crick base pairing between template and product strands at the active site. The degree of homology between viral RNA products and partners in recombination ranges from 3-30 nucleotides. When Leu420 reinforces interactions between homologous partners, three nucleotides from the active site, the polymerase can efficiently resume elongation and produce recombinant genomes. If the recombination partners are not homologous, then Leu420 will destabilize the binding geometry of the nascent RNA near the active site, reducing or precluding further elongation of the partial genome.

A1.4.2 L420A mutation in RdRP disables replicative RNA recombination.

Theory suggests that viral RNA recombination is advantageous because it can en masse purge deleterious mutations that arise during RNA replication (Chao, 1997). The replicative RNA recombination assay that we use requires nascent RNA products to move from one

template to another during RNA synthesis to produce infectious virus, as both individual templates have lethal mutations (Figure A1.3A). We find that an L420A mutation in the poliovirus polymerase disables RNA recombination ~50-fold without impacting template-dependent RNA replication (Figure A1.3). In contrast, neither G64S nor D79H mutations in the viral polymerase inhibited replicative RNA recombination, contrary to findings reported from other labs (Lowry et al., 2014; Woodman et al., 2016; Xiao et al., 2016). Technical differences in the design of recombination assays may be important in this regard. In our recombination assay (Kempf et al., 2016), functional $3D^{pol}$ is only expressed from one viral RNA template, the ΔCapsid Donor (Figure A1.3A). In contrast, functional $3D^{pol}$ is expressed from both RNA templates in the recombination assays from the Evans and Andino labs (Lowry et al., 2014; Woodman et al., 2016; Xiao et al., 2016). The expression of functional $3D^{pol}$ from both viral RNA templates may confound the recombination assay, increasing the concentrations of the different viral RNAs and allowing for competition between the two replicons, in addition to detecting recombination events. Expressing functional $3D^{pol}$ from only one of the recombination partners may alleviate these confounding issues. It is important to note that we have not yet resolved these technical discrepancies and future investigations are required to better understand the basis for quantitative differences in viral RNA recombination associated with D79H and G64S $3D^{pol}$ mutations in the two systems.

A1.4.3 L420A mutation in RdRP exacerbates ribavirin-induced error catastrophe.

One key disadvantage of asexual reproduction is the accumulation of mutations in progeny genomes, which can be detrimental to fitness or viability. Error catastrophe can arise due to repeated genetic bottlenecks (Duarte et al., 1992) or as a consequence of ribavirin's antiviral effects (Crotty et al., 2001; Graci and Cameron, 2002). We used ribavirin to assess the

impact of viral RNA recombination on error catastrophe (Figure A1.4) and found that the L420A recombination mutation in $3D^{pol}$ exacerbated ribavirin-induced error catastrophe. This effect was evident independent of whether a G64S mutation was used to increase the fidelity of RNA synthesis. A G64S mutation increases viral polymerase fidelity and renders an otherwise wildtype virus resistant to ribavirin-induced error catastrophe (Figures A1.4 and A1.5, Table A1.1). However, this was not the case when RNA recombination was disabled by an L420A mutation, showing that viral RNA recombination is required to counteract error catastrophe.

One caveat to these interpretations is the possibility that an L420A polymerase increases the incorporation of ribavirin into viral RNA. This would confound the interpretation that the L420A mutation renders virus susceptible to ribavirin-induced error catastrophe due to its impact on viral RNA recombination. Ribavirin incorporation is expected to increase or decrease coincident with changes in the speed and fidelity of the viral polymerase (Arnold et al., 2005). Biochemical assays of polymerase nucleotide discrimination partially address this caveat, suggesting that the L420A mutation slightly reduced polymerase fidelity, while polymerase containing both G64S and L420A mutations has a high-fidelity phenotype, like G64S (Figure A1.5 and Table A1.1). Importantly, virus containing the high fidelity G64S L420A polymerase exhibits increased ribavirin sensitivity like that of virus containing the L420A mutation alone (Figure A1.4). These data indicate that a high fidelity polymerase does not mediate ribavirin resistance when RNA recombination is disabled by an L420A mutation. Biochemical studies of RTP utilization indicate that RTP is a poor substrate for the polymerase, with incorporation rates being 3000- to 6000-fold slower than for the cognate rNTPs (Crotty et al., 2000; Graci and Cameron, 2002). We find that ribavirin incorporation into RNA products by purified polymerase is extremely low – precluding (to date) reliable direct comparisons of RTP utilization by

wildtype and L420A polymerases. A comprehensive biochemical analysis of the $3D^{pol}$ catalytic cycle is ongoing to further explore how the L420A mutation impacts fidelity and ribavirin incorporation. Furthermore, it is possible that mechanisms other than ribavirin-induced error catastrophe (Graci and Cameron, 2006), such as an inhibition of viral RNA synthesis by ribavirin (Vo et al., 2003), could influence the replication of viruses containing an L420A polymerase. Future experiments with other mutagens could provide insights in this regard. However, compelling evidence of ribavirin-induced error catastrophe in the literature (Crotty et al., 2000, 2001; Graci and Cameron, 2002) combined with our experimental data (Figures A1.3-A1.5) make us conclude that viral RNA recombination is debilitated by an L420A mutation and that viral RNA recombination counteracts ribavirin-induced error catastrophe.

A1.4.4 $3D^{pol}$ L420 is conserved across picornavirus species groups.

It is notable that an L420A mutation can disable RNA recombination ~50-fold (Figure A1.3B) without impairing the template-dependent RNA replication that sustains virus replication in cells (Figure A1.3C). This suggests that some polymerase features evolved specifically to mediate sexual reproduction, as would be expected (Chao, 1997). There are currently 94 species groups in the *Picornaviridae* family (www.picornaviridae.com), and because viral RNA recombination is advantageous one would expect the viral polymerases across these groups to have evolutionarily conserved features that facilitate recombination. Consistent with this notion, $3D^{pol}$ L420 is conserved in sequence and structure across picornavirus species groups (Table A1.2) (Chen et al., 2013; Ferrer-Orta et al., 2007; Gong et al., 2013; Kempf and Barton, 2015; Kempf et al., 2013; Love et al., 2004; Vives-Adrian et al., 2014; Wang et al., 2017). Based on the sequence and structural conservation, we predict that all picornaviruses have the capacity to use RNA recombination to avoid error catastrophe. Consistent with this, an L421A $3D^{pol}$ mutation in

Table A1.2 Conservation of 3D^{pol} L420 sequence and structure[a]

[a] Modified from Kempf and Barton (Kempf and Barton, 2015).

[b] Polio L420, and corresponding residue in other viruses, in bold.

Species group	Virus	Thumb α-helix[b]	PDB	Citation
Enterovirus A	EV-A71	^{410}NTQDHVRSLCL**L**AWHN^{425}	4IKA	33
Enterovirus B	CVB3	^{410}NTQDHVRSLCL**L**AWHN^{425}	4K4X	26
Enterovirus C	Polio	^{409}NTQDHVRSLCL**L**AWHN^{424}	4K4T	26
Enterovirus D	EV-D68	^{405}NTQDHVRSLCY**L**AWHN^{420}	5XEO	37
Rhinovirus A	HRV-16	^{408}QMQEHVLSLCH**L**MWHN^{423}	4K50	26
Rhinovirus B	HRV-14	^{408}NTQDHVRSLCM**L**AWHS^{423}	1XR5	35
Aphthovirus	FMDV	^{419}TIQEKLISVAG**L**AVHS^{434}	2E9T	19
Cardiovirus	EMCV	^{411}TLSEKLTSITM**L**AVHS^{426}	4NZO	36

EV-A71, analogous to the L420A $3D^{pol}$ mutation in poliovirus, inhibits EV-A71 RNA recombination coincident with an increase in sensitivity to ribavirin (Woodman et al., 2018b). It is theoretically satisfying that picornaviruses, which evolve into species groups due to frequent RNA recombination in nature (Brown et al., 2003; Oberste et al., 2004b; Simmonds, 2006), can avoid error catastrophe via RNA recombination. Viral RNA recombination has the potential to enrich beneficial mutations while purging deleterious mutations (Xiao et al., 2016). Thus, it is not surprising that picornaviral polymerases have conserved elements, including L420, to mediate RNA recombination.

A1.5 Summary

Why do RNA viruses recombine (Simon-Loriere and Holmes, 2011)? Our data indicate that viral RNA recombination counteracts error catastrophe, substantiating long held theories regarding the advantages and disadvantages of asexual and sexual RNA replication strategies (Barton, 1998; Chao, 1997; Drake and Holland, 1999).

APPENDIX 2

AN EXTENDED PRIMER GRIP OF PICORNAVIRUS POLYMERASE FACILITATES SEXUAL RNA REPLICATION MECHANISMS[3]

Author's contributions to the following text: Authors Kempf and Barton conceived of this project and wrote the manuscript. Author Kempf conducted all virology related experiments. I contributed the in vitro biochemistry data, summarized in table A2.3.

A2.1 Introduction

RNA viruses arose billions of years ago, becoming ubiquitous parasites of cellular life (Krupovic et al., 2019). The RNA-dependent RNA polymerase of RNA viruses is monophyletic, providing a means to consider the evolutionary history of all RNA viruses, to compare distinct groups of RNA viruses, and to classify RNA viruses (Kamer and Argos, 1984; Wolf et al., 2018). Picornaviruses, whose origins predate the radiation of eukaryotic supergroups (Koonin et al., 2008), have been studied extensively. Poliovirus, in particular, has been studied intensively for 100 years (Racaniello, 2006). The poliovirus RNA-dependent RNA polymerase was initially purified and biochemically characterized in 1977 (Flanegan and Baltimore, 1977). Since then, labs across the globe have studied the poliovirus RNA-dependent RNA polymerase in great detail [reviewed in (Peersen, 2017)]. Picornavirus RNA-dependent RNA polymerases are multifunctional, catalyzing distinct steps of viral RNA replication, from making primers (VPg-uridylylation) (Chen et al., 2013; Paul et al., 2003; Steil and Barton, 2008; Sun et al., 2012; Yang

[3] This appendix is published in the Journal of Virology as: Kempf BJ, Watkins CL, Peersen OB, Barton DJ. 2020. An Extended Primer Grip of Picornavirus Polymerase Facilitates Sexual RNA Replication Mechanisms. *J Virol*. https://doi.org/10.1128/JVI.00835-20.

et al., 2002), to replicating viral RNA (Van Dyke and Flanegan, 1980; Van Dyke et al., 1982), to polyadenylation of progeny RNA genomes (Kempf et al., 2013). Solving the atomic structure of poliovirus polymerase (Thompson and Peersen, 2004), and its elongation complex (Gong and Peersen, 2010), led to unprecedented opportunities, including our ongoing work to understand viral RNA replication mechanisms and error catastrophe (Kempf et al., 2019).

Picornaviruses have both asexual and sexual RNA replication mechanisms, both of which are catalyzed by the viral RNA-dependent RNA polymerase (Figure A2.1). Asexual RNA replication mechanisms involve one parental template whereas sexual RNA replication mechanisms involve two or more parental templates. Previously, we established that an L420A polymerase mutation exacerbates ribavirin-induced error catastrophe coincident with defects in sexual RNA replication mechanisms (Figure A2.1) (Kempf et al., 2016, 2019). Asexual RNA replication mechanisms are advantageous because vast amounts of progeny can be produced very quickly from one parental RNA template. However, asexual RNA replication mechanisms, in conjunction with error-prone RNA-dependent RNA polymerases (Drake and Holland, 1999; Ward and Flanegan, 1992), can be disadvantageous, contributing to a loss of fitness due to Muller's ratchet (Chao, 1990; Duarte et al., 1992). Errors introduced into viral RNA during asexual RNA replication cannot be easily removed in the absence of viral RNA recombination, except by reversion or negative selection (Agol and Gmyl, 2018). In conjunction with genetic bottlenecks, reiterative asexual RNA replication leads to error catastrophe, an overwhelming accumulation of mutations in viral RNA. Negative-strand RNA viruses like vesicular stomatitis virus, which do not recombine, are especially susceptible to Muller's ratchet (Duarte et al., 1992). In contrast, picornaviruses, which have relatively high levels of replicative RNA recombination (Jarvis and Kirkegaard, 1992; Kirkegaard and Baltimore, 1986), are somewhat

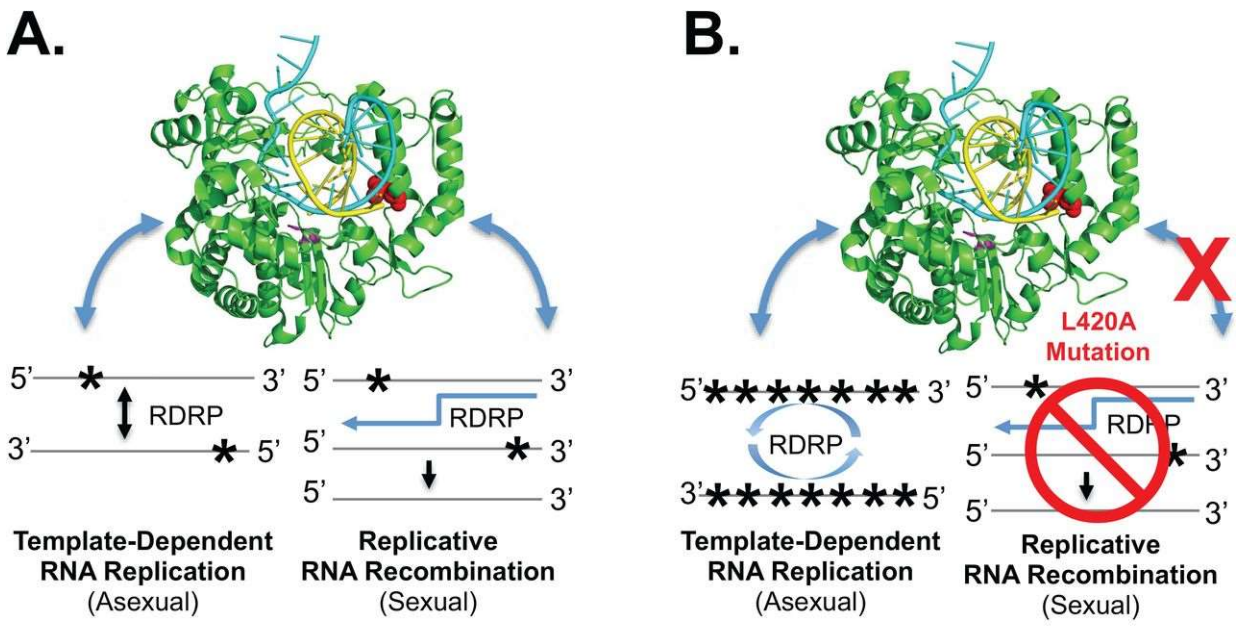

Figure A2.1: Picornaviruses have both asexual and sexual RNA replication mechanisms.

(A) Asexual RNA replication involves one parental template whereas sexual RNA replication involves two parental templates. Mutations (asterisks) introduced during asexual RNA replication can be removed by sexual RNA replication. (B) Ribavirin-induced error catastrophe. The fidelity of RNA synthesis (Pfeiffer and Kirkegaard, 2003, Arnold, et. al., 2005) and the frequency of RNA recombination (Kempf, et. al., 2019, Kempf, et. al., 2016) influence ribavirin-induced error catastrophe. Because an L420A mutation exacerbates ribavirin-induced error catastrophe coincident with defects in sexual RNA replication (Kempf, et. al., 2019), we hypothesized that ribavirin-resistant poliovirus selected from L420A background would acquire novel mutations that restore efficient sexual RNA replication. Modified from Kempf et al. 2019 (Kempf, et. al., 2019).

resistant to Muller's ratchet (Escarmís et al., 2008; Escarmís et al., 2002; Lázaro et al., 2002). Nonetheless, ribavirin, a mutagenic antiviral drug, can restrict picornavirus replication via error catastrophe (Crotty et al., 2000, 2001). Replicative RNA recombination, a form of sexual RNA replication (Barton, 1998; Chao, 1997), can counteract ribavirin-induced error catastrophe (Kempf et al., 2016, 2019), presumably by purging lethal mutations from viral RNA genomes (Dolan et al., 2018; Xiao et al., 2016).

A growing body of evidence indicates that sexual RNA replication mechanisms play an essential role in picornavirus speciation and ongoing genetic exchange between viruses within defined species groups (Benschop et al., 2008; Brown et al., 2003; Van Dung et al., 2014; McIntyre et al., 2013; Oberste et al., 2004b, 2004a; Simmonds and Welch, 2006). Mucosal surfaces in the respiratory (Martin et al., 2013) and gastrointestinal tracts (Van Dung et al., 2014) are important ecological environments for ongoing genetic exchange between related picornaviruses. Gastrointestinal bacteria may even enhance virus co-infection to promote recombination (Erickson et al., 2018). RNA sequence complementarity between genomes undergoing recombination (Kirkegaard and Baltimore, 1986; Lowry et al., 2014), among other factors (Runckel et al., 2013), influence the frequency of recombination, with higher rates of recombination when sequences are more alike. When two related viruses co-infect a cell, sexual RNA replication mechanisms lead to the generation of chimeric RNA genomes, which are often fit, culminating in sustained host-to-host transmission. Circulating vaccine-derived polioviruses, obstacles to poliovirus eradication, are products of sexual RNA replication mechanisms – formed when unfit portions of OPV RNA genomes recombine with more fit regions of non-polio group C enteroviruses (Combelas et al., 2011; Jegouic et al., 2009; Korotkova et al., 2017). When recombinant progeny are fit, as occurs in the case of circulating vaccine-derived polioviruses, the

recombinant strains circulate from host-to-host (Burns et al., 2013; Famulare et al., 2016). Thus, an important biological consequence of sexual RNA replication mechanisms is the frequent exchange of genetic material among related picornaviruses. Picornavirus species groups are sustained over time by repeated, ongoing genetic exchange between related viruses within the species group. A leucine 420 residue in the polymerase thumb domain, conserved across picornavirus species groups, is required for efficient RNA recombination / sexual RNA replication (Kempf et al., 2016, 2019).

The interplay between asexual and sexual RNA replication (Figure A2.1) is central to what Eigen deemed a grand challenge for the 21st century, elucidating the complex mechanisms of error catastrophe, which are variable from one type of virus to another (Eigen, 2002). Error catastrophe and lethal mutagenesis are commonly used in the literature to describe decreased virus titers associated with replication in the presence of ribavirin (Bull et al., 2007; Crotty et al., 2001; Eigen, 2002). In 2007, Bull et al. (Bull et al., 2007) set out to develop the theory of lethal mutagenesis, espousing that it be distinct from the theory of error catastrophe. In their theory of lethal mutagenesis, Bull et al. assume that viral RNA recombination is absent (Bull et al., 2007). By assuming that viral RNA recombination is absent, lethal mutagenesis theory fails to appreciate the important interplay between asexual and sexual RNA replication mechanisms. Asexual RNA replication mechanisms are efficient; however, they render viruses susceptible to error catastrophe / lethal mutagenesis (Barton, 1998; Chao, 1997; Crotty et al., 2001). Sexual RNA replication (aka recombination), while inefficient, counteracts the primary disadvantage of 113 asexual RNA replication, namely error catastrophe / lethal mutagenesis (Barton, 1998; Chao, 1997). Overall, we find that the theory of error catastrophe is more realistic because it better emphasizes the distinctions and interplay between asexual and sexual RNA replication strategies.

Using this theoretical framework, we propose that viral RNA recombination counteracts ribavirin-induced error catastrophe by purging ribavirin-induced mutations from viral RNA genomes (Figure A2.1). Viral RNA recombination can re-assemble mutant-free genomes to refresh the pool of RNA undergoing asexual RNA replication (Figure 1). Thus, consistent with theoretical underpinnings (Barton, 1998; Chao, 1997), genetic exchange during sexual RNA replication is advantageous because it provides a mechanism to counteract error catastrophe.

Building upon previous studies showing an L420A polymerase mutation exacerbates ribavirin-induced error catastrophe coincident with defects in sexual RNA replication mechanisms (Kempf et al., 2016, 2019), we predicted that ribavirin-resistant poliovirus selected from L420A parental strains will 125 acquire polymerase mutations that facilitate sexual RNA replication mechanisms (Figure A2.1). In the work presented here we used several ribavirin-sensitive and ribavirin-resistant parental viruses, in conjunction with serial passage in ribavirin, to select for ribavirin resistant poliovirus. Novel ribavirin resistance mutations, along with L420A revertants and pseudorevertants, were identified and characterized. These data implicate an extended primer grip of the viral polymerase in sexual RNA replication mechanisms.

A2.2 Materials and Methods

A2.2.1 Poliovirus and infectious cDNA clones.

Poliovirus type 1 (Mahoney), and mutant derivatives thereof, were derived from an infectious cDNA clone (Collis et al., 1992). Poliovirus RNAs were produced by T7 transcription of MluI-linearized cDNA clones (Ampliscribe T7, Cellscript Inc.) and transfected into HeLa cells to make infectious virus as previously described (Kempf et al., 2013, 2016, 2019).

A2.2.2 Serial passage of poliovirus in escalating concentrations of ribavirin.

Poliovirus was grown in HeLa cells in the presence of escalating concentrations of ribavirin (Sigma-Aldrich) using methods modified from those of Pfeiffer and Kirkegaard (Pfeiffer and Kirkegaard, 2003). HeLa cells were plated in 35mm 6-well dishes 24 hours before infection with WT or mutant parental strains of poliovirus at an MOI of 0.1 PFU per cell and incubated in 0 uM (P0), 100 uM (P1 to P4), 400 uM (P5 to P9) or 800 uM ribavirin (P10). Virus was harvested at 24 hpi by three freeze-thaw cycles and titered by plaque assay in HeLa cells. Three independent lineages of each parental strain were subjected to 10 serial passages in escalating concentrations of ribavirin. We expanded the population of poliovirus for each lineage after passage 10 by infecting a T150 flask of HeLa cells.

Viral cDNA was prepared from expanded passage 10 ($P10^e$) viruses from each lineage and sequenced to identify polymerase mutations fixed in the virus populations.

A2.2.3 One-step growth of poliovirus.

HeLa cells were plated in 35mm 6-well dishes 24 hours before infection with wildtype and mutant polioviruses. An MOI of 10 PFU per cell was used for one-step growth conditions. After 1 hour for virus adsorption, the inoculum was removed and the cells were incubated with 2 ml of culture media at 37°C. Total virus was harvested by three freeze-thaw cycles at the indicated times post-infection. Titers were determined by plaque assays.

Mean titers from triplicates were plotted versus time post-infection with standard deviation error bars. Statistical significance was assessed using the Holm-Sidak method of pairwise comparisons from GraphPad Prism (La Jolla, CA).

A2.2.4 Ribavirin dose-response assays.

HeLa cells were plated in 35mm 6-well dishes 24 hours before infection with wildtype and mutant polioviruses. An MOI of 0.1 PFU per cell was used for ribavirin-dose response assays. After 1 hour for virus adsorption, the inoculum was removed and the cells were incubated with 2 ml of culture media at 37°C. Total virus was harvested by three freeze-thaw cycles at 24 hours post-infection. Titers were determined by plaque assays.

Mean titers from triplicates were plotted with standard deviation error bars. Statistical significance was determined using the Holm-Sidak method of pairwise comparisons from GraphPad Prism (La Jolla, CA).

A2.2.5 Viral RNA recombination and replication controls.

Viral RNA recombination assays and replication controls were performed in L929 cells as previously described (Kempf et al., 2016, 2019). For viral RNA recombination, L929 murine cells were co-transfected with two viral RNAs, each of which contained a lethal mutation: ΔCapsid Donor and $3D^{pol}$ ΔGDD Recipient. ΔCapsid Donor is a subgenomic replicon containing an in-frame deletion of VP2 and VP3 capsid gene sequences (Δ nucleotide positions 1175 to 2956) (Collis et al., 1992; Kempf et al., 2016). $3D^{pol}$ ΔGDD recipient is a full-length poliovirus RNA containing a 9-base deletion in $3D^{pol}$ (Δ 6965GGU GAU GAU6973). Deleting three catalytic residues from the viral polymerase (ΔGDD) results in a noninfectious, RNA replication incompetent derivative of poliovirus (Kempf et al., 2016). Wild-type and mutant derivatives of ΔCapsid Donor RNA were used, with the following $3D^{pol}$ substitution mutations: G64S, Y275H, G64S L420A, L420A, L420I, L420V, M392A, K375R, R476K, M299I, M323I, T353I and ΔGDD. Mutations were engineered into ΔCapsid Donor cDNA clones as previously described (Thompson and Peersen, 2004; Ward and Flanegan, 1992). L929 cells were plated in 35-mm 6-

well dishes ~24 h before transfection, with ~106 cells per well. Two micrograms of viral RNA (1 ug each of ΔCapsid Donor and $3D^{pol}$ ΔGDD Recipient) was transfected into each well in triplicate (i.e., three independent samples for every experimental condition) (TransMessenger Transfection Reagent; Qiagen). Following transfection, 2 ml of culture medium (Dulbecco modified Eagle medium containing 100 units of penicillin and 100 ug per ml of streptomycin, 10% fetal bovine serum, and 10 mM MgCl2) was added to each well, and the cells were incubated at 37°C in 5% CO2. Virus was harvested at 72 h post-transfection, recovered after three rounds of freezing and thawing, cleared of cellular debris by centrifugation at 3,000 rpm, and quantified by plaque assay. RNA recombination is evident when infectious virus is recovered from cells co-transfected with two noninfectious RNAs. L929 murine cells, which naturally lack the poliovirus receptor, prevent multiple rounds of virus amplification beyond that within co-transfected cells (Kempf et al., 2016; Lowry et al., 2014). Virus was harvested from co-transfected cells at 72 hours post-transfection, an endpoint when co-transfected cells have produced as much recombinant virus as possible. For replication controls, L929 cells were transfected with full-length infectious poliovirus RNAs containing wildtype or mutant polymerase, as indicated. Infectious poliovirus recovered from the transfected cells was titered by plaque assay in HeLa cells.

Mean titers from triplicates were plotted with standard deviation error bars. Statistical significance was determined using Tukey's test for single-step multiple comparisons from GraphPad Prism (La Jolla, CA).

A2.2.6 Biochemical characterization of purified polymerase.

Biochemical characteristics of purified polymerase were examined as previously described (Gong et al., 2009; Hobdey et al., 2010; Kortus et al., 2012; McDonald et al., 2016).

Briefly, proteins were expressed in *E. coli* and purified through metal affinity, ion exchange, and gel filtration chromatography. Initiation rates are based on the time needed to form a +1 product after mixing 5 uM polymerase with 0.5 uM "10+1–12 RNA" and 40 uM GTP in reaction buffer containing 50 mM NaCl, 4 mM $MgCl_2$, 25 mM HEPES (pH 6.5), and 2 mM Tris(2-carboxyethyl)-phosphine hydrochloride (TCEP), all at room temperature. Elongation complex stability measurements are based on diluting a 15-minute initiation reaction 10-fold into the same buffer with 300 mM NaCl and then testing the amount of elongation competent complex present at time points up to 4 hours and fitting the resulting data to a single exponential decay function. Kinetics assays were done using rapid stopped-flow fluorescent methods in 75 mM NaCl, 50 mM HEPES pH7, and $MgCl_2$ in 4 mM excess over total NTP concentration. Processive elongation rates were determined with 26-nt long template RNA bearing a 5′-fluorescein end label (Gong et al., 2009) and the single cycle data were obtained with an RNA whose fluorescence reports on translocation of a template strand 2-aminopurine base from the +2 to +1 site following incorporation of a single CMP or 2′-dCMP (McDonald et al., 2016). The Discrimination Factor is the ratio of catalytic efficiencies of CTP and dCTP incorporation reactions, i.e. (k_{pol}/K_M) CTP / (k_{pol}/K_M) dCTP.

A2.3 Results

A2.3.1 Serial passage of poliovirus in ribavirin.

We adapted the methods of Pfeiffer and Kirkegaard (Pfeiffer and Kirkegaard, 2003) to select for ribavirin resistance in poliovirus (Figure A2.2). We used serial passage in escalating concentrations of ribavirin (Figure A2.2A), beginning with several ribavirin-sensitive and ribavirin-resistant parental viruses (Figure A2.2B): ribavirin-sensitive L420A virus, ribavirin-

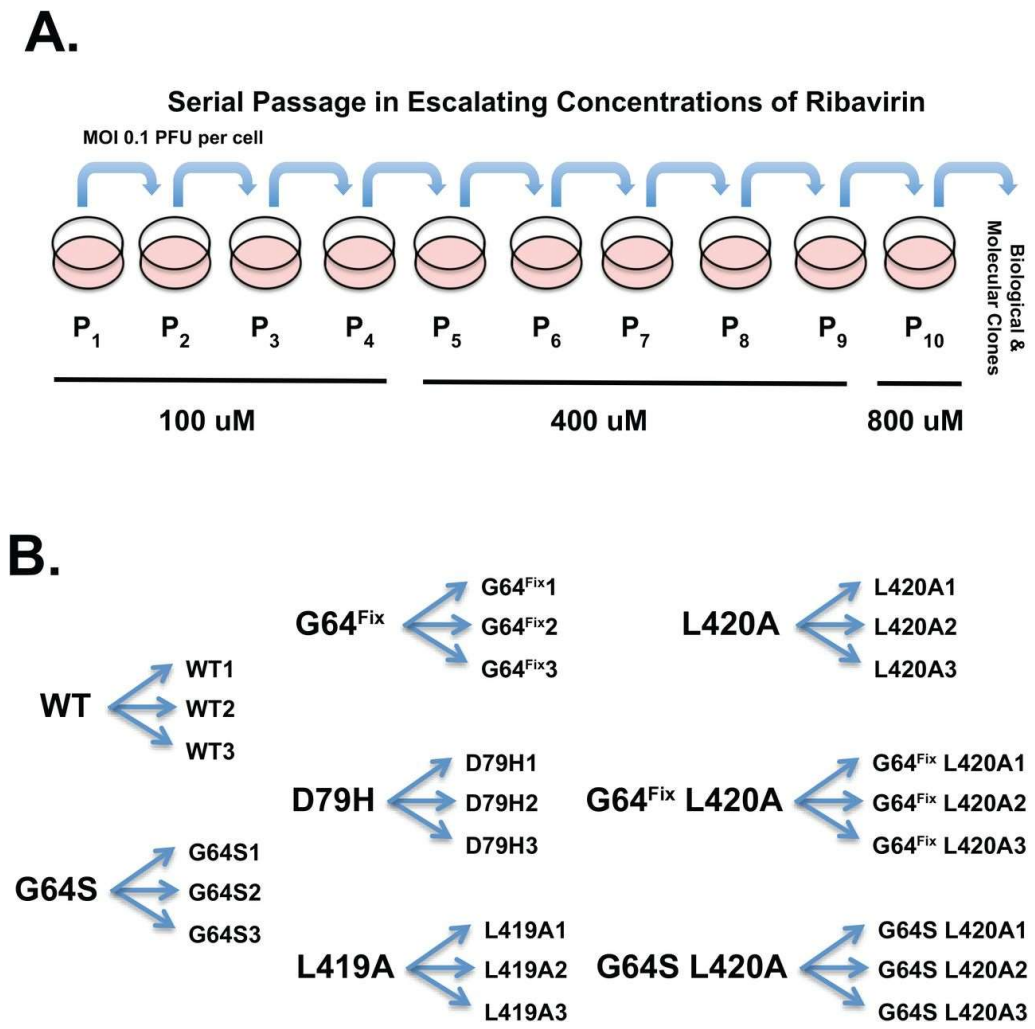

Figure A2.2: Selection of ribavirin-resistant poliovirus.

(A) Serial passage of poliovirus in escalating doses of ribavirin. Methods adapted from Pfeiffer and Kirkegaard, 2003 (Pfeiffer and Kirkegaard, 2003). (B) Diagram showing the eight parental and 24 progeny virus strains used in this study.

resistant G64S virus and virus with a G64 codon mutation ($G64^{Fix}$) designed to inhibit emergence of G64S-mediated resistance. Altogether, this study contained eight parental strains with three independent lineages per strain: wildtype (WT), G64S, $G64^{Fix}$, D79H, L419A, L420A, $G64^{Fix}$ L420A and G64S L420A. Poliovirus with these mutations have normal one-step growth phenotypes in HeLa cells and all of the mutations are stably maintained in virus populations in the absence of ribavirin (Kempf et al., 2019). A G64S mutation in the poliovirus polymerase mediates resistance to ribavirin by increasing the fidelity of RNA synthesis (Arnold et al., 2005; Pfeiffer and Kirkegaard, 2003) while an L420A mutation in the poliovirus polymerase increases sensitivity to ribavirin by inhibiting replicative RNA recombination / sexual RNA replication (Kempf et al., 2016, 2019). Virus with a D79H polymerase mutation was included due to conflicting reports: one suggesting the mutation inhibits replicative RNA recombination (Xiao et al., 2016) and another refuting these conclusions (Kempf et al., 2019). Virus with an L419A polymerase mutation was included because the L419 residue is immediately adjacent to the L420 residue, and alanine substitution mutations at these sites have been functionally characterized in biochemical and virological assays (Kempf et al., 2013, 2016, 2019). We used an MOI of 0.1 PFU per cell for each serial passage, harvesting virus by freeze thaw at 24 hpi. We used an MOI of 0.1 PFU per cell to avoid genetic bottlenecks, so that irrelevant (random) mutations that have nothing to do with ribavirin resistance would be less likely to accumulate and fix in the populations. At an MOI of 0.1 (~ 10^5 PFU per 35 mm well), advantageous (ribavirin resistance) mutations need to outcompete parental strains (due to positive selection) to become fixed in the population.

In order to maintain an MOI of 0.1 PFU per cell for each serial passage, we monitored the titer of poliovirus recovered from each infection (Figure A2.3). P_0 titers ~10^9 PFU per ml

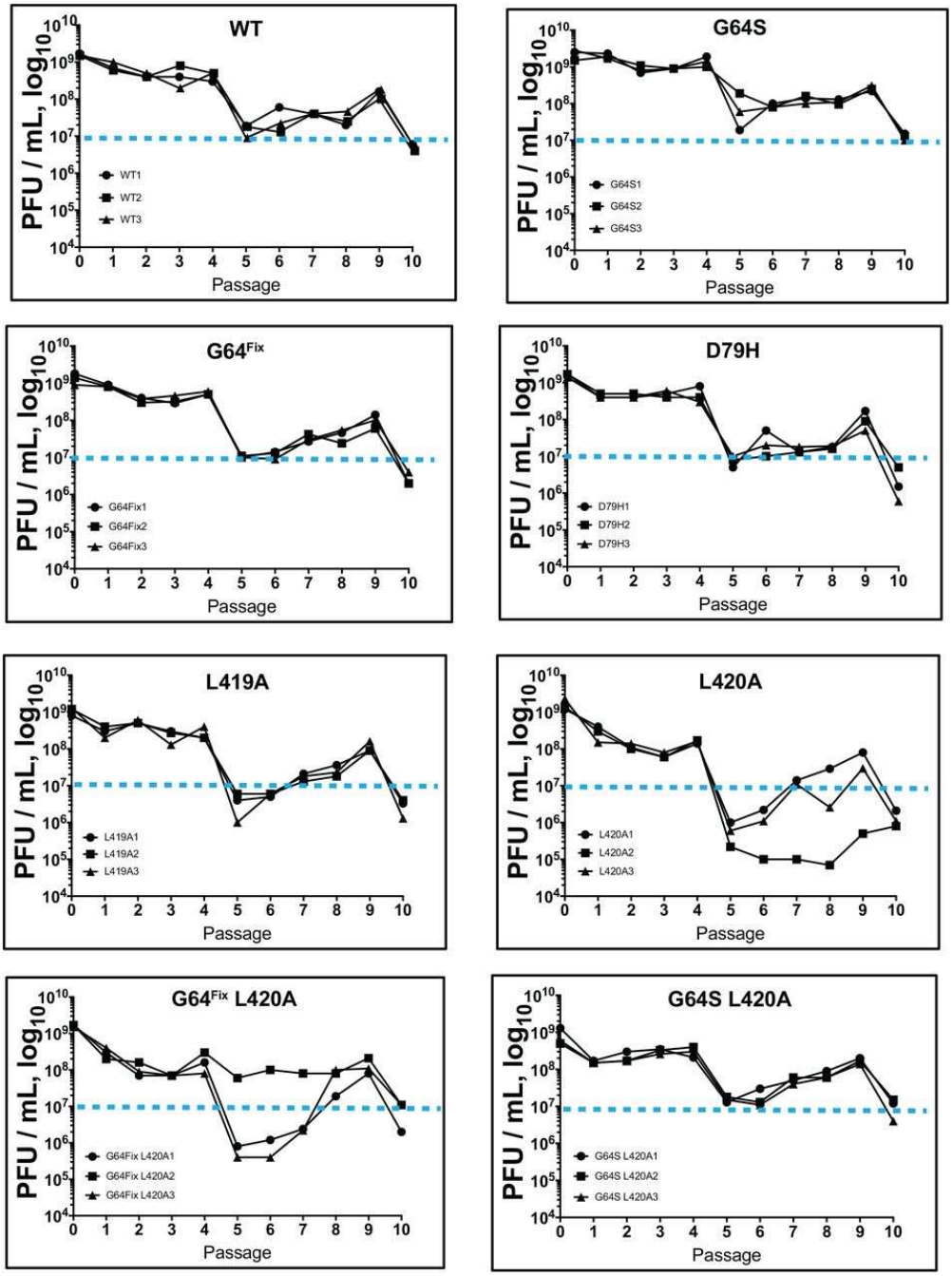

Figure A2.3: Titers of poliovirus recovered during serial passage in ribavirin.

HeLa cells were infected with WT or mutant parental strains of poliovirus at an MOI of 0.1 PFU per cell, with three independent lineages per virus, and passaged 10 times (P1-P10). Virus was incubated in 0 uM (P0), 100 uM (P1 to P4), 400 uM (P5 to P9) or 800 uM ribavirin (P10). Virus was harvested at 24 hpi and titered by plaque assay at each passage. Dashed line provides reference point for 100-fold decrease in titers as compared to no drug at P_0.

were obtained in the absence of ribavirin (Figure A2.3, P_0). Under most circumstances, the titers for each lineage of virus were similar from passage to passage, with lower titers recovered as the concentrations of ribavirin increased: titers of 10^8 to 10^9 PFU per ml for passage 1-4 in 100 uM Rb, a wider range of titers from 10^5 to 10^8 PFU per ml for passage 5-9 in 400 uM Rb and titers ~ 10^6 to 10^7 PFU per ml for passage 10 at 800 uM Rb. In some cases, the titers of poliovirus recovered for independent lineages varied considerably from one passage to another: L420A lineage 2 (L420A2) titers dropped precipitously at passage 5 (~ 10^5 PFU per ml) and remained low through passage 8 whereas L420A lineages 1 and 3 dropped at passage 5 but then increased incrementally from passage 5 through passage 9. Similar divergence was apparent in $G64^{Fix}$ L420A lineages: $G64^{Fix}$ L420A lineage 2 ($G64^{Fix}$ L420A2) titers did not drop at passage 5 whereas $G64^{Fix}$ L420A lineages 1 and 3 dropped precipitously at passage 5, with an incremental recovery by passage 9. In contrast, titers for G64S lineages 1-3 remained relatively high at all passages, as compared to other parental strains. Altogether, these data show that WT, ribavirin-resistant and ribavirin-sensitive parental strains behave differently during serial passage in escalating concentrations of ribavirin, and that individual lineages of ribavirin sensitive parental viruses become more resistant to ribavirin at different points during passage.

A2.3.2 Polymerase mutations selected by serial passage in ribavirin.

We expanded the population of poliovirus for each lineage after passage 10 and sequenced viral cDNA to identify polymerase mutations fixed in the virus populations (Table A2.1). No polymerase mutations were selected in 12 of 24 lineages, including all lineages of WT and ribavirin-resistant G64S parental viruses. These data indicate that genetic bottlenecks were avoided - no mutations were detected in expanded passage 10 ($P10^e$) populations from WT or G64S parental viruses. In contrast, we identified revertants, pseudorevertants or second site

Table A2.1: Polymerase mutations selected by serial passage in ribavirin.

* Poliovirus recovered from passage 10 (Figures A2.2 & A2.3) was expanded by infecting a T150 flask of HeLa cells. The $3D^{pol}$ region of expanded P10 virus ($P10^{e}$) was converted to cDNA and sequenced.

Parental virus phenotype	Lineage	$P10^{e}$ mutation(s)[a]
WT	1	None
	2	None
	3	None
G64S	1	None
	2	None
	3	None
G64Fix	1	None
	2	None
	3	$3D^{pol}$ M392V and $C \rightarrow U_{6163}$
D79H	1	None
	2	$3D^{pol}$ T353I
	3	None
L419A	1	None
	2	None
	3	$3D^{pol}$ G64S
L420A	1	$3D^{pol}$ L420V and M299I
	2	$3D^{pol}$ L420V and $3C^{pro}$ A172S
	3	$3D^{pol}$ L420I and M323I
$G64^{Fix}$ L420A	1	$3D^{pol}$ L420 reversion, M323I, and $C \rightarrow U_{6163}$
	2	$3D^{pol}$ L420 reversion
	3	$3D^{pol}$ L420 reversion and G64S
G64S L420A	1	$3D^{pol}$ L420 reversion
	2	$3D^{pol}$ L420 reversion
	3	$3D^{pol}$ L420 reversion

mutations fixed in expanded passage 10 ($P10^{e}$) virus populations in 9/9 lineages of ribavirin-sensitive parental viruses (Table A2.1). Ribavirin-sensitive parental viruses include L420A, $G64^{Fix}$ L420A and G64S L420A (Kempf et al., 2019). Polymerase mutations were also detected in 1/3 lineages of $G64^{Fix}$, D79H and L419A parental viruses, parental viruses with ribavirin sensitivity like WT virus (Table A2.1). The precise nature of polymerase mutations varied from one ribavirin-sensitive parental virus to another. A variety of polymerase mutations were selected from L420A and $G64^{Fix}$ L420A parental viruses (Table A2.1. Lineage 1-3 of L420A and $G64^{Fix}$ L420A viruses) while G64S L420A parental viruses had one singular outcome – reversion to L420 while maintaining the G64S resistance mutation (Table A2.1, G64S L420A Lineages 1-3). These data indicate that distinct genetic backgrounds and codon potential played a role in the nature of selected ribavirin-resistance mutations. When G64S was present alone (G64S parental virus), or within an otherwise ribavirin-sensitive parental population (G64S L420A parental virus), the only outcome was L420A reversion to WT while maintaining G64S alleles.

In contrast, when G64S was absent from the parental virus, or a codon mutation was used to inhibit G64S emergence, a variety of novel polymerase mutations were selected during serial passage in escalating concentrations of ribavirin. Thus, by using a variety of parental viruses, a variety of polymerase mutations were selected during serial passage in escalating concentrations of ribavirin. Altogether, polymerase mutations were selected in 12/24 lineages, including L420A revertants (L420), L420A pseudorevertants (L420V and L420I), G64S and several novel ribavirin resistance mutations (M299I, M323I, M392V and T353I) (Table A2.1). It was surprising to find that G64S was selected infrequently and even more surprising to see it selected from a $G64^{Fix}$ parental strain. As noted above, by passing virus at an MOI of 0.1 PFU per cell,

ribavirin-resistant mutants that arise in the population must outcompete parental strains to become fixed in the population. Under the conditions of our experiment, G64S mutations became fixed in the populations on occasion, but infrequently overall.

The polymerase mutations selected during serial passage in escalating concentrations of ribavirin were located at several sites within polymerase elongation complexes, but all were in close proximity to RNA or the active site (Figure A2.4). G64S, a well-studied ribavirin-resistance mutation (Arnold et al., 2005; Pfeiffer and Kirkegaard, 2003), is distal from the active site, near the back of the palm domain (Figure A2.4. G64S). L420A revertants and pseudorevertants are present in an alpha-helix in the thumb region of the polymerase, where they pack into the minor groove of the dsRNA product as it exits the polymerase (Figure A2.4, L420). L420, and presumably pseudorevertants thereof (L420V and L420I), interact with the ribose of viral RNA products 3 bases from the active site (Kempf et al., 2016). M392V is found between L420 and the primer grip in polymerase elongation complexes (Figure A2.4, M392V). The location of the M392V ribavirin-resistance mutation underpins several important insights from this investigation. T353I is in motif D at the back of the palm domain and near the NTP entry tunnel and G64 (Figure A2.4, T353I). M299I is on the motif B helix and buried in the palm domain, not far from M323I that is found in the motif C beta-sheet containing the active site YGDD motif (Figure A2.4).

A2.3.3 Ribavirin sensitivity of parental and progeny virus.

We compared the titers of parental and progeny virus populations grown in the presence and absence of 600 uM ribavirin (Figure A2.5). Expanded passage 10 ($P10^e$) viruses from each lineage were compared to their respective parental viruses. Titers ~ 10^9 PFU per ml were obtained for all viruses when they were grown in the absence of ribavirin (Figure A2.5). WT

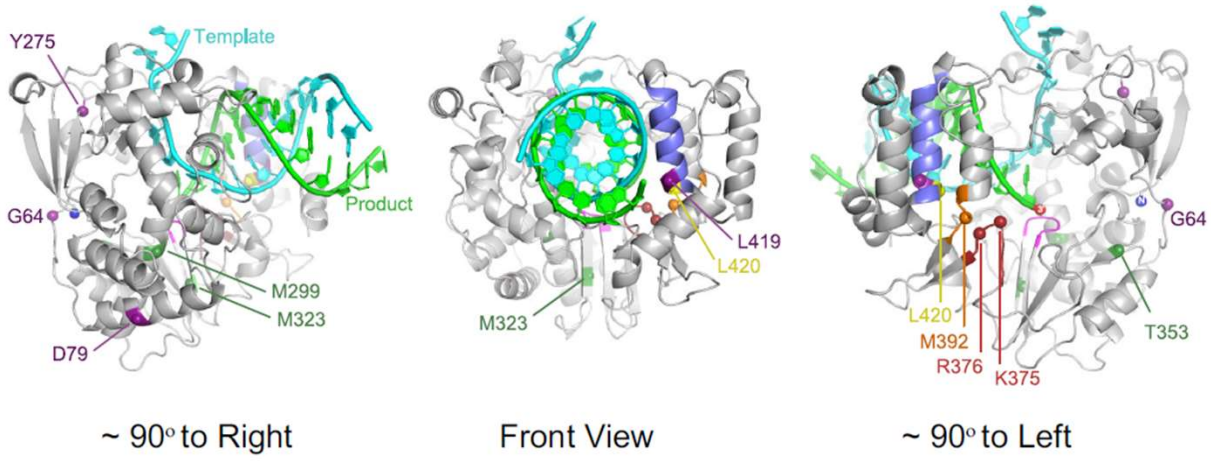

Figure A2.4: Location of polymerase mutations within elongation complexes.

The atomic structure of poliovirus polymerase elongation complexes (Gong and Peersen, 2010) was used to highlight the locations of mutations. Polymerase is shown in grey, RNA template in cyan, RNA products in green and locations of mutations in purple (G64, D79, Y275 and L419), dark green (M299, M323, T353), orange (M392) and yellow (L420). Classic primer grip residues are shown in red (K375 and R376). L419 (purple) and L420 (yellow) residues are adjacent to one another at the base of a thumb alpha-helix (blue) that interacts with the minor groove of the RNA product helix. Poliovirus polymerase PDB entry 3OL6 was rendered using PyMOL molecular graphics (Schrodinger, LLC).

poliovirus, and the $P10^e$ viruses derived from WT after serial passage in escalating concentrations of ribavirin, remained sensitive to inhibition by ribavirin, with titers below 10^7 PFU per ml when grown in 600 uM ribavirin (Figure A2.5, WT). G64S parental virus, and the $P10^e$ viruses derived therefrom, retained resistance to ribavirin, with titers well above 10^7 PFU per ml when grown in the presence of 600 uM ribavirin (Figure A2.5, G64S). Expanded passage 10 ($P10^e$) virus from other parental strains exhibited a spectrum of sensitivity to ribavirin, as compared to WT and G64S viruses (Figure A2.5). $G64^{Fix}$ parental virus was similar to WT poliovirus; however, $G64^{Fix}$ lineage 3 was more resistant to ribavirin (Figure A2.5, $G64^{Fix}$), presumably due to the selected M392V mutation (Table A2.1). D79H parental virus was similar to WT poliovirus, as reported (Xiao et al., 2016), yet D79H lineage 2 was modestly resistant to ribavirin (Figure A2.5, D79H panel), presumably due to the selected T353I mutation (Table A2.1). L419A parental virus was similar to WT poliovirus, as reported (18, 19), yet L419A lineage 3 was more resistant to ribavirin (Figure A2.5, L419A panel), no doubt due to the selected G64S mutation (Table A2.1). L420A parental virus was more sensitive to ribavirin than WT virus, as reported (Kempf et al., 2016, 2019); however, L420A lineage 1, 2 and 3 viruses grew to higher titers in the presence of ribavirin (Figure A2.5, L420A panel), presumably due to the selected polymerase mutations in each lineage (Table A2.1). $G64^{Fix}$ L420A parental virus was more sensitive to ribavirin than WT virus, as reported (Kempf et al., 2016, 2019); however, $G64^{Fix}$ L420A lineage 1, 2 and 3 viruses grew to higher titers in the presence of ribavirin (Figure A2.5, $G64^{Fix}$ L420A panel), presumably due to the selected polymerase mutations in each lineage (Table A2.1). Similarly, ribavirin-sensitive G64S L420A parental virus and its selected progeny exhibited ribavirin phenotypes consistent with the genotypes of the respective virus populations (Figure A2.5 and Table A2.1). Altogether, these data suggest that the polymerase

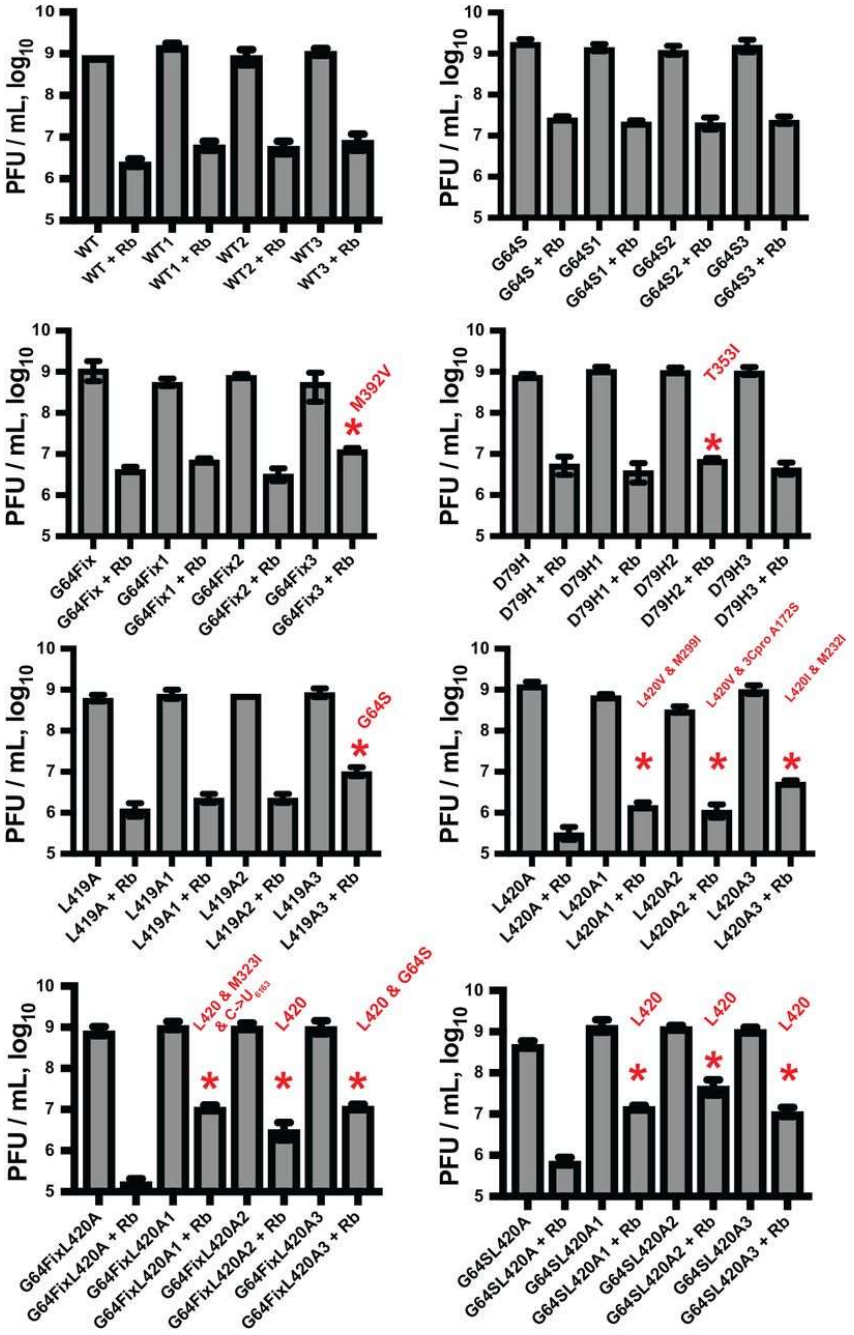

Figure A2.5: Ribavirin sensitivity of parental and progeny virus.

Titer of parental and progeny virus grown in the absence and presence of 600 uM ribavirin. Expanded passage 10 (P10^{e}) viruses from each lineage were used to infect HeLa cells at an MOI of 0.1 PFU per cell. Infected cells were incubated for 24 hrs in the absence or presence of 600 uM ribavirin. Virus was harvested by freeze-thaw and titered by plaque assay on HeLa cells. The names of parental strains and lineage 1, 2 and 3 strains are indicated on the X-axis. Polymerase mutations selected during serial passage and identified in Table 1 are annotated on the graphs accordingly (in red).

mutations selected during serial passage in escalating concentrations of ribavirin (Table A2.1) correlate with measurable resistance to ribavirin (Figure A2.5).

A2.3.4 Virus derived from infectious cDNA clones.

The polymerase mutations identified during serial passage in escalating concentrations of ribavirin were reverse engineered into infectious cDNA clones of poliovirus to derive genetically defined virus populations (Table A2.2, Ribavirin Selected Mutations). In addition, several engineered variants originating from our work, and that of others, were similarly generated (Table A2.2, Engineered Variants) (Kempf et al., 2016, 2019; Pfeiffer and Kirkegaard, 2003; Xiao et al., 2016). Many of these mutations lie within an extended primer grip region of the polymerase, consisting of L420, M392 and primer grip residues (K375 and R376). All of the mutations listed here were stably maintained in infectious virus, except for alanine substitution mutations in the primer grip residues (K375A and R376A) that reverted back to wildtype residues, precluding further evaluation. However, the charge retaining K375R and R376K mutants at these sites were stable. Virus containing polymerase mutations were assessed in one-step growth assays (Figure A2.6), ribavirin dose-response assays (Figure A2.7A), viral RNA recombination assays (Figure A2.7B) and replication controls (Figure A2.7C).

A2.3.5 One-step growth of wildtype and mutant polioviruses.

Before assessing ribavirin sensitivity and resistance, we measured one-step growth of wildtype and mutant polioviruses in HeLa cells (Figure A2.6). HeLa cells were infected at an MOI of 10 PFU per cell, and virus yields were determined at 0, 1, 2, 3, 4, 6, 8 and 24 hpi (Figure A2.6). Each of the viruses grew by 4 orders of magnitude between 3 and 24 hpi, reaching titers ~10^9 PFU per ml. While some mutant viruses had noticeably lower titers than wildtype at 6 or 8

Table A2.2: Virus derived from infectious cDNA.

* Polymerase mutations identified during ribavirin selection, along with engineered variants, were cloned into infectious cDNA. Poliovirus was recovered from HeLa cells transfected with RNA derived from each clone. cDNA from the $3D^{pol}$ region was sequenced to determine whether mutations were stably maintained in virus. Citations (Kempf, et. al., 2019, Kempf, et. al., 2016, Collis, et. al. 1992, Pfeiffer and Kirkegaard, 2003, Acevedo, et. al., 2018).

Polymerase*	Ribavirin Selected	Engineered Variant	Stably Maintained	Origin
WT			Yes	Collis et al. 1992
G64S	✓		Yes	Pfeiffer and Kirkegaard, 2003
L420A		✓	Yes	Kempf et al. 2016
G64S L420A		✓	Yes	Kempf et al. 2019
L420V	✓		Yes	This study
L420I	✓		Yes	This study
M392A		✓	Yes	This study
M392L		✓	Yes	This study
M392V	✓		Yes	This study
K375R		✓	Yes	This study
R376K		✓	Yes	This study
K375A		✓	No	This study
R376A		✓	No	This study
M299I	✓		Yes	This study
M323I	✓		Yes	This study
T353I	✓		Yes	This study
Y275H		✓	Yes	Acevedo et al. 2018

hpi, overall differences between WT and mutant viruses were not statistically significant. A Holm-Sidak comparison of each mutant to wildtype yielded the following p-values for Y275H (0.5368), M299I (0.3686), M323I (0.6709), T353I (0.3425), K375R (0.1099), R376K (0.4372), M392A (0.7341), M392L (0.4467), M392V (0.4391), L420I (0.4590) and L420V (0.9703). These Holm-Sidak comparisons include all time points, effectively comparing one virus growth curve to another. In another statistical analysis, we compared WT and mutant virus titers at individual time points using unpaired t-tests. Statistically significant differences (p-values < 0.05) were not obtained in unpaired t-tests, except for some spurious differences at 0, 1 and 2 hpi. The error bar on the WT virus titer at 8 hpi is larger than error bars at other time points because one of three WT virus titers was unusually higher at 8 hpi, making this data point shift higher, with a larger error bar (Figure A2.6). Because of this, differences in WT and mutant virus titers are noticeable at 8 hpi but are not statistically significant. These one-step growth curves do not rule out the possibility of subtle differences in replication rates between WT and mutant viruses; however, the wildtype and mutant viruses exhibited similar one-step growth phenotypes in highly permissive HeLa cells.

A2.3.6 Ribavirin sensitivity and resistance.

Poliovirus-infected HeLa cells were used to assess ribavirin sensitivity and resistance (Figure A2.7A). Wildtype and mutant polioviruses reached titers ~10^9 PFU per ml in the absence of ribavirin (Figure A2.7A, 0 uM ribavirin). Wildtype poliovirus titers decreased incrementally as the dose of ribavirin increased, from titers ~10^9 PFU per ml in the absence of ribavirin to titers ~ 2 x 10^6 PFU per ml in 1000 uM ribavirin. Consistent with other studies (Graci et al., 2012), we normalized the ribavirin dose-response data using the titers of each virus at 0 uM ribavirin (Figure A2.7A, Percent 0 uM Ribavirin Titers for y-axis). Wildtype poliovirus titers decreased by ~1000-fold as ribavirin treatment

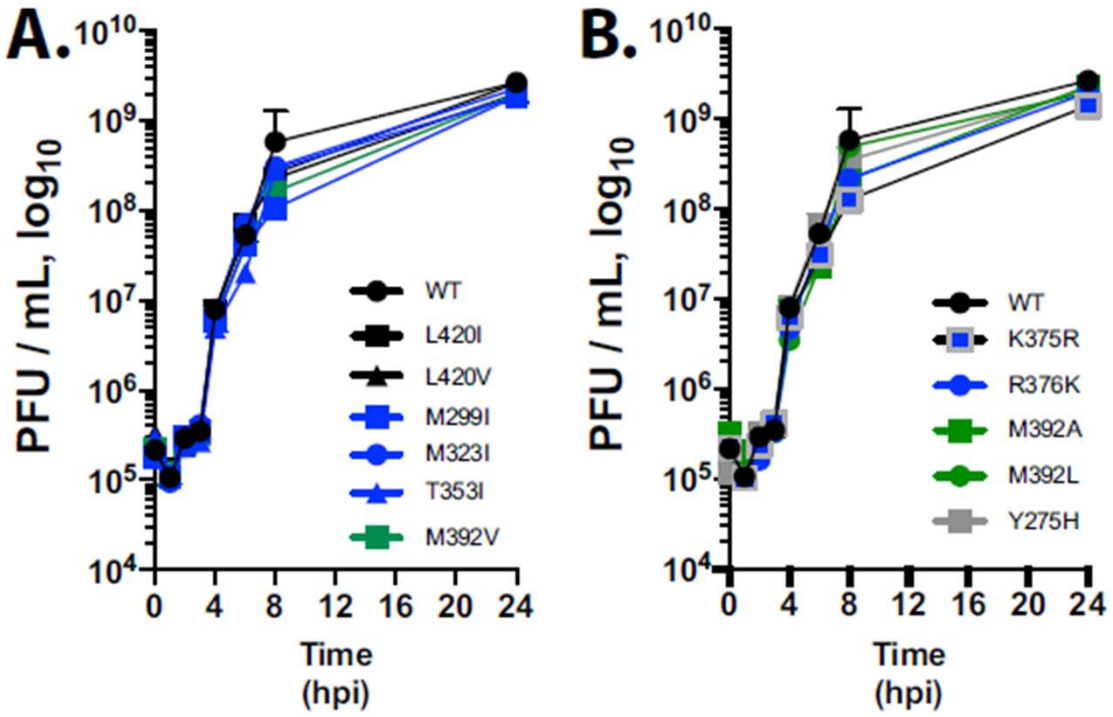

Figure A2.6: One-step growth of wildtype and mutant polioviruses.

Polymerase mutations selected by serial passage in ribavirin were engineered into infectious cDNA clones. Viruses derived from infectious cDNA clones were compared under one-step growth conditions: HeLa cells were infected with wild-type or mutant poliovirus at an MOI of 10 PFU per cell. Virus was harvested at the indicated times by freeze-thawing cells. Titers were determined by plaque assay and plotted versus time (hpi, hours post-infection). Mean titers from triplicates were plotted with standard deviation error bars, although some error bars are too small to show.

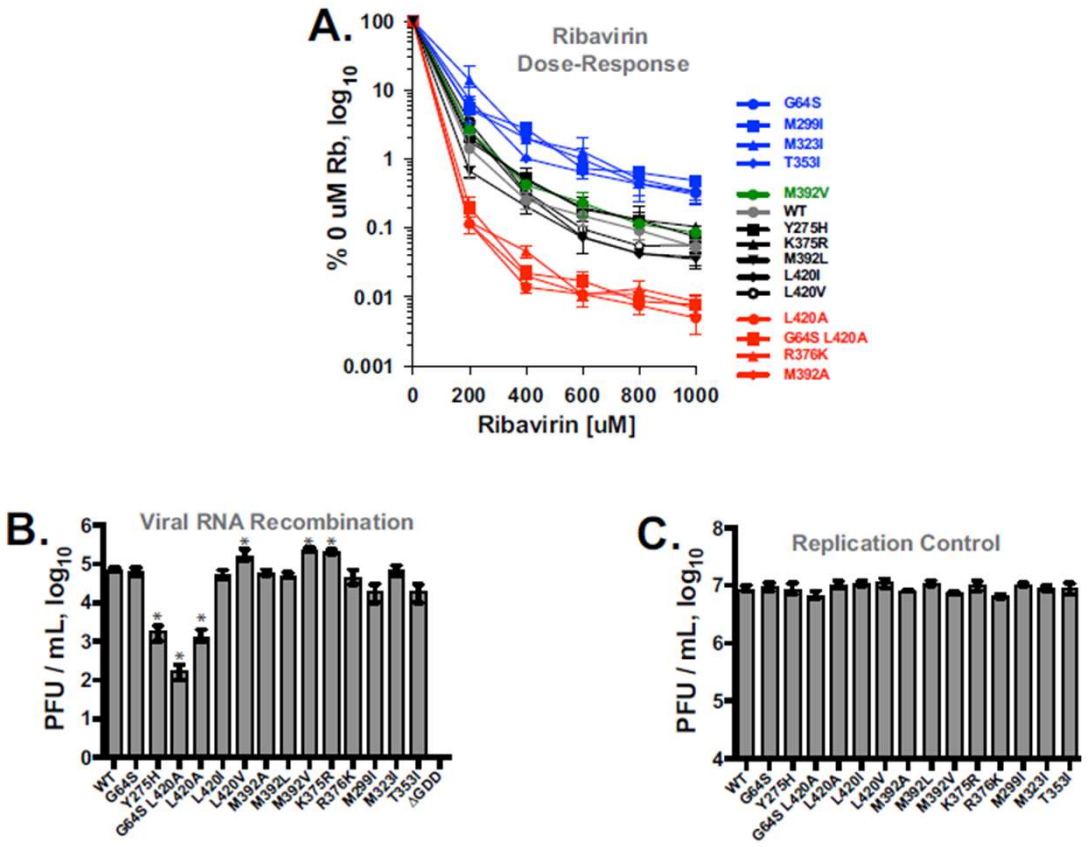

Figure A2.7: Polymerase mutations influence ribavirin-induced error catastrophe and sexual RNA replication mechanisms.

(A) Polymerase mutations influence ribavirin sensitivity and resistance. HeLa cells were infected with wild-type or mutant poliovirus at an MOI of 0.1 PFU per cell and incubated for 24 h with the indicated concentrations of ribavirin. Each condition was performed in triplicate. After three freeze-thaw cycles, viral titers were determined by plaque assay and plotted versus ribavirin concentration. Three ribavirin-responsive phenotypic clusters were observed: a ribavirin-resistant cluster (blue), a wildtype cluster (grey & black) and a ribavirin-sensitive cluster (red). M392V (green) congregated with viruses in the WT cluster. Mean titers from triplicates were plotted with standard deviation error bars, although some error bars are too small to show. Holm-Sidak p-values ranging from 0.01 (200 uM) to 0.000002 (1000 uM) when comparing viruses in the WT cluster to those in the G64S or L420A clusters. (B) Polymerase mutations influence the frequency of viral RNA recombination mechanisms, i.e. sexual RNA replication. ΔCapsid "Donor" and ΔGDD "Recipient" RNAs were co-transfected into murine cells as previously established (Kempf, et. al., 2019, Kempf, et. al., 2016). The titer of virus produced in co-transfected murine cells was determined by plaque assays in HeLa cells. Mean titers from triplicates were plotted with standard deviation error bars. *, P values are <0.05 compared to the wildtype. (C) Replication controls showing titers from poliovirus RNA containing polymerase mutations (x-axis) after transfection into murine cells. At 72 hpt, the amount of poliovirus produced within the murine cells was determined by plaque assay. Mean titers from triplicates were plotted with standard deviation error bars.

increased from 0 to 1000 uM (Figure A2.7A, WT). Poliovirus with a G64S polymerase mutation resisted inhibition by ribavirin, as previously reported (Kempf et al., 2016, 2019; Pfeiffer and Kirkegaard, 2003), with titers decreasing by ~100-fold in 1000 uM ribavirin (Figure A2.7A, G64S). In contrast, poliovirus with an L420A polymerase mutation was more sensitive to inhibition by ribavirin, with titers decreasing by ~10,000-fold in 1000 uM ribavirin (Figure A2.7A, L420A). Likewise, poliovirus containing both G64S and L420A mutations was more sensitive to ribavirin, with titers decreasing by ~10,000-fold in 1000 uM ribavirin (Figure A2.7A, G64S L420A). Thus, the ribavirin-resistant (G64S) and ribavirin-sensitive (L420A) controls behaved as previously reported (Kempf et al., 2016, 2019).

The viruses used in this investigation segregated into three ribavirin-responsive phenotypic groups (Figure A2.7A): a ribavirin-resistant cluster (blue), a wildtype cluster (black) and a ribavirin-sensitive cluster (red). Holm-Sidak p-values were significant at all ribavirin concentrations when comparing viruses in the WT cluster to those in the G64S or L420A clusters: p-values ranging from 0.01 (200 uM) to 0.000002 (1000 uM). Ribavirin-resistant polioviruses included those with G64S, M299I, M323I and T353I (Figure 7A, ribavirin-resistant cluster in blue). Although an M392V mutation was selected during serial passage in ribavirin, poliovirus with an M392V mutation was not as resistant to ribavirin as G64S: rather, it grouped with other viruses in the WT cluster (Figure A2.7A, M392V in green). Furthermore, M392V resistance to ribavirin was not statistically significant when compared to WT poliovirus: p-values of 0.262 (200 uM), 0.173 (400 uM), 0.393 (600 uM) and 0.173 (1000 uM). Ribavirin-sensitive polioviruses included those with L420A, G64S L420A, R376K and M392A polymerase mutations. Polioviruses in the wildtype cluster included WT, Y275H, K375R, M392V, M392L, L420I and L420V. These results indicate that the polymerase mutations selected during serial

passage in escalating concentrations of ribavirin (G64S, L420V, L420I, M392V, M299I, M323I, and T353I) are indeed ribavirin-resistance mutations, although one mutant (M392V) was not statistically different from WT. By comparison to the selected M392V mutation, an M392A variant was ribavirin-sensitive and an M392L variant had wildtype sensitivity to ribavirin. L420A and G64S L420A parental strains were ribavirin-sensitive viruses, whereas L420V and L420I containing viruses exhibited wildtype sensitivity to ribavirin. Conservative substitutions in the primer grip exhibited divergent ribavirin phenotypes: poliovirus containing a K375R mutation had wildtype sensitivity to ribavirin whereas poliovirus containing a R376K mutation was ribavirin-sensitive. These data indicate that polymerase mutations influence ribavirin sensitivity and resistance, with some mutations rendering virus more resistant to ribavirin while others render virus more sensitive to ribavirin.

A2.3.7 Viral RNA recombination.

Based on our previous investigation (Kempf et al., 2019), we predicted that ribavirin-resistant poliovirus might acquire polymerase mutations that facilitate sexual RNA replication mechanisms as a means to avoid ribavirin-induced error catastrophe. Consequently, we assayed the impact of polymerase mutations on the frequency of replicative RNA recombination (Figure A2.7B). In replicative RNA recombination assays, we co-transfect murine cells with two viral RNAs, each of which contains a lethal mutation: 1, a subgenomic replicon with an in-frame capsid deletion carrying the mutant polymerase, and 2, a full-length poliovirus RNA with a lethal deletion of the active site GDD motif in the polymerase (Kempf et al., 2016, 2019). Wildtype or mutant polymerase is produced from the subgenomic replicon in the co-transfected cells. The amount of infectious poliovirus produced in the co-transfected murine cells corresponds with the

frequency of viral RNA recombination. By comparing the titers of poliovirus, we assess the impact of polymerase mutations on the frequency of viral RNA recombination (Figure A2.7B).

WT polymerase established wildtype levels of replicative RNA recombination in our experiments, with titers of poliovirus ~ 10^5 PFU per ml (Figure A2.7B, WT). While Lowry et al. (Lowry et al., 2014) report that a G64S mutation inhibits viral RNA recombination, we consistently find that a G64S mutation has no impact on the frequency of viral RNA recombination under the conditions of our experiments (Figure A2.7B, G64S). Technical differences in viral RNA recombination assays between the Barton and Evans labs may be responsible for these divergent outcomes, as previously reported (Kempf et al., 2016, 2019). In contrast, an L420A mutation inhibited replicative recombination by two orders of magnitude, with poliovirus titers ~10^3 PFU per ml (Figure A2.7B, L420A). Likewise, a G64S L420A polymerase supported decreased levels of replicative RNA recombination, with titers near 10^2 PFU per ml (Figure A2.7B, G64S L420A). These outcomes for WT, G64S and L420A polymerases are similar to those previous reported (Kempf et al., 2019): WT and G64S polymerases support wildtype magnitudes of viral RNA recombination whereas an L420A polymerase mutation significantly inhibits replicative RNA recombination.

L420A pseudorevertants (L420V and L420I), selected during serial passage in escalating concentrations of ribavirin, restored replicative RNA recombination frequencies to wildtype levels (Figure A2.7B, L420I and L420V). All the other polymerase mutations selected during serial passage in escalating concentrations of ribavirin maintained replicative RNA recombination frequencies at near wildtype levels, albeit with some variation in titers both above and below 10^5 PFU per ml (Figure A2.7B, M392V, M299I, M323I, M392V and T353I).

Engineered polymerase mutations had divergent ribavirin and recombination phenotypes: M392A was ribavirin-sensitive with wildtype levels of recombination while M392L exhibited wildtype ribavirin sensitivity and wildtype levels of recombination (Figure A2.7, M392A and M392L). Likewise, K375R exhibited wildtype ribavirin sensitivity and wildtype levels of recombination whereas R376K was ribavirin sensitive with wildtype levels of recombination (Figure A2.7, K375R and R376K). Finally, consistent with data reported by Acevedo et al. (Acevedo et al., 2018), a Y275H mutation inhibited viral RNA recombination without impacting ribavirin sensitivity or resistance (Figure A2.7A and A2.7B).

As a control for the viral RNA recombination assay, we examined the impact of polymerase mutations on virus replication in murine cells (Figure A2.7C, Replication Controls). Full-length infectious RNA containing each of the polymerase mutations (Figure A2.7C, x-axis) was transfected into murine cells. Infectious poliovirus was recovered at 72 hpi by freeze-thaw and titered plaque assay (Figure A2.7C, Replication Control). These data show that the polymerase mutations do not inhibit virus replication within murine cells. Consequently, the decreased titers of virus recovered from viral RNA recombination assays of Y275H, L420A and G64S L420A polymerases are due to defects in sexual RNA replication mechanisms (i.e. viral RNA recombination) rather than defects in asexual RNA replication mechanisms.

Altogether, these results indicate that ribavirin resistant poliovirus selected during serial passage in escalating concentrations of ribavirin restored (or maintained) efficient viral RNA recombination mechanisms. Furthermore, polymerase with engineered amino acid substitution mutations at M392 and primer grip residues (K375 and R376) exhibited divergent ribavirin and recombination phenotypes, with some mutations increasing sensitivity to ribavirin without negatively impacting viral RNA recombination frequencies.

A2.3.8 Biochemical phenotypes of wildtype and mutant polymerase.

Having examined viral RNA recombination and ribavirin sensitivity, we next examined the biochemical phenotypes of wildtype and mutant polymerases (Table A2.3). We used purified polymerase to assess the effects of individual mutations on various biochemical parameters, including RNA synthesis initiation, elongation complex stability, processive elongation rate, and single nucleotide addition cycle rate. The fidelity of nucleotide addition was indirectly assayed via a nucleotide Discrimination Factor (DF) that is based on the relative catalytic efficiency of rNTP versus 2'-dNTP incorporation (Campagnola et al., 2015). Previous studies established that the speed of polymerase elongation tends to correlate with the fidelity of RNA synthesis: as the rate of elongation increases the fidelity of RNA synthesis decreases (Arnold et al., 2005; Campagnola et al., 2015; Peersen, 2017).

In a previous study (Kempf et al., 2019), we compared the biochemical phenotypes of WT, G64S, L420A and G64S L420A polymerases (Table A2.3). A G64S mutation reduced the rate of RNA elongation in both single nucleotide (23 per second for G64S versus 30 per second for WT) and processive elongation assays (12 nt/s for G64S versus 22 nt/s for WT), with a corresponding increase in the fidelity of RNA synthesis (DF of 220±40 for G64S polymerase versus 120±10 for WT). Consequently, G64S is considered a high-fidelity polymerase (Arnold et al., 2005; Pfeiffer and Kirkegaard, 2003). An L420A mutation decreased elongation complex stability, increased RNA elongation rates, and exhibited a wildtype discrimination factor of 100±10. Thus, L420A polymerase is considered to have normal (wildtype) fidelity. Polymerase containing both G64S and L420A mutations has hybrid phenotypes, with faster elongation rates like L420A polymerase and an increased discrimination factor of 220±50. Consequently, G64S L420A is considered a high-fidelity polymerase, like G64S polymerase. Consistent with our

Table A2.3: Biochemical phenotypes of purified polymerase.

*Polymerase was expressed and purified. Biochemical phenotypes were determined as previously reported (Hobdey, et. al., 2010, Kortus, et. al., 2012).

**Kempf et al. (2019) (Kempf, et. al., 2019), with permission.

Polymerase*	Initiation (min)	EC Stability (Min)	Processive Rate (nt/sec)	Processive K_M (μM)	Single NTP Rate (per sec)	Single NTP K_M (μM)	Discrimination Factor
From Kempf et al. 2019**							
WT	6 ± 1	88 ± 9	22 ± 1	67 ± 2	30 ± 1	37 ± 2	120 ± 10
G64S	8 ± 1	24 ± 7	12 ± 1	46 ± 4	23 ± 1	39 ± 5	220 ± 40
L420A	8 ± 1	28 ± 4	30 ± 1	75 ± 3	40 ± 1	39 ± 3	100 ± 10
G64S L420A	6 ± 3	40 ± 8	26 ± 1	84 ± 9	28 ± 1	36 ± 6	220 ± 50
From this study							
WT	4 ± 1	130 ± 20	26 ± 1	72 ± 4	40 ± 1	46 ± 3	100 ± 10
L420V	4 ± 1	170 ± 10	29 ± 1	82 ± 5	41 ± 1	40 ± 3	99 ± 9
L420I	4 ± 1	60 ± 2	33 ± 1	90 ± 5	45 ± 1	41 ± 2	107 ± 8
M392A	3 ± 1	20 ± 1	30 ± 1	90 ± 8	33 ± 1	38 ± 3	120 ± 10
M392L	4 ± 1	39 ± 4	25 ± 1	74 ± 9	38 ± 1	34 ± 1	102 ± 6
M392V	4 ± 1	11 ± 1	22 ± 1	74 ± 7	42 ± 1	51 ± 3	160 ± 10
K375R	5 ± 1	51 ± 2	20 ± 1	96 ± 8	36 ± 1	44 ± 3	200 ± 20
R376K	2 ± 1	30 ± 3	65 ± 2	160 ± 10	48 ± 1	46 ± 3	17 ± 1
M299I	4 ± 1	180 ± 30	18 ± 1	66 ± 4	35 ± 1	42 ± 2	120 ± 10
M323I	4 ± 1	150 ± 20	18 ± 1	64 ± 4	37 ± 1	50 ± 3	130 ± 10
T353I	5 ± 1	90 ± 20	18 ± 1	73 ± 2	38 ± 1	49 ± 3	160 ± 10
Y275H	13 ± 6	95 ± 3	27 ± 1	95 ± 8	35 ± 1	42 ± 2	140 ± 10

previous report (Kempf et al., 2019), an L420A mutation inhibits viral RNA recombination (Figure A2.7B, L420A) and renders virus more susceptible to ribavirin-induced error catastrophe (Figure A2.7A, L420). Furthermore, the high-fidelity G64S L420A polymerase fails to render virus resistant to ribavirin when viral RNA recombination is substantially reduced (Figure A2.7, G64S L420A). These observations for WT, G64S, L420A and G64S L420A polymerases provide context to interpret the biochemical data for the new polymerase mutants identified in this study (Table A2.3).

As compared to WT polymerase, the L420V, M299I and T353I polymerases form more stable elongation complexes (Table A2.3). In contrast, mutations in the extended primer grip of the polymerase significantly reduced elongation complex stability (Table A2.3, L420A, L420I, M392A, M392L, M392V, K375R and R376K). The M392V mutation reduced elongation complex stability by 10-fold to 11±1 min versus 130±20 min for WT.

Primer grip mutations (K375R and R376K) exhibited divergent biochemical phenotypes (Table A2.3). The K375R mutation reduced the rate of RNA elongation in both single nucleotide (36 per sec versus 40 per sec for WT) and processive elongation assays (20 nt per sec versus 26 nt per sec for WT), with a corresponding increase in nucleotide discrimination (DF of 200±20). In contrast, the R376K mutation increased the rate of RNA elongation in both single nucleotide (48 per sec) and processive elongation assays (65 nt per sec), with a dramatic decrease in the fidelity of RNA synthesis (DF of 17±1 for R376K polymerase). The low fidelity R376K polymerase rendered poliovirus more susceptible to ribavirin (Figure A2.7A, R376K), as one might expect, however, the high fidelity K375R polymerase exhibited normal (wildtype) sensitivity to ribavirin (Figure A2.7A, K375R). The R376K polymerase also initiated RNA synthesis 2-fold faster than WT polymerase. Together, these data show that conservative

substitution mutations in the primer grip (K375R and R376K) result in divergent biochemical phenotypes (Table A2.3) and divergent ribavirin sensitivity (Figure A2.7A).

Polymerase mutations selected during serial passage in ribavirin tend to have reduced rates of RNA elongation and increased discrimination factors (Table A2.3, G64S, M392V, M299I, M323I and T353I). The M299I, M323I and T353I mutations reduced the rate of RNA elongation in processive elongation assays (18 nt per sec for M299I, M323I and T353I versus 26 nt per sec for WT), with corresponding increases in nucleotide discrimination (Discrimination Factors of 120±10 for M299I polymerase, 130±10 for M323I polymerase and 160±10 for T353I polymerase). The M392V mutation reduced the rate of RNA elongation in processive elongation assays (22 nt per sec for M392V versus 26 nt per sec for WT) and increased the fidelity of RNA synthesis (Discrimination Factor of 160±10 for M392V polymerase). Increases in the fidelity of RNA synthesis are consistent with increased resistance to ribavirin (Figure A2.7A, G64S, M392V, M299I, M323I and T353I). Altogether, these data indicate that the polymerase mutations selected during serial passage in escalating concentrations of ribavirin tend to increase the fidelity of RNA synthesis and decrease rates of RNA elongation.

We included a Y275H mutation in this study based on the report from Acevedo et al. (2018) (Acevedo et al., 2018). Consistent with their data, we find that the Y275H mutation inhibits viral RNA recombination (Figure A2.7B) without increasing sensitivity to ribavirin (Figure A2.7A). These phenotypes are distinct from those associated with an L420A polymerase mutation, where decreased viral RNA recombination is mechanistically linked with increased sensitivity to ribavirin (Figure A2.7A and A2.7B, L420A). The biochemical phenotypes of Y275H are important in this regard (Table A2.3). The Y275H mutation inhibits the initiation of RNA synthesis by 3-fold (13±6 min for Y275H versus 4±1 min for WT). None of the other

mutations in our panel exhibit this phenotype. The proximity of Y275 to the template entry channel of the polymerase suggests this defect in the initiation of RNA synthesis may arise from defects in template RNA binding (Kortus et al., 2012; Shi et al., 2020). In the discussion, we elaborate on the structural and functional distinctions of Y275 and L420 residues in the polymerase, especially as they relate to viral RNA recombination and ribavirin-induced error catastrophe.

A2.4 Discussion

In this study, we investigated the interplay between sexual RNA replication mechanisms and ribavirin-induced error catastrophe. Sexual RNA replication involves two parental templates, wherein mutant-free genomes can be re-assembled from viral RNAs containing mutations (Figure A2.1). An L420A polymerase mutation inhibits sexual RNA replication by ~50-fold (Figure A2.7B) coincident with increased susceptibility to ribavirin-induced error catastrophe (Figure A2.7A). During serial passage in ribavirin, L420A revertants and pseudorevertants (L420I and L420V) regained efficient viral RNA recombination coincident with ribavirin resistance (Figure A2.7). These data reinforce previous studies suggesting that viral RNA recombination counteracts error catastrophe (Kempf et al., 2019). We conclude that sexual RNA replication mechanisms counteract ribavirin-induced error catastrophe by purging ribavirin-induced mutations from viral RNA genomes (Figure A2.1).

The structural orientation of L420 interacting with primer-template duplexes provides insights into its role in sexual RNA replication mechanisms (Figure A2.8). Within this region of the structure, residues K375 and R376 constitute the classic primer grip electrostatic interaction that brackets the phosphodiester bonds of the nascent RNA product one base from the active site. L420 is located on the thumb domain and expands these RNA interactions via a hydrophobic

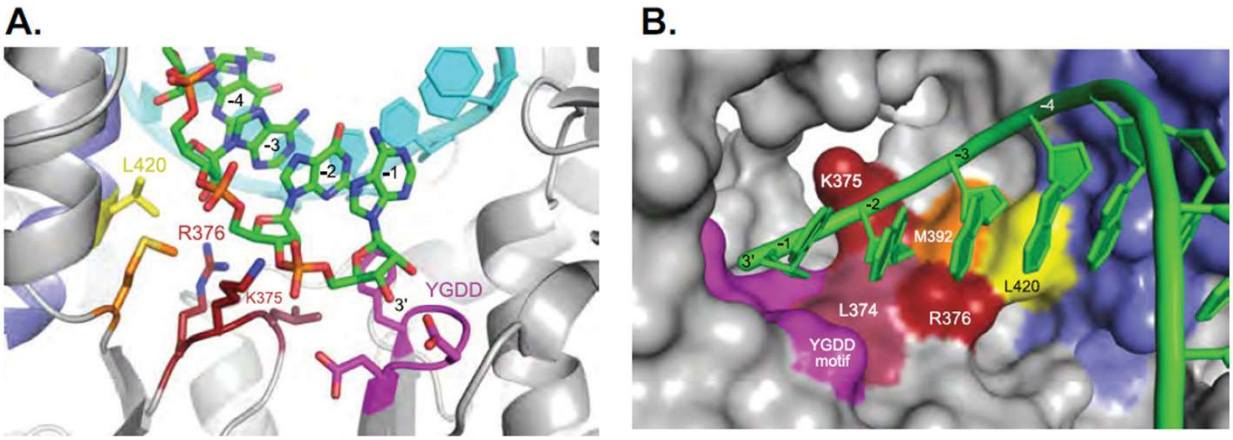

Figure A2.8: An extended primer grip in the viral polymerase mediates sexual RNA replication mechanisms.

Poliovirus polymerase L420 (yellow), M392 (orange) and classic primer grip residues K375 & R376 (red) constitute an "extended primer grip". Altogether, these residues work coordinately to hold nascent RNA products on homologous RNA templates near the catalytic site of the polymerase. (A) Back view of extended primer grip with the active site YGDD residues (fuchsia), template (cyan), and product (green) RNAs. (B) Surface representation showing how extended primer grip residues form a continuous surface in direct contact with the 3' terminal nucleotides of the RNA primer strand (in green). The template strand was omitted for clarity. Poliovirus polymerase PDB entry 3OL6 rendered using the PyMOL molecular graphics system (Schrodinger, LLC).

contact with the product strand ribose group three bases from the active site. M392, which mutated during our serial passage experiments, is found immediately between L420 and the classic primer grip. Altogether, residues L420, M392, and K375/R376 form an extended surface for direct interactions with the RNA product strand at the third, second and first bases from the active site, respectively (Figure A2.8). We refer to this structural element as an "extended primer grip". These protein-RNA interactions facilitate both asexual and sexual RNA replication mechanisms by modulating the orientation and dynamics of the primer (i.e. product) RNA strand in the active site. Notably, the L420A mutation specifically disrupts sexual RNA replication mechanisms without inhibiting asexual RNA replication mechanisms (Figure A2.1).

By using several ribavirin-sensitive and ribavirin-resistant parental viruses, serial passage in escalating doses of ribavirin led to the selection of both known and novel ribavirin resistance mutations (Table A2.1). These data reinforce and build upon prior studies regarding ribavirin-induced error catastrophe (Crotty et al., 2000; Eigen, 2002; Kempf et al., 2019; Pfeiffer and Kirkegaard, 2003). A G64S mutation was selected in two instances, reinforcing its well-established role as a modulator of RNA synthesis fidelity (Arnold et al., 2005; Pfeiffer and Kirkegaard, 2003). In addition, three novel ribavirin-resistance mutations were identified: M299I, M323I and T353I. Two important phenotypes were shared by all polymerases containing these ribavirin-resistant mutations: efficient viral RNA recombination and increased fidelity of RNA synthesis. Thus, both asexual (fidelity and nucleotide discrimination) and sexual (viral RNA recombination) RNA replication mechanisms influence ribavirin sensitivity and resistance.

A2.4.1 L420A revertants and pseudorevertants.

Because an L420A polymerase mutation exacerbates ribavirin-induced error catastrophe coincident with defects in sexual RNA replication mechanisms (Kempf et al., 2019), we

predicted that ribavirin-resistant poliovirus selected from L420A parental strains would acquire polymerase mutations that facilitate sexual RNA replication mechanisms. Consistent with this prediction, L420A revertants or pseudorevertants (L420I and L420V) were selected in every lineage of virus containing an L420A mutation: 3/3 lineages of L420A parental virus, 3/3 lineages of $G64^{Fix}$L420A parental virus and 3/3 lineages of G64S L420A parental virus (Table A2.1). Importantly, L420A revertants and the L420I and L420V pseudorevertants restored efficient sexual RNA replication mechanisms coincident with ribavirin resistance (Figure A2.7). These data reinforce the correlation between efficient viral RNA recombination and resistance to ribavirin-induced error catastrophe (Kempf et al., 2019).

A2.4.2 Novel ribavirin-resistance mutations.

Serial passage in escalating concentrations of ribavirin leads to the selection of ribavirin resistant polymerase mutations in both poliovirus and foot-and-mouth-disease virus (FMDV): a G64S mutation in the poliovirus polymerase (Pfeiffer and Kirkegaard, 2003) and an M296I mutation in the FMDV polymerase (Arias et al., 2008; Sierra et al., 2007). By using several ribavirin-sensitive and ribavirin-resistant parental viruses, we obtained both known and novel ribavirin resistant polymerase mutations following serial passage of poliovirus in escalating concentrations of ribavirin (Figure A2.2 and Table A2.1). G64S mutations were selected twice, M323I was selected twice, and M299I, T353I and M392V were each selected once. The M299I mutation in poliovirus is distinct from the ribavirin resistant M296I mutation selected in FMDV; the poliovirus M299 residue corresponds to FMDV residue I309 (Arias et al., 2008; Sierra et al., 2007). Our biochemical data indicate that polymerase mutations selected during serial passage in ribavirin tend to have reduced rates of RNA elongation and increased discrimination factors (Table A2.3). M299I, M323I and T353I mutations reduced the rate of RNA elongation in

processive elongation assays from 26 to ≈18 nt per sec, with slight increases in the fidelity of RNA synthesis with DF values of 120, 130, and 160 for M299I, M323I, and T353I polymerases, respectively.

The M392V polymerase mutation selected during serial passage in ribavirin was not statistically resistant to ribavirin in poliovirus re-derived from an infectious cDNA clone, although it was perhaps trending toward slight resistance (green in Figure A2.7A). This mutation increased the polymerase Discrimination Factor (Table A2.3) without changing the frequency of viral RNA recombination (Figure A2.7B), consistent with an impact on the fidelity of viral RNA synthesis. Curiously, a mutation at an analogous site in the EV-71 polymerase (M393L) is reported to resist the antiviral drug NITD008 (Deng et al., 2014). NITD008 is an adenine analogue whereas ribavirin is a general purine analogue. Both NITD008 and ribavirin are pro-drugs converted into triphosphate forms *in vivo* where they function as NTP substrates in the active site of the polymerase. Deng and colleagues note that the EV-71 M393 side chain interacts with the primer grip of the polymerase (Deng et al., 2014). They suspect that the M393L mutation influences the polymerase active site in a way that inhibits NITD008 antiviral activity. We find that a M392V mutation decreased the rate of RNA elongation coincident with an increase in the poliovirus polymerase Discrimination Factor (Table A2.3) without changing the frequency of viral RNA recombination (Figure A2.7B). These data suggest that the M392V mutation can influence the rate of catalysis in the active site of the polymerase, consistent with the conclusions of Deng and colleagues (Deng et al., 2014).

Two important phenotypes were shared by all of the ribavirin-resistant polymerase mutations in poliovirus: Efficient viral RNA recombination (Figure A2.7) and increased fidelity of RNA synthesis (Table A2.3). A G64S mutation renders poliovirus resistant to ribavirin-

induced error catastrophe by increasing the fidelity of RNA synthesis (Arnold et al., 2005; Pfeiffer and Kirkegaard, 2003), but a G64S mutation is insufficient for ribavirin resistance when viral RNA recombination is disabled by an L420A mutation (Kempf et al., 2019). Consequently, poliovirus with both G64S and L420A polymerase mutations is highly sensitive to inhibition by ribavirin (Figure A2.7, G64S L420A). Furthermore, when subjected to serial passage in ribavirin, the L420A mutation virus reverted in 3/3 lineages (Table A2.1), resulting in the classic ribavirin-resistant G64S poliovirus. While polymerase speed, fidelity and efficiency of recombination are intricately intertwined (Kim et al., 2019; Li et al., 2019), our data suggest that polymerase residue L420 is required for efficient viral RNA recombination and for counteracting ribavirin-induced error catastrophe (Kempf et al., 2019). Furthermore, even though polymerase speed and fidelity are intricately intertwined, recent studies suggest that polymerase speed is more important than fidelity for virulence (Fitzsimmons et al., 2018). Thus, it is possible to attribute specific biochemical functions of the polymerase to important biological outcomes.

Remarkably, despite substantial selective pressure, neither WT parental strains nor G64S parental strains acquired any new polymerase mutations during serial passage in escalating concentrations of ribavirin (Figure A2.2 and Table A2.1). We used an MOI of 0.1 PFU per cell for each serial passage in ribavirin to avoid genetic bottlenecks. Under these conditions, ribavirin resistant variants only become fixed in the population if they outcompete the parental strain from which they arise. Serial passage of poliovirus at lower MOIs might increase the selective pressure of ribavirin, as decreased virus populations are more sensitive to inhibition by ribavirin (Graci et al., 2012). Because sexual RNA replication via viral RNA recombination requires two parental templates, decreased populations of virus should reduce the frequency of viral RNA recombination, effectively mimicking the effects of an L420A mutation.

Population size and genetic diversity among homologous partners in recombination are important to effectively overcome ribavirin-induced error catastrophe. It is intuitive that small populations (which are inherently less diverse) are likely to be more susceptible to ribavirin-induced error catastrophe. When the population size and genetic diversity decrease coordinately, it becomes less probable that fit recombinants can be re-assembled from a limited and genetically debilitated (mutated) population of RNA templates. Graci et al. (Graci et al., 2012) showed that coxsackievirus B3 was more susceptible to ribavirin at a low MOI vs. a high MOI, implicating population size as an important variable. Larger and more diverse populations of otherwise homologous virus within cells make it more probable that RNA recombination can re-assemble fit genomes from ribavirin-mutated genomes.

A2.4.3 An extended primer grip in the viral polymerase.

By using a codon mutation to inhibit the rapid emergence of the G64S classic mutation, we identified M392V as a novel ribavirin resistance mutation, albeit statistically insignificant resistance to ribavirin. The M392V mutation provided statistically insignificant ribavirin-resistance as compared to the magnitudes of ribavirin resistance found in the G64S virus cluster (Figure A2.7A, G64S, M299I, M323I and T353I). Structurally, M392 is found between L420 and the primer grip and L420, M392, and K375/R376 interact directly with viral RNA product strand at the third, second and first bases from the active site, respectively (Figure A2.8). These protein-RNA interactions at the core of the polymerase facilitate both asexual and sexual RNA replication mechanisms. A L420A mutation specifically disrupts sexual RNA replication mechanisms without inhibiting asexual RNA replication mechanisms (Figure A2.1) (Kempf et al., 2016, 2019) while K375A and R376A mutations impaired asexual RNA replication mechanisms to a sufficient degree that both mutations were unstable (Table A2.2). These results

are consistent with the lethal effects of clustered charge to alanine substitution mutations at these locations in the polymerase (Diamond and Kirkegaard, 1994). More conservative substitution mutations, K375R and R376K, were stably maintained in virus and neither of these mutations disabled viral RNA recombination, but they had divergent effects on ribavirin sensitivity, which the R376K mutation increased whereas the K375R mutation had little impact.

Due to the interesting location of the M392V mutation within the elongation complex, we engineered two additional mutations, M392A and M392L, at this residue. Both were stably maintained in virus, but divergent phenotypic effects were observed. Virus with the M392A mutation was highly susceptible to inhibition by ribavirin, virus with the M392L mutation had no impact on ribavirin sensitivity, and virus with the M392V mutation had statistically insignificant resistance to ribavirin. None of these M392 mutations had significant impacts on the magnitudes of viral RNA recombination. However, biochemical data provided key information: the M392V mutation decreased elongation rates coincident with increased nucleotide discrimination (Table A2.3). These data suggest that the M392V mutation impacts asexual RNA replication mechanisms by slowing the rates of elongation coincident with very slight increases in the fidelity of RNA synthesis.

Together, the data from the mutations at L420, M392, and the primer grip suggest that these residues form an extended primer-grip surface on the polymerase that senses the complementarity of the product duplex in the immediate vicinity of the active site (Figure A2.8). Recombination involves resumption of elongation with a new primer-template pairing in the active site, and this extended primer grip would energetically favor a properly structured dsRNA helix in priming region, and thus facilitate recombination by selecting for proper base pairing at the critical step when elongation resumes on the new template. Consistent with this, the

recombination mutations generally show reduced elongation complex stability that is indicative of impaired polymerase-RNA contacts.

A2.4.4 Reconciling distinct phenotypes of Y275H and L420A mutations.

A L420A mutation renders poliovirus susceptible to ribavirin-induced error catastrophe coincident with defects in viral RNA recombination (Kempf et al., 2016, 2019). Consequently, we conclude that viral RNA recombination counteracts ribavirin-induced error catastrophe (Kempf et al., 2019). In this study, we find that L420A revertants and pseudorevertants (L420I and L420V) restored efficient sexual RNA replication mechanisms coincident with increased resistance to ribavirin, as compared to L420A parental strains (Figure A2.7). These data reinforce the correlation between efficient viral RNA recombination and resistance to ribavirin-induced error catastrophe (Kempf et al., 2019). However, data from a Y275H mutation seem to be inconsistent with the conclusion that viral RNA recombination counteracts ribavirin-induced error catastrophe (Figure A2.7, Y275H). A Y275H mutation inhibited viral RNA recombination without rendering poliovirus more susceptible to ribavirin (Figure A2.7A and A2.7B). These results are similar to those reported by Acevedo and colleagues (Acevedo et al., 2018). Importantly, the Y275 and L420 residues are structurally and functionally distinct, providing insights to help reconcile their alternate effects on ribavirin-induced error catastrophe. The Y275 residue is located in the fingers domain and the Y275H mutation is 4-fold slower for template binding and initiation (Table A2.3), suggesting a role in template binding, and the Y275H elongation complex could not be purified in sufficient yield for crystallization studies. In addition, a recent enterovirus 71 elongation complex structure shows a downstream template strand base inserted into a surface pocket with Y275 at its bottom (Y276 in EV71 polymerase) (Shi et al., 2020). Importantly, this RNA binding role for Y275 is mechanistically upstream of

L420's role in discriminating between homologous and non-homologous base pairing of the primer-template duplex near the active site. Thus, Y275 helps recruit RNA templates in a presumably sequence independent manner whereas L420 enforces sequence-dependent steps of homologous recombination. Further work will be required to better understand these two distinct steps of viral RNA recombination, and their divergent effects on ribavirin sensitivity.

A2.4.5 Picornavirus species groups and sexual RNA replication mechanisms.

The ancient origin of picornavirus RNA-dependent RNA polymerase suggests that viral RNA recombination has been a characteristic feature of viruses for billions of years (Koonin et al., 2008; Krupovic et al., 2019; Wolf et al., 2018). Thus, both asexual and sexual RNA replication mechanisms likely arose coordinately very early during virus evolution. Genetic exchange is a common event among modern picornaviruses and inevitable when two related viruses with shared RNA sequence homology co-infect cells at mucosal surfaces in the respiratory or gastrointestinal tracts (Van Dung et al., 2014; Martin et al., 2013), a process facilitated by bacteria (Erickson et al., 2018). The extended primer grip of the poliovirus polymerase highlighted by the work presented herein is conserved across picornavirus species groups and likely mediates genetic exchange between viruses in nature. Intraspecies recombination is well documented for enterovirus species A (McWilliam Leitch et al., 2012; Oberste et al., 2004b), species B (McWilliam Leitch et al., 2010; Oberste et al., 2004a), species C (Brown et al., 2003) and species D viruses (Tan et al., 2016; Yip et al., 2017), as well as rhinovirus species groups (McIntyre et al., 2010, 2013), parechoviruses (Benschop et al., 2008) and non-human enteroviruses (Van Dung et al., 2014). cVDPVs arise, in part, due to improved fitness from genetic exchange with non-polio group C enteroviruses in the field (Burns et al., 2013; Combelas et al., 2011; Famulare et al., 2016; Jegouic et al., 2009; Korotkova et al., 2017).

Now that we appreciate the molecular and ecological conditions that contribute to genetic exchange between picornaviruses, we may be able to exploit recombination to combat human disease. Recombination-deficient oral poliovirus vaccines are one option to consider (Van Damme et al., 2019); however, new approaches to exploit viral RNA recombination deserve consideration.

A2.5 Summary

An extended primer grip in picornavirus polymerases contributes to the fidelity of RNA synthesis and to efficient sexual RNA replication mechanisms. This region of the polymerase interacts with nascent RNA products near the active site, effectively sensing the degree of RNA sequence complementarity between RNA products and RNA templates. With the potential to discriminate between homologous and non-homologous RNA templates, this region of the polymerase mediates sexual RNA replication mechanisms, thereby counteracting error catastrophe and contributing to genetic exchange between related picornaviruses.