Flexibility and distributive synthesis regulate RNA priming and handoff in human DNA polymerase α-primase

By

John James Cordoba

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

December 16, 2023

Nashville, Tennessee

Approved:

Walter J. Chazin, Ph.D.

Ivelin Georgiev, Ph.D.

Brandt Eichman, Ph.D.

Katherine Friedman, Ph.D.

Yi Ren, Ph.D.

Acknowledgements

I would first like to acknowledge my advisor, Dr. Walter Chazin, who has supported and encouraged me even when I felt unencourageable. His enthusiasm for science, constant perspective on what's most important, and supportive and gentle mentoring style provided a constant sense of security and counterbalance throughout any of the doldrums I encountered. I'd also like to acknowledge his long and storied career (not to mention his pinpoint accuracy) on the basketball court and for nurturing a pickup basketball league that was a constant outlet for me. Walter has always encouraged a steady and balanced approach to life and work, values that appreciate and that have made it an easy fit to work as part of his research group.

I'd like to thank my committee members, Drs. Ivelin Georgiev, Brandt Eichman, Kathy Friedman, and Yi Ren for providing patient and consistent guidance throughout my candidacy. I appreciate their insight and support.

I'd like to thank the researchers at the SIBYLS beamline, particularly Dr. Michal Hammel, Dr. Greg Hura, and Kathryn Burnett and the members of the Vanderbilt cryoEM facility, Dr. Scott Collier, Melissa Chambers, and Mariam Hader for providing a wealth of support, for expertly maintaining and expanding the facilities that made this work possible, and for providing guidance during the troubleshooting of the many problems I encountered.

I'd also like to thank current and past members of the Chazin lab who provided guidance and camaraderie throughout my time at Vanderbilt. Drs. Alex Blee, Swati Balakrishnan, and Areetha D'Souza were phenomenal office mates who I'm glad to have shared stories, experiences, advice, and a very successful miniature plant nursery with. I'd like to also thank my mentors and friends in the pol-prim squad Drs. Elwood Mullins, Lauren Salay, and Marilyn Holt for providing guidance without which this project would not have been possible. Lauren was instrumental in helping me get started on this project and was a constant source of infectious enthusiasm that almost got me to like pol-prim. Elwood was a source of mentorship, endlessly confident and competent in the lab, and without his blazing the trail on electron microscopy of pol-prim, I might have been working on this project for many more years. Marilyn bequeathed the SAXS of pol-prim project to me right when I joined the lab. In spite of this, she has become one of my dearest friends in the world. Her inclusiveness, earnestness, excitement, and tendency

to over-imbibe have been a much-needed source of entertainment but also of community. I especially thank her for listening to me rant about pol-prim well after she had graduated, moved onto better things, and probably wanted nothing to do with it anymore.

I'd also like to thank my QCB/CPB comrades, especially Drs. Hope Woods, Shannon Smith, and Bob Chen. All three greatly outachieved me throughout grad school and I'm lucky to have been friends with such brilliant scientists. Having Shannon and Hope just down the hall, ready to take a walk around campus to catch up and blow off steam was a joy, and I thank them both for putting up with me. Thank you to them for making me feel like an honorary part of the Meiler lab, even though I'm not sure I wanted to be. Thank you to Bob for being a really cool guy. I will forever remember the eight hour drive we took on a Tuesday night to go see a concert in Bloomington, IN for how it solidified our bond. Thank you for always being up for anything or nothing, for consuming content and cheap tobacco with me and for always having an interesting thought to share.

Finally, I'd like to thank friends and family members who supported me throughout this process: Alexis Myers for being an unexpected partner in crime and for always answering my phone calls to commiserate about grad school and relationships and for always providing sound advice whenever I needed it; Ted Summer for always providing a sense of family and like-mindedness no matter how infrequently we are able to see each other; Tyler Peyman for being a for being such a beautiful creative soul, for letting me pretend to be one with you sometimes, and for being the best Craigslist roommate anyone could ask for; Peter and Alex Lilios for being the best Greeks in the world; to my parents for their love and support and for understanding every time I missed out on a family vacation because I had experiments to run; and my brother Alex for being there in spite of everything I put him through over the years. I could not have made it to the end without their love, patience, advice, and support and I am eternally grateful.

Contents

List of figures

List of Tables

List of Abbreviations

SEC	Size-exclusion chromatography
SAXS	Small angle X-ray Scattering
R_G	Radius of Gyration
P_X	Porod Exponent
D_{max}	Maximum Dimension
MW	Molecular weight
EM	Electron Microscopy
CTF	Contrast Transfer Function
NMR	Nuclear Magnetic Resonance
ITC	Isothermal Titration Calorimetry
MST	Microscale Thermophoresis
Kd	Dissociation constant
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
NTP	Nucleoside triphosphate
ATP	Adenosine triphosphate
Pol	polymerase
Pol-prim	polymerase α-primase
PRIM1	Primase large subunit
PRIM2	Primase small subunit
POLA1	Polymerase alpha catalytic subunit
POLA2	Polymerase alpha accessory subunit
CTD	C-terminal domain
NTD	N-terminal domain
PCNA	Proliferating cell nuclear antigen
MCM	Minichromosome maintenance protein
GINS	Go Ichi Ni San complex

Cdc45	Cell division cycle protein 45
CMG	Cdc45-MCM2-7-GINS complex
Ctf4	Chromosome transmission fidelity factor 4
AND-1	Acidic Nucleoplasmic DNA-binding protein 1
RPA	Replication factor A
CaM	Calmodulin
Prp40	Pre-mRNA processing factor 40 homolog B
FF3	FF domain 3

CHAPTER 1. INTRODUCTION[1]

Overview

Propagation of the genetic material is a fundamental chemical process that is responsible for the inheritance of physical traits across biological kingdoms. The faithful replication of DNA, the molecule that encodes genetic information, is thus an essential requirement for life. DNA replication is a complex, multistep process; first the double-stranded DNA is unwound by helicases, enabling complementary copies of each strand to be synthesized by DNA polymerases (1). At the junction between the parental duplex DNA and the unwound single-stranded templates (termed the replication fork), the coordination of a substantial number of proteins is required to maintain orderly and accurate replication of the DNA.

In eukaryotes, three replicative polymerases (pols) perform the bulk of daughter strand synthesis: pol α-primase, pol δ, and pol ε (2–5). Pols δ and ε are processive polymerases that are able to proofread their products generating accurate daughter strands, but neither is able to function without having a ~30 nucleotide complementary primer already in place. Pol α-primase (pol-prim) is responsible for synthesizing this primer and is the only replicative polymerase capable of initiating synthesis on ssDNA template (6).

Pol-prim synthesizes a chimeric primer through a dual RNA-then-DNA mechanism in which the primase half of the enzyme first synthesizes 7-10 nucleotides of RNA followed by extension of the primer with ~20 nucleotides of DNA by the pol α half of the enzyme. Synthesis of the primer across these dual active sites and in the context of the highly complex replisome

[1] This chapter has been adapted from a review chapter in the 3rd edition of the Encyclopedia of Biochemistry titled "DNA polymerase alpha-primase: Biochemical and Structural Mechanisms" published in 2020 and updated corresponding to recent developments in the field.

requires careful coordination. Although pol-prim was one of the first polymerases discovered and has been studied for decades, elucidation of the structural mechanisms to explain the regulation of primer synthesis remains an area of ongoing investigation. The central line of inquiry for this dissertation research is to elucidate the structural and biochemical mechanisms of specifically the RNA primer synthesis half-reaction of pol-prim.

Orchestrating Replication: Organization of the replisome and coordination of the Replicative Pols

Fundamentally, replication consists of the unwinding of double-stranded parent DNA into two single strands, after which the replicative machinery synthesizes complementary daughter strands, thus creating two complements of duplex DNA. However, as the replicative polymerases operate in a uniformly 5'-3' direction (reading the parent strand in a 3'-5' direction), only one of the strands can be synthesized in a continuous fashion (the leading strand) while the other must be synthesized discontinuously (the lagging strand) (7). On both strands, pol-prim must synthesize a chimeric RNA/DNA primer for the processive pols δ and ε to duplicate the bulk of DNA. The discontinuous nature of the lagging strand requires repeated priming by pol-prim, and ultimately many thousands of primers are synthesized during replication. Canonically, Pol ε has been thought to be responsible for synthesizing the bulk of the leading strand while Pol δ synthesizes the lagging strand, but more recent evidence suggests that pol δ may be responsible for synthesizing the initial portion of the leading strand as well immediately after priming by pol-prim (8,9). Importantly, pols δ and ε have both much higher processivity as a function of the processivity clamp PCNA and much higher fidelity due to their inherent proofreading ability compared to pol-prim (5,10). Pol-prim is highly error-prone and as such, most of the primer is removed during Okazaki fragment maturation on the lagging strand (11,12). During this process, pol δ performs strand displacement synthesis, replacing the primer synthesized by pol-prim. As

this strand-displacement is length limited, it is essential that the length of the primer synthesized by pol-prim is highly regulated, lest a large amount of error-prone pol-prim synthesized RNA/DNA make it into the genome, resulting in mutations. Thus, the length regulation ability of pol-prim and its coordination with the other polymerases is highly relevant to genomic integrity and is implicated in some mechanisms of carcinogenesis. After the primer has been displaced by pol δ, it is removed by Fen1 nuclease and the Okazaki fragment is ligated to the rest of the synthesized DNA by DNA ligase (11,12).

Although polymerases are the key enzymes directly responsible for DNA synthesis, careful coordination of replication is accomplished through a highly complex protein machine that has been termed the replisome. A dizzying array of proteins are associated with the replisome and its composition is temporally variable as replication proceeds through its various stages and as the replisome encounters potential obstacles (13). At its most simplified level, the replisome consists of several core components (Figure 1): the CMG helicase consisting of the Cdc45, MCM2-7, and GINS subcomponents that are together responsible for the unwinding of DNA (14–17); Pols δ and ε together with the PCNA processivity clamp and RFC clamp loading protein which are responsible for the bulk of DNA synthesis (10,18,19); RPA which protects single-stranded DNA after it has been unwound (20–23); Pol-prim which synthesizes the chimeric RNA/DNA primer; and Ctf4/AND-1 scaffolding protein which recruit and organize a number of replication factors including the CMG helicase and pol-prim (24–28). Further factors that are associated with the replisome include MCM10 which acts as a scaffolding protein and regulates fork progression (29), and the fork protection complex (FPC) consisting of Claspin, Timeless, and Tipin (30).

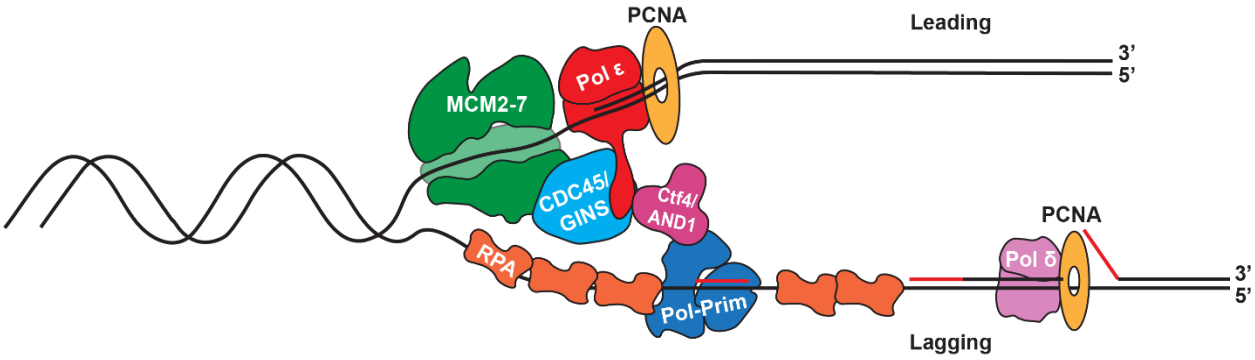

Figure 1.1 Schematic of a replication fork.

The CMG complex consisting of MCM2-7, CDC45, and GINS is responsible for unwinding the double-stranded parent DNA. Following separation of the strands, single-stranded DNA is bound by RPA. Ctf4 (yeast) or AND1 (metazoan) acts as an interaction hub for multiple replication factors. Pol-prim synthesizes primers (red) on both the lagging and leading strands for processive polymerases δ (lagging strand) and ε (leading strand). Processivity clamp PCNA increases the processivity of both by encircling double-stranded DNA.

Recent EM structures of replisome complexes have established several key interactions between these components. Importantly for the present discussion, interactions for pol-prim have been detected between Ctf4/AND-1 (24,26,27), MCM10 (31), Tipin (32), GINS (33), and RPA (20,34). Spatially, these binding partners appear to be far apart at the replication fork, and the nature of organization or order of binding to pol-prim remains unclear. One series of cryoEM structures of a yeast "replication factory" in which pol-prim was coordinated with one or more CMG helicases through Ctf4 suggests that simultaneous replication may occur for two replication forks as part of a spatially coordinated bidirectional replication bubble (26). Ultimately though, further structural investigation is needed to properly assess the organization of the replication fork throughout replication. Nonetheless, it remains apparent that replication is reliant on a highly structurally dynamic interconnected web of protein-protein interactions and that the coordination of all these factors is essential for genomic integrity.

Polymerase α-Primase Basics

Pol-prim is comprised of two different polymerases, primase and pol α, whose activities are distinct but tightly coordinated. Primase is a DNA dependent RNA polymerase that initiates daughter strand synthesis by generating the initial primer of 7-10 ribonucleotides. This initial RNA primer is the intramolecularly transferred to pol α, which extends the primer by ~20 deoxyribonucleotides (35).

Primase and pol α each contain a catalytic and regulatory subunit. In primase, the catalytic subunit PRIM1 (a.k.a. p48, p49, PriS, Pri1) is constitutively bound to the regulatory subunit PRIM2 (a.k.a. p58, PriL, Pri2) through its N-terminal domain (PRIM2N). Pol α consists of a catalytic subunit POLA1 (a.k.a. p180, Pol1) and a regulatory subunit POLA2 (a.k.a. p68, p70, B subunit, Pol12). The C-terminal domain of POLA1 (POLA1C) is a critical interaction center scaffolding between PRIM2N and POLA2, while connected through a flexible linker to the catalytic domain of POLA1 (POLA1cat). All four subunits are necessary for priming in cells (36). However, the primase and pol α subunits can be generated independently which enables their in vitro biochemical activities to be characterized. These studies have been useful for dissecting the action of pol-prim and defining the respective roles of the four subunits.

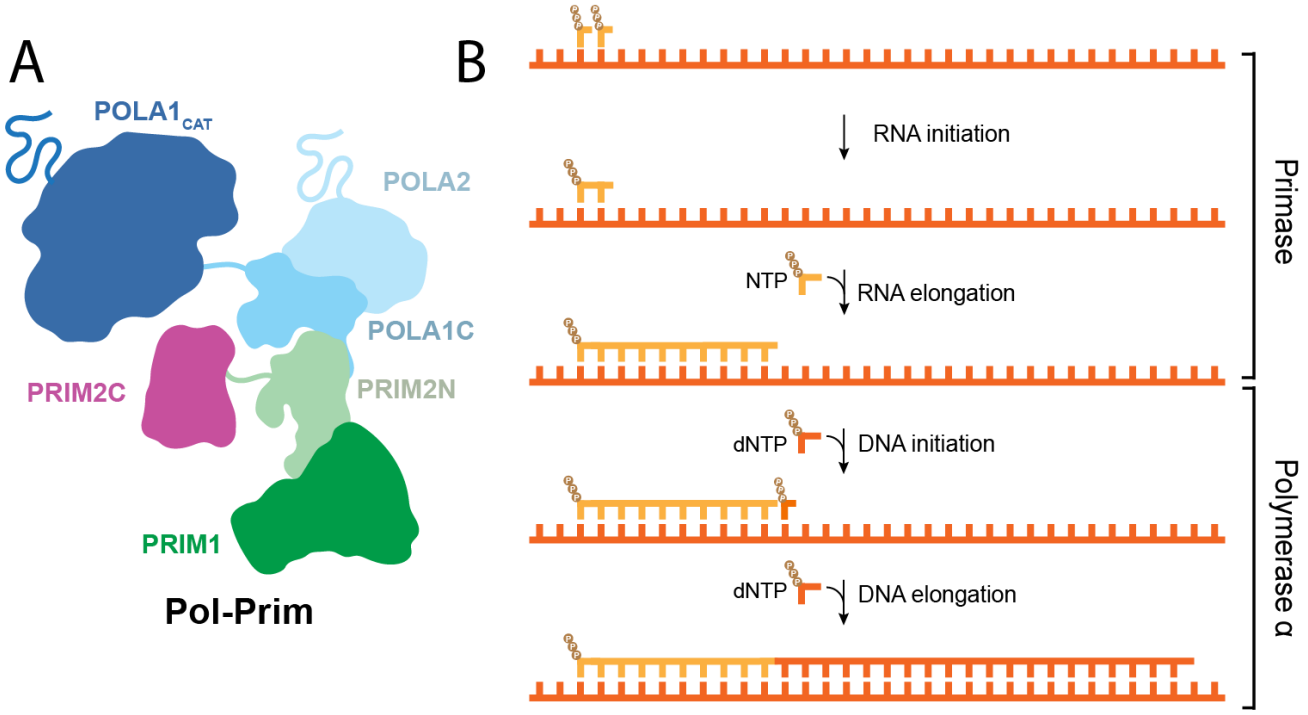

Figure 1.2 Pol-prim schematic.

A) Domain schematic of pol-prim. B) Schematic of chimeric primer synthesis by pol-prim.

The isolated primase catalytic subunit PRIM1 is able to generate complementary RNA strands in the presence of single-stranded DNA, ribonucleotides and divalent metal cations (typically Mn^{2+} or Mg^{2+}), albeit inefficiently (37,38). It uses the common two-metal catalytic mechanism to synthesize these primers. In the presence of PRIM2, PRIM1 synthesizes primers more efficiently. PRIM2N functions as a scaffold and is tightly bound to PRIM1. The PRIM2 C-terminal domain (PRIM2C) is flexibly tethered to PRIM2, and aids in initiation of primer synthesis (37). It is essential for generating primers of appropriate length (7-10 nucleotides), i.e. is necessary for primer length "counting."

The catalytic subunit of pol α contains three functional domains: POLA1N, POLA1cat, and POLA1C. The disordered POLA1N is key for interactions with other replisome components. Beyond this, it is unclear if POLA1N has additional functional roles; in most in vitro studies

POLA1N is truncated to improve solubility (39). The POLA1cat domain is highly conserved among eukaryotes and adopts the same fold as pol δ and ε (40). This domain in isolation elongates RNA or DNA primers with deoxyribonucleotides (41). In the cell, elongation of the initial RNA primer with deoxyribonucleotides is limited to ~20 nucleotides; the processivity of this domain in vitro is still under debate (40–43). The POLA1cat domain is flexibly tethered to the POLA1C domain, which in turn associates tightly with the PRIM2N domain and the regulatory POLA2 subunit. POLA2 stabilizes the POLA1C domain and is important for proper targeting of pol-prim during the cell-cycle (44,45). It also interacts with other replisome factors in metazoan systems such as AND-1, but notably not the yeast orthologue Ctf4 (24). Instead, the interaction between yeast pol-prim and Ctf4 is mediated by POLA1 (27,28).

Three-dimensional structure of Pol-prim

Structural characterization of pol-prim has been extensive over the past 10-15 years, initially using a "divide and conquer" crystallographic strategy. High-resolution 3D structures have been determined for each of the globular domains of pol-prim alone and/or as part of larger multidomain complexes. Several crystal structures have also been determined in complex with model RNA and DNA substrates. More recently, cryoEM structures of the full pol-prim heterotetramer have been determined that inform much of the understanding of how the flexible domains are organized in both the absence and presence of model substrates.

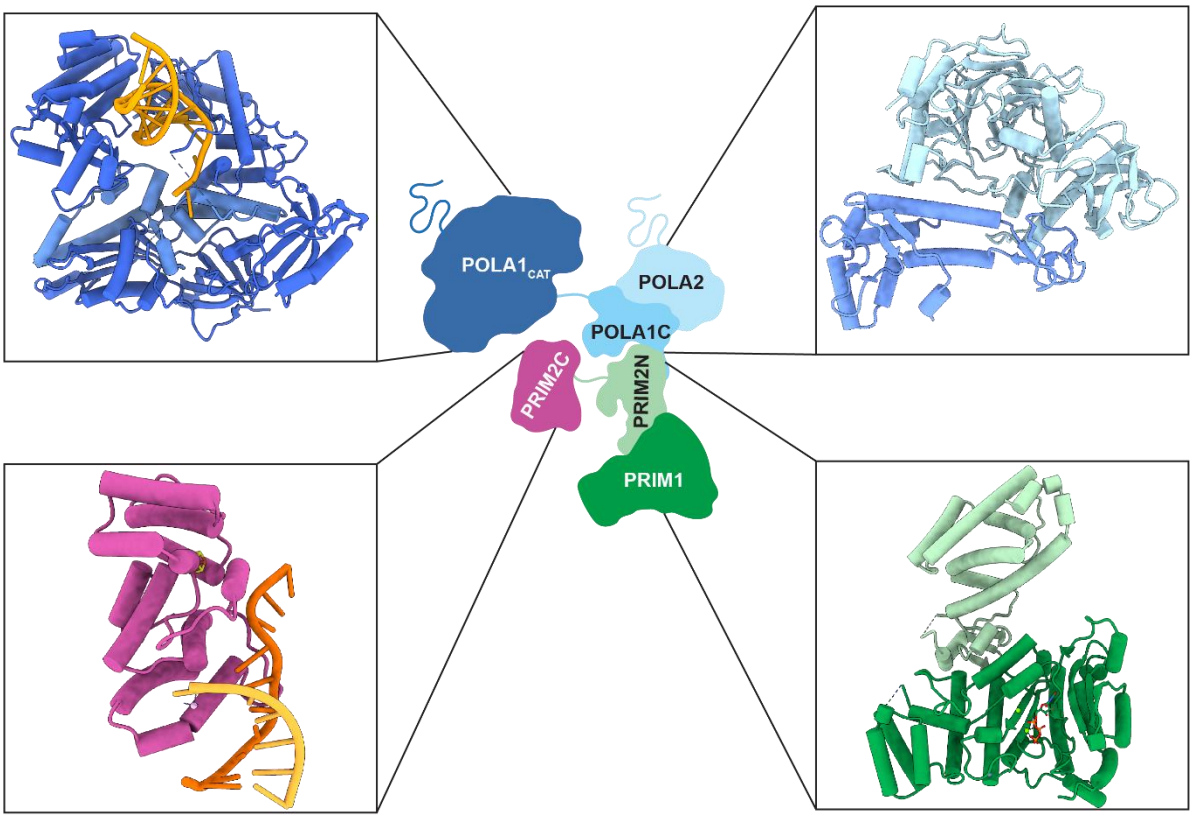

Figure 1.3. Crystal structures of constituent subdomains of pol-prim.

Structure of POLA1C/POLA2 (PDB:4Y97), PRIM2N/PRIM1 (PDB:4BPU), PRIM2C (PDB:5F0Q), and POLA1CAT (PDB: 5IUD).

Structures of Primase

The primase heterodimer contains three globular domains: One that covers the large majority of PRIM1 and two others in PRIM2 (PRIM2N, PRIM2C), which are tethered by a ~20 residue flexible linker (46). The PRIM1 domain in eukaryotes adopts a characteristic archaeo-eukaryotic primase fold, in which a conserved mixed helix-sheet region forms the core around a typical aspartate-divalent metal active site (47–49). Several crystal structures have been determined with different combinations of NTPs and catalytic metals (Table 1.1) (50–52). Incoming NTPs are stabilized in the active site by key contacts between the negatively charged

phosphate groups of the nucleotide triphosphate and the positively charged Arg and Lys side-chains of PRIM1. Importantly, no structures have been determined of PRIM1 bound to either a single stranded template or RNA-primed DNA substrate representing a critical gap in knowledge. Eukaryotic primases have a helical accessory subdomain that diverges in both sequence and structure from archaeal primases (49,53,54). A zinc motif is located in the interface of the core and accessory regions neighboring the catalytic center. The interface with PRIM2N is on the opposite side of the Zn^{2+} binding module.

The PRIM2N domain serves as a critical scaffold mediating contact between primase and the POLA1 subunit of pol α. The domain consists of a mixed sheet-helix subdomain that serves as the interaction platform with PRIM1 and primary helical subdomain that interacts with POLA1C. PRIM2N is unstable in the absence of PRIM1 and is tightly bound to it with an interface of ~1600 $Å^2$ of predominantly hydrophobic surface. In eukaryotes, PRIM2N is tethered to the unique PRIM2C domain by a ~20 residue flexible linker. PRIM2C itself is a compact helical domain with a 4Fe-4S cluster buried within the protein (55–57). The domain has a nucleotide binding site with a highly basic groove. A structure of PRIM2C bound to an RNA-primed DNA substrate has been determined, showing the key interactions with the 5' terminal triphosphate of the RNA primer strand and the 3' overhang of the template strand of the RNA/DNA duplex (58).

A crystal structure has also been determined of the full-length primase (PRIM1/PRIM2) in an "open" configuration in which the PRIM2C domain is located far from the PRIM1 active site (46). It is highly unlikely that this is an active configuration for primase. Rather primase must occupy a "closed" configuration to generate the initial dinucleotide, with the PRIM2C

template and nucleotide binding sites positioned close to the PRIM1 catalytic center. Such a change in configuration is enabled by the flexible linker tethering PRIM2C and PRIM2N.

Structures of Pol α

The pol α hetero dimer contains three globular domains: the POLA1 large central catalytic (POLA1cat) and C-terminal domains (POLA1C) and a central core domain of POLA2. The two domains in POLA1 are tethered by a ~30 residue flexible linker. Consequently, POLA1cat is structurally independent from the rest of the pol-prim complex. In addition to these structurally characterized globular domains, significant N-terminal disordered domains exist for both POLA1 and POLA2. These domains function as interaction sites for binding to other replication factors.

Crystal structures of POLA1cat revealed a classic B-family polymerase hand fold, consisting of an active site located in the palm domain, a fingers domain responsible for binding incoming nucleotides, a thumb domain that grasps the duplex, and an inactive exonuclease domain. Structures of POLA1cat have been determined in complex with several different model primer substrates including RNA/DNA and DNA/DNA duplexes (40,59,60). These structures provided preliminary insight into primer length counting by pol α in that the interaction between the RNA/DNA duplex and pol α was more extensive than that of the DNA/DNA duplex.

No crystal structures have been determined for the isolated POLA2 or POLA1C domains, but there are several of their complex (61,62). These structures show a compact slab-like structure in which the helical POLA1C domain forms a saddle shape that abuts the triangular POLA2 domain. The interaction between the two domains is stabilized by a large interface that occludes ~4500 Å and is mediated by a number of polar and hydrophobic interactions. Two

metal binding motifs are observed on opposite sides of POLA1C that are occupied by zinc in all available structures, though preliminary work in the Chazin lab has indicated that these may be filled by Fe-S clusters under anaerobic conditions.

Early Structures of Pol-Prim Tetramer

Though informative, structural data for the individual domains and subcomplexes are not sufficient to define the mechanism and organization of pol-prim as it synthesizes chimeric RNA-DNA primers. The first step towards capturing the structural organization of the pol-prim tetramer was a low-resolution negative-stain EM structure of yeast pol-prim (63,64). A bilobal structure was observed with significant interlobal flexibility evident from variability in the relative positioning of the two lobes in different 2D class averages. This structure was informative despite its low resolution because it revealed the ability of pol-prim to undergo significant configurational rearrangements, which are commonly accepted as being essential to the coordination and regulation of primer synthesis across the two polymerase active sites.

Subsequently, a substantially higher resolution 3.6 Å crystal structure was determined for human pol-prim in the absence of substrate or cofactors (58). This structure was valuable for visualizing the organization of the globular tetramer core (a.k.a. platform) of pol-prim that consists of the PRIM1, PRIM2N, POLA1C, and POLA2 domains. However, the POLA1cat domain is seen to occupy an inactive conformation with the thumb domain splayed open while the PRIM2C substrate binding domain is occluded by POLA1cat and is unable to bind substrate for primer initiation. The conformation observed in this structure has been termed the autoinhibitory (AI) conformation and has been observed in multiple subsequent cryoEM structures in the absence of nucleotide substrate or in the presence of lower affinity substrates.

Insights into the Mechanism of RNA priming by Pol-Prim

Biochemical studies of pol-prim in vitro, along with 3D structures, have been vital in generating knowledge of the mechanisms by which primase and pol α generate primers. The studies have been greatly enabled by the fact that these two enzymes maintain at least some of their biochemical functions when separated from each other. Moreover, even the isolated catalytic domains, PRIM1 and POLA1cat, retain some biochemical activity. Because pol-prim is a complex dual enzyme and the components are easier to produce and work with than the intact tetramer, the majority of in vitro investigations of priming mechanisms have been performed with either isolated primase, isolated POLA1cat, or soluble fragments of the tetramer.

The initial RNA priming step

DNA primase has been the object of intense investigation because it is the sole enzyme capable of initiating oligonucleotide synthesis on a DNA template without a primer and is therefore responsible for initiating daughter strand synthesis in replication. In addition, it has the curious property of synthesizing initial RNA primers of limited length (7-10 nucleotides). These properties are fundamental to primase function and are observed for both the full pol-prim tetramer and the isolated primase heterodimer. Primase synthesizes the RNA primer in two steps: formation of the initial dinucleotide and extension of from the dinucleotide. Simultaneous binding of the template DNA, of nucleotides that are incorporated into the primer, and catalytic metals are required to generate the initial dinucleotide; this is the rate-limiting step in RNA priming. The subsequent elongation of the dinucleotide is much more rapid (65,66).

Critical mechanistic insights into primase function have been obtained using two biochemical techniques: (1) de novo RNA synthesis (initiation) assays requiring synthesis

without a pre-made primer; (2) elongation assays that extend from an existing primer. In a de novo RNA synthesis assay, primase is incubated with radiolabeled nucleotides, template DNA, and catalytic metals (Mg^{2+} or Mn^{2+}). The results are interpreted in terms of the total amount of products at different timepoints or by analyzing the product length and product abundance over time. In an elongation assay, primase is incubated with radiolabeled nucleotides, primed-template RNA/DNA, and catalytic metals. The results are interpreted by an analyzing the amount and rate of extension to primer length products and the presence of multimers of primer length products. Specific versions of these priming assays, typically by titrating important components, are easily performed and mechanistically informative.

RNA primer initiation

The primase active site was identified because the PRIM1 subunit contains catalytic residues that are highly conserved in eukaryotic primases. The location of the active sitewas confirmed by mutagenesis studies in yeast cells (49,50). Interestingly, even though PRIM1 contains the RNA catalysis active site, its intrinsic affinity for single stranded DNA and RNA primed templates is weak. Crosslinking experiments with photoactivatable nucleotides generated no observable template-PRIM1 species, only modified PRIM2C species (67). Recently, electrophoretic mobility shift assays revealed that the isolated PRIM2C domain has a very similar affinity for an RNA-DNA junction substrate as the intact primase heterodimer (68). These observations suggest that RNA priming requires both primase subunits, though further clarification of the binding affinities for each subunit would be helpful, as these two studies provide starkly different affinity estimations.

Priming assays analyzing the de novo RNA synthesis activity of primase with mutations in PRIM2C indicate that this domain has a key role in primer synthesis and counting; deletion of

the domain impaired initiation of primer synthesis and completely abolished counting. Alanine scanning mutagenesis coupled with de novo RNA synthesis assays identified residues R302, R306, and K314 as essential for primer initiation and elongation (37,69). The crystal structure of the PRIM2C-product complex shows that R302 and R306 make strong contacts with the 5' nucleotide (58).

Taken together, these results confirm that both the PRIM1 and PRIM2C domains are required for priming, and suggest a mechanism by which dinucleotide synthesis occurs. The PRIM2C domain binds the template DNA and the 5' nucleotide, and PRIM1 binds the 3' nucleotide and provides the active site for polymerization of the two nucleotides complementary to the template. In this model, PRIM2C must be positioned over the active site of PRIM1 for initiation (generation of the initial dinucleotide) to occur. Remarkably, despite the large number of crystal and EM structures available, none have shown primase in this configuration. The only available crystal structure of free primase occupies an open configuration with PRIM2C extended far away from the PRIM1 domain (46). A recent EM structure of pol-prim in the presence of template DNA likewise shows PRIM2C separated from PRIM1 (70). Hence, a significant reorganization, mediated by the flexible linker tethering PRIM2C and the PRIM1-PRIM2N-POLA1C-POLA2 tetramer core, would be required to attain the active configuration. We and others propose that the positioning of PRIM2C is the rate limiting step of RNA primer initiation (66). However, direct evidence is lacking; structures of primase in the active configuration are required to convincingly determine the mechanism of RNA primer initiation.

RNA primer elongation and counting

Once the initial dinucleotide is formed, its extension to 7-10 nucleotides is relatively rapid. However, generation of the primer length product is relatively inefficient overall. De novo

RNA synthesis assays with isolated primase revealed that the most abundant product is a dinucleotide rather than primer length primers of 7-10 nucleotides (35,46,66,71). This indicated that once the dinucleotide is formed, the primase-template-dinucleotide complex more readily dissociates than extends on to a primer length product. The abundance of dinucleotide products has been attributed to the low annealing temperature of the dinucleotide to the template and may also be influenced by the low affinity of PRIM1 for substrate. Support for this hypothesis was provided by a series of de novo RNA synthesis assays with isolated primase, which invariably produced many times more dinucleotide products for every primer-length product (35,37,66).

After extension to approximately 7-10 nucleotides, primase pauses and the substrate is handed off to pol α (35,46,65,66). In de novo RNA synthesis assays with isolated primase, a low level of primer multimers were observed, whose abundance increased over time. These were believed to arise because primase can either dissociate from the primer-length primer or re-engage the same or a different primer and extend it by another primer length. However, it turns out that primer multimers are not generated in assays with the intact pol-prim tetramer. Hence, these multimers are viewed as artifacts of working with isolated primase in vitro.

Termination of elongation at only 7-10 nucleotides is one of the most notable aspects of RNA priming. The mechanism by which primase is able to "count" the primer length has been a subject of intense speculation, has been a major focus of this thesis work, and is discussed in more detail in Chapter III. Based on available data, the unique 4Fe-4S PRIM2C domain appears to be key to the ability of eukaryotic primers to count. Early studies using de novo RNA synthesis assays noted PRIM2C deletion and missense mutations drastically altered counting ability (37).

Subsequent crystal structures, binding assays, and primer elongation assays by the Tahirov group suggested that the interaction of PRIM2C with the triphosphate of the 5' nucleotide is particularly important. One informative study revealed that primase preferentially extends, and counting is enhanced for, a primer that has a triphosphate on the 5' nucleotide relative to a primer that has a monophosphate on the 5' nucleotide (58). These observations led to the proposal that the PRIM2C domain remains bound to the 5' triphosphate throughout primer synthesis. In this model, PRIM2C and PRIM1 interact with the primer throughout elongation, moving apart as the primer elongates. This is an attractive hypothesis as it is supported by data indicating the PRIM2C has a role in initiation, elongation, and counting. Importantly, while elongation and counting are enhanced for primers with a triphosphate on the 5' nucleotide, primase does not count efficiently when extending primers lacking this triphosphate (39,72).

Over the years, three mechanisms for counting have been proposed. In all cases, the flexible tethering of PRIM2C through the ~20 residue flexible linker to the tetramer core plays a major role. In the simplest model, the linker physically limits the distance between the active site in PRIM1 and the template binding region of PRIM2C, preventing primer extension beyond a defined primer length. However, de novo RNA synthesis assays with PRIM2C mutants with different linker lengths had no apparent effect on the ability of primase to synthesize products of the appropriate length (46). Additionally, structural modeling of the organization of primase bound to an elongated primer shows that there is significant slack in the flexible linker (58). Instead, the linker is likely most extended during the RNA initiation step. Another model proposed by the Tahirov group attributes counting to PRIM2C remaining bound to the 5' nucleotide throughout the RNA priming step and encountering steric hindrance near the end of its catalytic cycle (58). In this model, as the primer elongates, PRIM2C moves away from the

PRIM1 active site until is hindered from further movement by tetramer core. One issue with this model is that it is inconsistent with the generation of primer multimers that are observed in primer elongation assays. Additionally, this mechanism provides a basis for the upper limit of primer length but does not adequately explain the lower bound of primer length at 7.

A third proposed mechanism for counting includes the movement of PRIM2C away from the active site of PRIM1 as the RNA primer grows, but also invokes a role for change in the redox state of the 4Fe-4S cluster in PRIM2C. This model is derived from the observation that PRIM2C is more tightly associated with substrate when the 4Fe-4S cluster is oxidized, than when it is reduced (73). In this model, the cluster switches from oxidized to the reduced state after the primer reaches 7-10 nucleotides, thereby driving disengagement of the primer by primase. However, recent structures of pol-prim during DNA primer extension by POLA1cat show that PRIM2C remains bound to the primer during this stage, a scenario that would not occur in the redox RNA primer counting mechanism (42,43). Additionally, theoretical modeling of the redox potential of PRIM2C and potential tyrosine charge-hopping pathways within the domain to nucleotide substrate suggest that a DNA charge transport mechanism is unlikely (74).

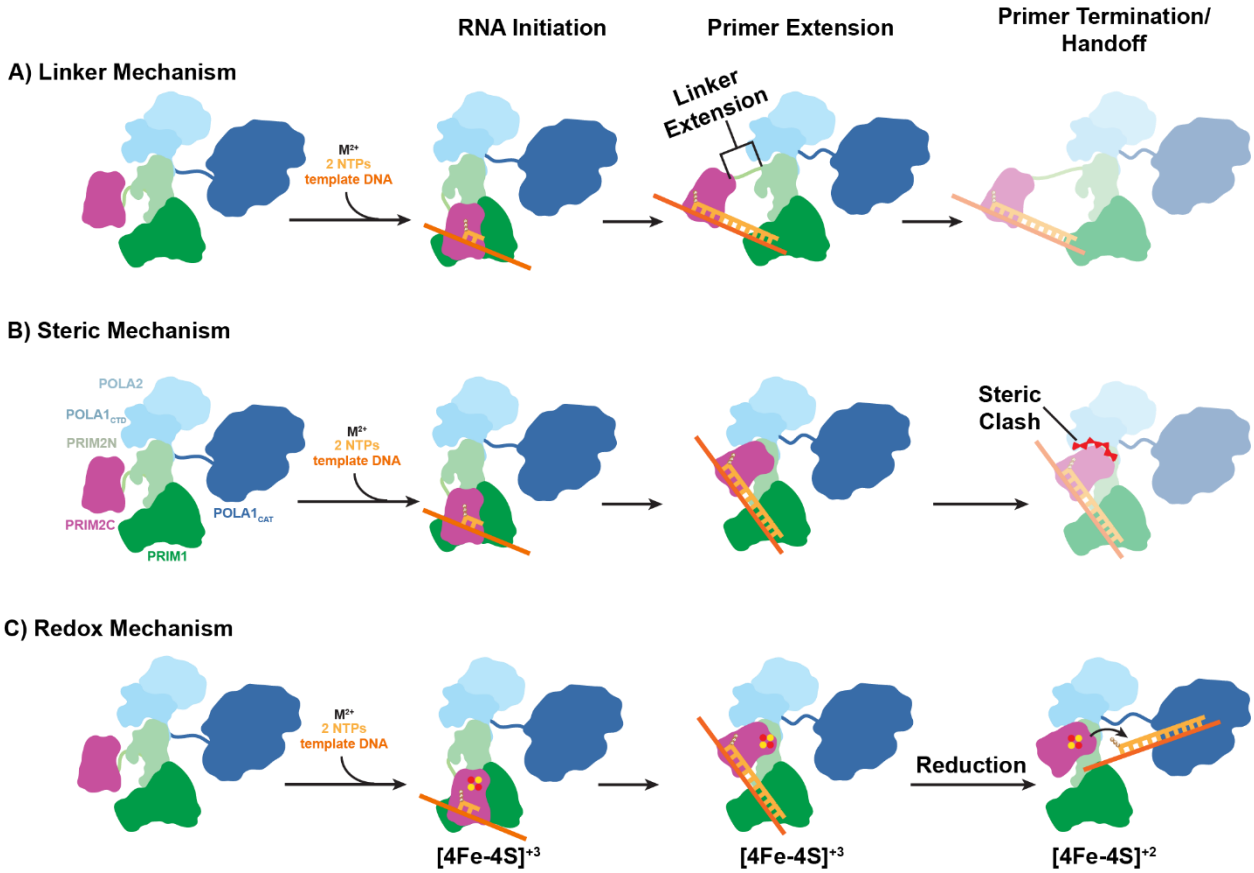

Figure 1.4 Proposed mechanisms of RNA primer length regulation.

In all three mechanisms PRIM2C plays a key role A) PRIM2C and PRIM1 become spatially separated as the primer is elongated, gradually extending the flexible linker between PRIM2C and PRIM2N to a maximum length that restricts further primer synthesis. B) PRIM2C becomes separated from PRIM1 as the primer is extended, but the path of PRIM2C results in a steric clash with PRIM2N at longer primer lengths, restricting further synthesis. C) PRIM2C binds the primer with high affinity in the oxidized state, but upon reduction binds with much lower affinity, allowing dissociation of the primer and transfer to POLA1CAT.

In summary, although several hypotheses have been proposed, a comprehensive explanation for how primase counts remains elusive. While each of the three proposed mechanisms has some support, the dearth of relevant structural information has inhibited understanding for structural mechanisms of primer length counting and architectural organization of the primase subunits during primer synthesis.

Structure of Pol-prim during RNA priming

In the past two years, a flurry of cryoEM structures has been reported of the full pol-prim tetramer, greatly expanding our understanding of the structural transitions that occur during primer synthesis (42,70,75–78). A trio of papers were first published describing pol-prim in complex with the telomeric factor CST while another three papers have described pol-prim in the absence of other replication factors, but in the presence of nucleotide substrates. Of these studies, several structures recapitulated the autoinhibitory conformation observed in the absence of substrate, while most others captured novel configurations bound to DNA priming substrates (either chimeric RNA/DNA primed or DNA primed DNA templates). However, of the 20+ structures of pol-prim deposited to the PDB from these papers, only three were relevant to the RNA priming half-reaction of pol-prim. In the first, the structure of pol-prim in the presence of template ssDNA and CST was determined by the Lim group (75). In it, CST scaffolded and stabilized the tetramer core and high-resolution density was observed for all subunits except PRIM2C, for which only low-resolution density was observed suggesting it occupied a range of configurations. After 3D variability analysis, density for PRIM2C was observed in one configuration termed the pre-initiation state in which it was positioned near to the PRIM1 catalytic center bound to ssDNA, poised for primer initiation.

Two other structures were determined by the O'Donnell group, one again in the presence of ssDNA template and the other in the presence of an RNA primed template (70). In the presence of template DNA, architecture of pol-prim was starkly different from that of the autoinhibitory complex as all subunits were available for substrate binding, suggesting that the presence of template DNA alone may drive the unfolding of the auto-inhibitory conformation. An alternative possibility is that for free pol-prim there is an equilibrium between the

autoinhibitory conformation and an unfolded conformation that is capable of binding the template DNA. This DNA binding then stabilizes the unfolded conformation and structural transition back to the autoinhibitory state. Notably however, no density was observed for the single-stranded DNA in the structure and the location of the PRIM2C domain was not near the catalytic center of PRIM1, suggesting that the configuration captured in this structure is not the configuration adopted for RNA initiation. Importantly, for the RNA primed template structure, the RNA primer used in the assay does not appear to have been triphosphorylated. As the 5' triphosphate is critical for binding PRIM2C, this is a major cause for skepticism of the structure. To ensure binding of the primed template substrate to the protein, the authors employed a formaldehyde crosslinking agent that resulted in covalent crosslinks between the PRIM2C domain and the substrate, but in an orientation that is dissimilar from the published crystal structure of PRIM2C or EM structures of PRIM2C bound to longer primed template substrates. Finally, in this structure, the 3' of the substrate is not located near the PRIM1 catalytic site. In fact, no part of the substrate interacts with the PRIM1 domain suggesting that the captured configuration does not represent an actively elongated substrate.

Altogether, initial structural work has begun to illuminate the architectural organization of pol-prim during RNA catalysis but are hampered by limitations in structural studies to date. A large gap in knowledge remains in trying to understand the biochemical and structural mechanisms of RNA primer synthesis.

CHAPTER 2. STRUCTURAL INVESTIGATION OF MODULAR PROTEINS BY INTEGRATIVE METHODS

Overview

The three-dimensional structure of proteins ultimately dictates their biological functions. As such, investigation of protein structure has in the past several decades matured into an integral part of any biological inquiry, providing a basis for understanding how the infinite chemical reactions that fuel life are orchestrated at the molecular level. Until recently, the field of structural biology had a tendency towards parochialism: X-ray crystallographers employed a divide-and-conquer approach that could provide snapshots of rigid domains but lacked the ability to observe flexibility while NMR spectroscopists could provide a wealth of information about flexibility in solution but were often limited by size constraints. Twin currents in the past decade have pushed the field of structural biology to new frontiers and provided fertile grounds for an explosion of structural knowledge. The first is drastic technical advancements in cryoEM and computational structural prediction. In the case of cryoEM, the "resolution revolution" has seen a drastic increase in the attainable resolution for structures determined by EM and has allowed for investigation of large multi-protein complexes while advancement in classification algorithms has allowed for the characterization of structural flexibility. Meanwhile, the development of machine learning methods for structure predictions, beginning with the release of AlphaFold, has seen a proliferation of high-confidence structural models from simulations that require limited user expertise. The second major current in structural biology, which has been both reliant on and a motivation for the above-mentioned technical advancements, has been the increasing appreciation for the highly complex nature of the protein machinery that drives cellular processes. In most cases, proteins function not as lone rigid globular domains that are amenable

to crystallography, but as multi-domain intricate, tightly coordinated, flexible protein machines. As discussed in Chapter I, this is especially true in the case of DNA replication and at a smaller scale for pol-prim. Ultimately, modern structural biology relies on an integrative approach that leverages the insight from multiple experimental techniques to illuminate structural mechanisms.

Solution Scattering

Scattering techniques can be powerful tools to understand biophysical properties of proteins and other macromolecules. Small-Angle X-ray Scattering (SAXS) in particular can provide a wealth of information on the size, shape, and flexibility of proteins (79). While X-ray crystallography relies on diffraction from an ordered crystal matrix, SAXS is a solution method in which scattering of the molecule is measured from all particles present in the solution, regardless of orientation, conformation, or oligomeric state. While this averaging of scattering signal ultimately results in lower resolution information, the combination of this data with higher resolution domain structures can provide powerful insight into the organization of flexible, modular proteins.

SAXS basics

Similar to X-ray diffraction techniques, acquisition of SAXS data involves exposing the sample to a high flux, coherent x-ray beam, which is scattered by electrons distributed throughout the molecule (79–81). The scattered x-rays then travel forward to the detector, undergoing constructive and destructive interference until forming a unique intensity pattern. Several major factors differentiate scattering from x-ray crystal diffraction techniques. First, the sample in this case is a (preferably monodisperse) solution of macromolecules or macromolecular complexes rather an a highly organized crystal. Second, due to the solution

nature of the experiment and unlike the diffraction pattern from an ordered crystal, the scattering intensity is radially symmetric about the detector center as a consequence of random molecular orientation in solution. This symmetric nature allows radial integration and simplification of the data from a two-dimensional detector image to a one-dimensional scattering curve. The scattering curve, I(q), is the integrated intensity at a given momentum transfer angle, q, which can be defined several ways but for these purposes corresponds to the scattering angle, 2θ, by the equation $q = 4\pi \sin(\theta) / \lambda$, where λ is the x-ray wavelength used in the experiment. The intensity of scattering decreases rapidly beyond 1° (momentum transfer, q, values beyond 0.3 $Å^{-1}$) and data beyond this angle is typically truncated, thus only small-angle scattering is considered. Due to this rapid decay in scattering, high-resolution information is lost, as the scattering angle q is inversely related to the resolution due to the relationship between the detected scattering image and a real-space representation. Similar to an x-ray crystallography diffraction pattern, the image recorded at the detector is the Fourier transform of the real-space shape of the illuminated sample. Below q = 0.3 $Å^{-1}$, the decay in signal is dictated by core biophysical properties including size, shape, surface-to-volume ratio, and electron density of the scattering particle. Various transformations of the first order I(q) curve can inform different second-order properties related to the above-mentioned biophysical properties.

The first transformation provides direct information about the size of the particle. The Guinier transformation of I(q) allows calculation of a particle radius of gyration (R_g) within the low-q Guinier ragion of the curve (defined as $q*R_g < 1.3$) (79,82). The calculation of the R_g is enabled by the relationship between size of the particle and scattering intensity as defined by the following equation:

$$I(q) = I(0)e^{\frac{-q^2R_g^2}{3}}$$

However, the calculation of R_g is reliant on the linear behavior of the scattering curve; the intensity value at angle 0°, I(0), is not measurable as it cannot be distinguished from the high-intensity radiation that passes through the sample unscattered. Thus, the I(0) must be extrapolated using the following equation:

$$I(0) = N(\Delta\rho V)^2 = \frac{C\Delta\rho^2 v^2 MW}{N_A}$$

in which N is the number of scattering particles per unit volume, Δρ is the contrast, and V is the particle volume. This can alternatively be expressed in terms of C, the mass per unit volume, the molecular weight MW, partial specific volume of the particle, v, and Avogadro's number, N_A. However, this linear relationship only applies for relatively globular particles below $q*R_g < 1.3$, while simultaneously being a pre-requisite for accurate calculation of size by the Guinier method. Thus, visual inspection of the Guinier region can also be a powerful and rapid diagnostic tool for sample quality, as it can provide evidence for the presence of aggregation or degradation of the particle from X-ray exposure as determined by non-linear behavior at low-q values.

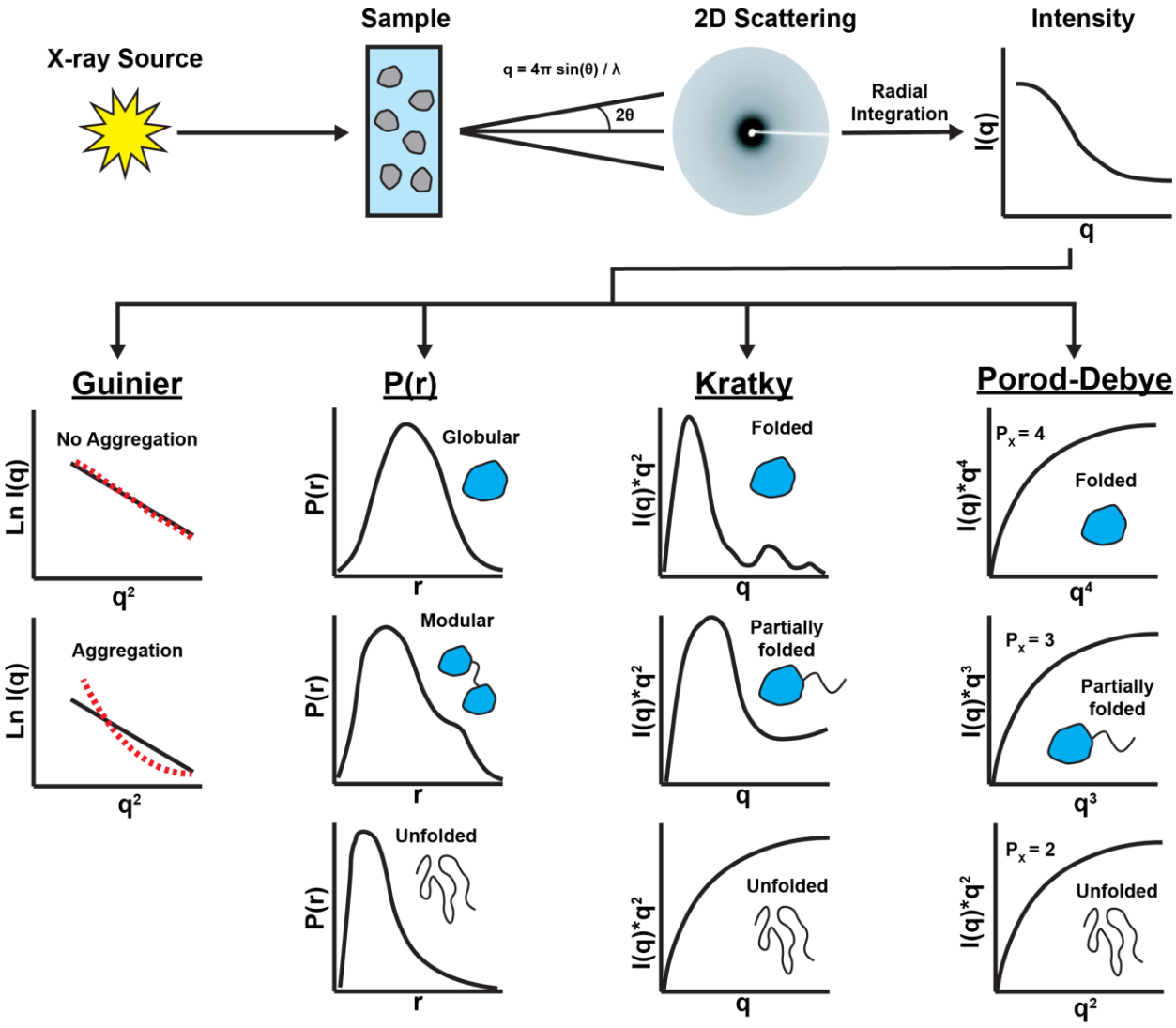

Figure 2.1 Schematic of SAXS experiment and basic data transformations.

A solution of the sample of interest is illuminated by a high-power x-ray source, producing a 2-dimensional scattering pattern at the detector. This scattering pattern is then radially integrated to produce a 1-dimensional scattering intensity curve. From the raw intensity data, mathematical conversion can be applied to obtain key information about the biophysical properties of the sample. First the Guinier transformation can be used to compare the experimental data (red dotted line) to an ideal sample to determine the presence of aggregation. The P(r) transformation provides is a plot of the interatomic distances present in the sample and provides information about the shape and size of the sample. The Kratky and Porod-Debye analyses are influenced by the surface-to-volume ratio of the particle, and are therefore sensitive to changes in the size, shape, and foldedness of the particle.

A second transformation of the data into the Pairwise distribution function, P(r), is enabled by the equation:

$$I(q) = 4\pi \int_0^{Dmax} P(r) \frac{\sin(qr)}{qr} dr$$

where r represents all interatomic distances in a scattering particle, and D_{max} represents the maximum linear dimension or longest interatomic distance in the particle.(79) The calculation of the P(r) distribution involves an indirect Fourier transform of the I(q) curve and frequently requires manual manipulation of D_{max} parameter to produce an adequate fit. This transformation provides invaluable information about the size and shape of the protein. Shape changes especially are readily detectable from the shape of the P(r) distribution. An additional relationship allows the calculation of an independent R_g parameter directly from the P(r) distribution, providing both a good metric for fit of the P(r) distribution to the experimental data and providing another diagnostic of sample quality. Disagreements between the R_g derived from the P(r) distribution and the Guinier method can be a sign of aggregation or interparticle association.

The final more complex set of transformations provides information about the volume, shape, and especially flexibility of the molecule (83). The Kratky transformation involves multiplication of the I(q) plot by q^2. Integration of this curve provides a parameter termed the Porod Invariant, Q, which is related to the volume of the particle:

$$V = 2\pi^2 \cdot \frac{I(0)}{Q}$$

However, this relationship is limited to highly globular particles and breaks down when applied to flexible polymers and extensions. This relationship is defined by the relationship between the

decay in scattering intensity of a globular particle vs. an unfolded particle by the Porod-Debye law which states that the scattering of a folded particle should decay as q^{-4} scaled by the particles surface area, S, concentration, c, and the contrast of the particle, $\Delta\rho$:

$$I(q) \approx \Delta\rho \cdot \frac{2\pi}{q^4} \cdot S$$

Transformation of the data to $q^4 \bullet I(q)$ vs. q allows estimation of whether the particle is globular and reaches a Porod Plateau or contains unfolded character and reaches a plateau in a different q regime. The related Porod Exponent (P_x) parameter provides a simplified version of this metric and can be described phenomenologically; $P_x = 4$ corresponds to a globular compacted particle, $P_x = 3$ corresponds to a mostly globular particle with either significant interdomain flexibility or a significant unfolded character that impacts the surface-to-volume ratio, while $P_x = 2$ corresponds to a largely unfolded particle as in the case of an intrinsically disordered protein. Corresponding qualitative visual analysis can be applied to the Kratky plot. Calculation of the volume of the particle by this method enables another way of determining the molecular weight of the sample that can be compared to measurements derived from the Guinier approximation, as the MW-to-volume ratio is relatively fixed for globular proteins. It should be cautioned that this metric may be less reliable for proteins with a high degree of flexibility. As such, estimates of molecular weight by this metric should be limited to proteins for which a P_x parameter has been estimated that is as close to 4 as possible.

Data collection formats

The simplicity of SAXS data collection allows for versatility in experimental setup. At the simplest level, one takes a single sample and exposes it to an x-ray beam and repeats this process for several samples. As sample preparation requirements are usually minimal and

preliminary data analysis is relatively quick and can be easily automated, SAXS has the potential to be much higher throughput than other methods of structural analysis. Through implementation of an upstream automatic pipetting robot, the SIBYLS beamline has enabled collection of SAXS data from a 96-well plate format (81,84). In this high-throughput SAXS (HT-SAXS) experiment, a 96 well plate is provided containing an array of samples, from which the pipetting robot transfers the sample to the SAXS sample cell for data collection (Figure 2.2A).

As mentioned above, SAXS is particularly sensitive to aggregation and is unable to provide accurate structural information unless homogenous, monodisperse samples are provided. In many cases, small amounts of aggregation are present in protein samples that result in poor data-quality but are difficult to resolve. Another SAXS data collection format provides a powerful avenue for such samples by employing in-line Size exclusion chromatography directly upstream of data collection (SEC-SAXS) (80). In this format, the sample is injected onto an SEC column which is capable of separating out aggregated species or in the case of complex mixtures allows analysis of various subcomponents (Figure 2.2B). The eluent is then routed to a transparent sample cell from which SAXS measurements are performed as the sample flows off the column. An auxiliary line can be used in which MALS data collected concurrently as a secondary measure of sample quality and homogeneity. From this, the R_g can be measured rapidly for each point in the chromatogram to detect whether the sample is homogenous. SAXS data from a range of points for which the R_g is homogenous can then be averaged to produce a high signal-to-noise, high-quality SAXS dataset that is free from aggregation.

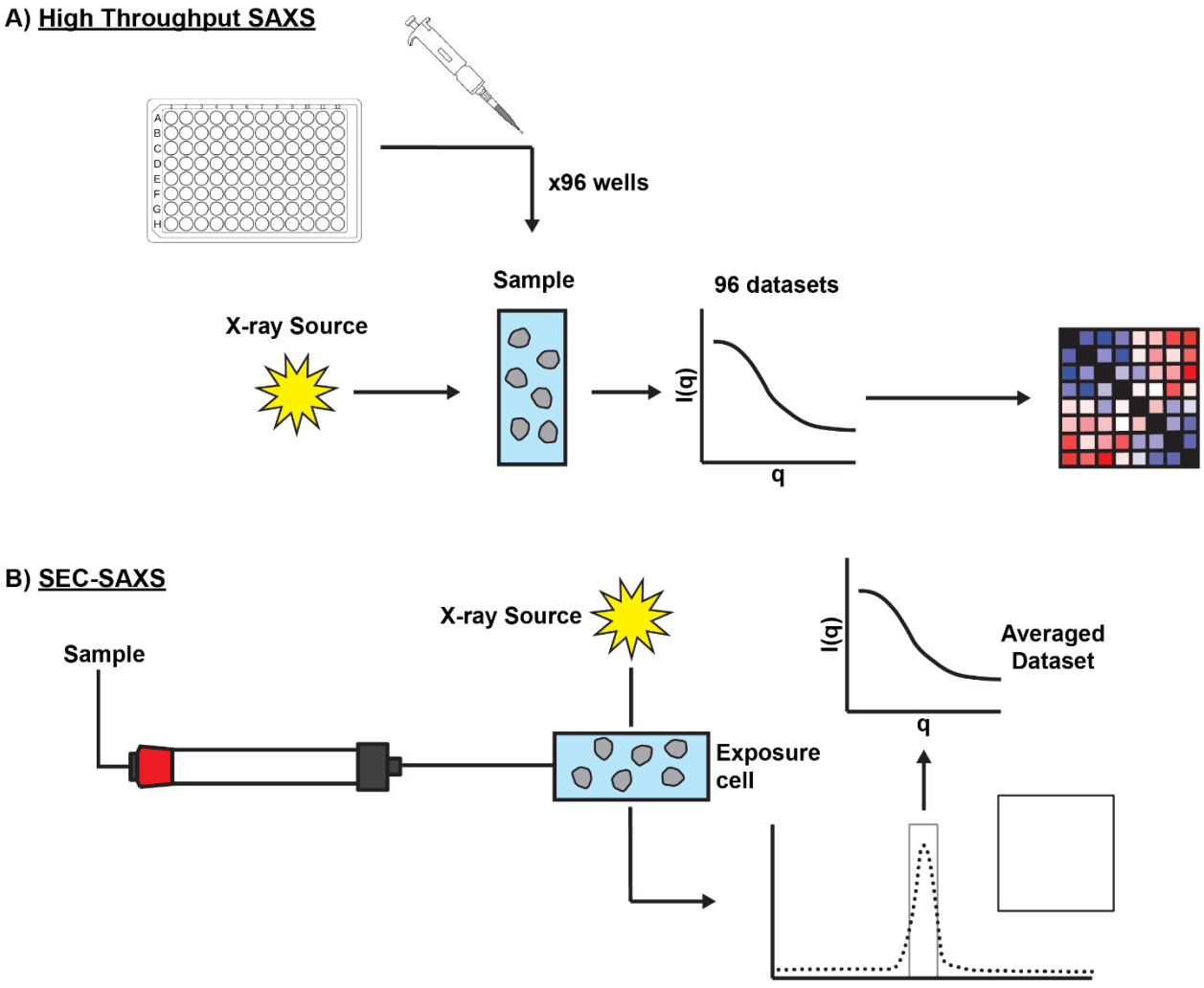

Figure 2.2 Two SAXS data collection formats used for this work.

A) In high-throughput SAXS (HT-SAXS) samples in a 96-well plate are provided, transferred to a sample-chamber by a pipetting robot, and successively collected. Data from individual samples can the be compared using a similarity plot. B) In SEC-SAXS, the sample is applied to a size-exclusion column which separates samples by size and removes aggregated species. Flow-through from the column is coupled to an exposure cell, where scattering data is collected over many frames (typically one exposure per second). Scattering data from individual frames can then be averaged across a homogenous peak.

Structural modeling using SAXS data

Combining lower-resolution SAXS data with modeling approaches can provide a powerful tool to analyze the structure of difficult proteins (flexible, low-affinity, heterogeneous)

for which other higher resolution methods are challenging. The most common form of structural modeling to be used in combination is Ab-initio structure determination. Several methods can be employed to produce a low-resolution (20-40 Å) molecular envelope of the macromolecule. Traditionally, ab-initio structural determination was performed using a bead-modeling approach such as DAMMIN/F. In this approach, dummy atoms are modeled in a simulated-annealing approach to find a configuration of atoms that fits the experimental data well (85). However, some ambiguity exists due to the one-dimensional nature of SAXS data in that divergent models can provide similarly good fits to the SAXS data. Thus, typically one creates multiple models and averages them to create a final molecular envelope. A more recent alternative approach employs an iterative structure factor retrieval algorithm that directly calculates an electron density map rather than creating a dummy atom model (86). This algorithm, known as DENSS, provides comparable molecular envelopes of comparable resolution to DAMMIN/F but provides an advantage in that differences in electron density can be visualized, especially useful for systems like hollow proteins, protein/DNA complexes etc. Importantly, due to the above-discussed surface-to-volume dependence of SAXS data, calculation of a molecular envelope is inadvisable for highly flexible proteins ($P_x < 3.0$).

Once a low-resolution molecular envelope has been produced, higher-resolution structural models from published structures or computational methods like AlphaFold can be placed into the density to gain an understanding of molecular architecture. Further insight can be obtained by directly comparing the high-resolution structural model to the experimental SAXS data rather than analyzing the fit of the structure into highly ambiguous low-resolution molecular envelopes. Back-calculation of a theoretical scattering curve is possible through methods like CRYSOL or FoXS, which both calculate an exclusion volume and hydration shell of the input

atomic model and perform an inverse-fourier transform to back-calculate the theoretical scattering curve of the model (87,88). This simulated model can then be compared directly to the experimental SAXS data through statistical methods such as a χ^2 analysis or calculation of more specialized statistical metrics. As SAXS data is highly reliant on size, it is very important that all missing atoms from published structures are added to the model before comparing directly to the experimental SAXS data. All loops and missing regions must be modeled, and accurate stoichiometry is essential for reliable comparison. This method provides a powerful tool for determining whether alternate conformations or configurations of a macromolecule are adopted in solution and evaluating the quality of structural models.

More sophisticated approaches to SAXS-based modeling can be employed that allow one to predict possible conformational ensembles (87,89). Typically, multi-state modeling employs a basic conformational sampling method that produces a large (>10,000) number of possible conformations of the macromolecule of interest. To speed up the conformational sampling step, one common approach is to provide user defined flexible residues to which the sampling step is restricted. This beads-on-a-string treatment is often an acceptable approximation for modular multidomain proteins. Following conformational sampling, the models produced are then subjected to SAXS-based scoring, allowing evaluation of how well each conformation fits the SAXS data. Finally, multi-state modeling allows combination and weighting of multiple models to represent an ensemble of conformations. Though powerful in providing an estimation of the theoretical range of conformations possible in solution, it is important to interpret these results with caution. First, this approach is susceptible to overfitting as the combination of multiple conformations can eventually provide a very well-scored ensemble model. However, as mentioned above, many possible combinations of data can yield similar scattering data so

individual models of the ensemble may or may not represent important conformations for the molecule. Likewise, it is important to recognize that for flexible proteins that occupy a continuous range of conformations rather than a series of discrete orientations, representation of the ensemble using a small number of orientations as representatives can be highly misleading.

Electron Microscopy

Electron Microscopy (EM) has long been a complementary technique used for the acquisition of low-resolution structural information of bio-macromolecules. Negative-staining using a heavy atom solution and or cryogenic EM in which the sample is frozen in vitreous ice are the two main methods of sample preparation. In the case of negative stain, a heavy atom staining solution such as Uranyl Acetate provides a high contrast fixation agent that facilitates imaging of otherwise low-contrast protein samples in a cost and time-efficient manner, though resolution is limited to 10-15 Å. Recent developments in cryoEM have greatly increased the achievable resolution of the technique to near atomic resolution and allow imaging of macromolecules in near-native conditions. Although cryoEM has seen an explosion in popularity in the past decade due to these technical developments, both techniques provide valuable information that can be integrated with other structural methods to gain insight into molecular structure biophysical properties.

EM basics

Although an in-depth discussion of the theoretical basis of electron microscopy is outside the scope of this dissertation, several key concepts must be introduced in order to apply this technique towards the investigation of a modular protein such as pol-prim. Similar to x-ray scattering techniques, EM relies on the interaction of a coherent beam of electrons with the

atoms of a particular sample of interest. Several differences between the physical properties of electrons and photons (x-rays) allow for acquisition of direct images of the sample rather than relying on diffraction. First, electrons interact much more strongly with matter, greatly increasing the signal-to-noise ratio of a given image compared to x-ray techniques. Importantly, electrons behave in a dual wave-particle manner similarly to photons and due to their charge can be manipulated or focused by strong electromagnetic fields in the form of lenses allowing for highly magnified images, similar to how glass lenses manipulate photons in a light microscope.

Electron microscopy of biological molecules typically employs Transmission Electron Microscope (TEM) (90). In TEM, an image is formed by illuminating the sample with an ordered bundle of electrons (an electron beam) (91). The electrons in this beam can then be scattered by interaction with the atoms present in a sample, either elastically or inelastically. Elastic scattering typically occurs in the case of an interaction between the incoming electron and the nucleus of an atom from the sample, in which case the energy (defined as the combined kinetic energy and momentum) of the incoming electron are not transferred to the sample atom due to the large difference in mass between the two particles. Inelastic scattering can occur when there is interaction between an electron from the beam and from the sample, in which the energy from the incoming atom is transferred to the sample and causes radiation damage. Importantly, the radiation damage caused that occurs during inelastic scattering events is a fundamental resolution limitation of electron microscopy, as fragile biological samples are quickly damaged before very high signal-to-noise images can be obtained. As elastically scattered electrons maintain coherent energy, they can be focused by lenses to form a high-resolution image, while inelastically scattered electrons are scattered and lose essentially random amounts of energy, resulting in an incoherent, difficult to focus beam. Scattering of an incoming electron increases as the number of

protons (Z) in an atom increases, with inelastic scatting increasing by $Z^{1/3}$ while elastic scattering increases by $Z^{4/3}$. Thus, heavier atoms provide a higher ratio of elastic scattering and therefore produce higher signal-to-noise images.

As an image is recorded by the detector of a microscope, the contrast and resolution are defined by the Contrast Transfer Function (CTF). At the most fundamental level, the CTF from a given image is defined by the following equation:

$$CTF = \sin\left(-\pi \Delta z \lambda k^2 + \frac{\pi C_s \lambda^3 k^4}{2}\right)$$

where Δz is the defocus of the image, λ is the wavelength of the electron as defined by the de Broglie equation and which is dependent on its accelerating voltage, C_S is the spherical aberration of the microscope, and k is the spatial frequency which is inversely related to the resolution. Thus, two fundamental and tunable parameters that affect contrast visible at a given resolution 1/k are the defocus and the accelerating voltage of the microscope. To simplify, higher accelerating voltages in more powerful microscopes (the Titan Krios in the Vanderbilt EM core is a 300 kEV microscope while the Glacios and TF20 microscopes are 200 kEV) offer the potential for higher resolution information, while possibly sacrificing contrast of lower-resolution features. Similarly, higher defocus values offer higher contrast of lower-resolution features while sacrificing contrast at higher resolutions.

The CTF is directly related to the power spectrum, which is a Fourier transformation of the original image. However, the power spectrum is also impacted by several phenomena that ultimately drastically reduce contrast at higher resolutions, collectively referred to as the envelope functions. The envelope function is the product of several factors: image quality including the effects from beam-induced motion and aberrations in the image/microscope, the

precision of the alignment, and for single-particle reconstruction the particle homogeneity and number of particles averaged.

Further limitations for biological TEM are related to the nature of atomic makeup of proteins and nucleic acids and their relative similarity in electron density to the water-based solution they are usually in. At the simplest level, this similarity in electron density results in signal for the molecule that is difficult to distinguish from the background solution. Additionally, as carbon, nitrogen, and oxygen are all relatively low atomic number, the ratio of elastic scattering to inelastic scattering is low, thus increasing dose to increase the overall signal results in a corresponding increase in inelastic scattering and correspondingly radiation damage. This damage destroys the sample at the atomic level, perturbing the chemical nature of the sample of interest from a native state. In order to overcome this low signal-to-noise ratio, several strategies are available.

Two sample-preparation strategies offer alternative routes to either limit radiation damage of the sample or increase the signal-to-noise of the sample. The first is cryogenic EM (cryoEM) in which the sample is plunged into liquid ethane and flash-frozen in vitreous ice (91–94). In this technique, native structure is mostly preserved as ice crystals that might damage the sample are avoided by the rapid rate of freezing. Additionally, the cryogenic temperatures reduce the effects of radiation damage, allowing for higher doses and increased signal-to-noise. Another sample preparation technique is to coat the sample of interest with a heavy atom stain such as uranyl formate (90,95,96). This process is called negative stain EM as the heavier atoms surround the protein and act as a dark background field, reflecting more electrons than the lower atomic number protein atoms which appear white in the micrograph as the beam more easily passes through them. This approach provides drastically improved contrast and preserves the sample,

protecting it from radiation damage. However, the grain size of the stain limits the overall achievable resolution from negative stain EM to roughly ~15-20 Å. Additionally, the harsh staining conditions can result in perturbations of the native state of the protein, sometimes resulting in complete collapse of the particle. Thus, care must be taken to verify that the structure obtained by negative stain represents an intact sample.

In addition to sample preparation methods, single-particle averaging is a strategy in which many two-dimensional projection images of a sample of interest are computationally aligned. The alignment of many thousands of particles from similar viewing angles greatly increases the signal-to-noise ratio over that of a single image of a particle. Mathematically, the Fourier transform of a 2D projection image corresponds to a center slice through the 3D reciprocal space representation of a 3D object. Thus, through the acquisition of many 2D projection images, a 3D volume can be reconstructed in reciprocal space, after which an inverse Fourier transform results in a 3D volume corresponding to the electron density of the particle.

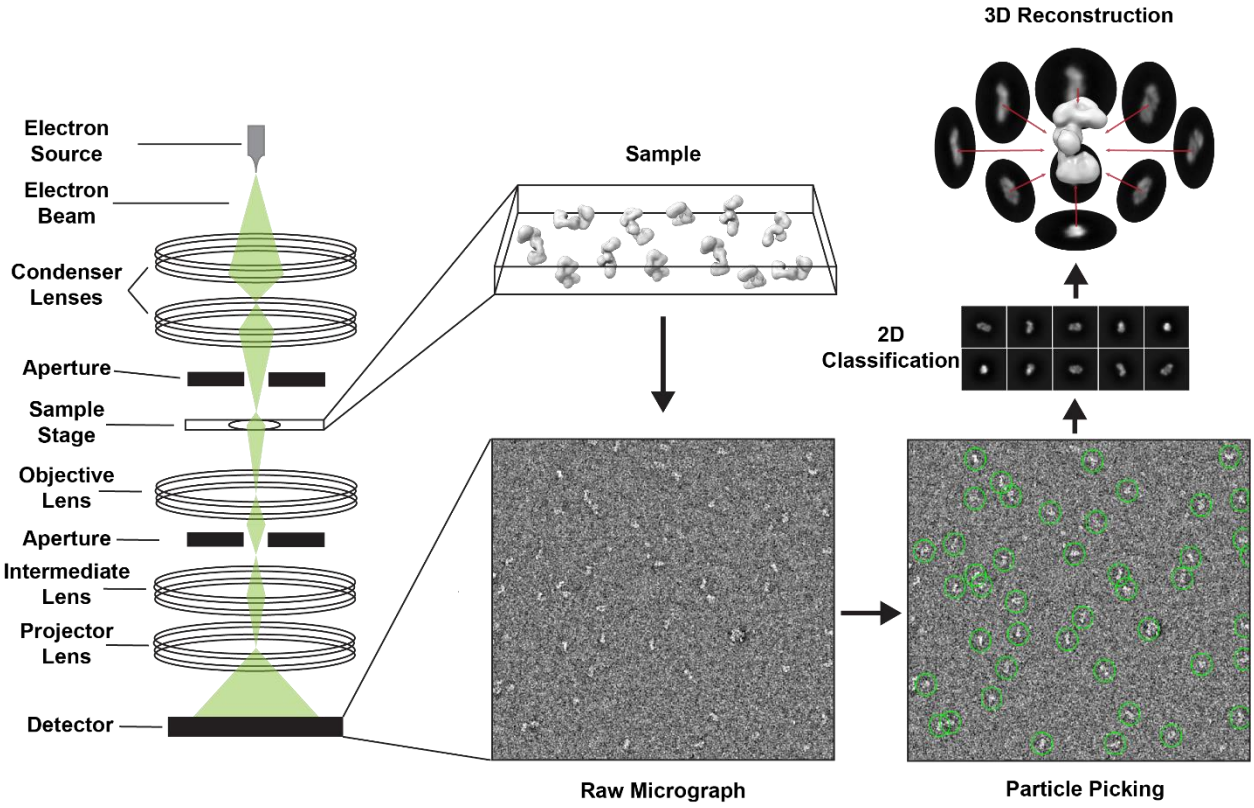

Figure 2.3 Schematic of Electron Microscope and basic data processing.

Left, simplified schematic of a transmission electron microscope. Right, schematic of data collection and initial processing.

When performing single-particle analysis, sample homogeneity is essential for alignment of individual particle images and achieving higher resolution. As such, it is essential that compositional heterogeneity is limited during the sample preparation stage. Additionally, in simple 3D reconstruction algorithms, protein flexibility can present a major barrier to achieving high-resolution, as components for which there is a large degree of flexibility can cause a blurring effect during reconstruction. Several methods exist for sorting, characterizing, and ultimately overcoming the limitations presented by protein flexibility (97–100). In addition to achieving a high-resolution reconstruction from a flexible sample, the flexibility itself can be described in high-resolution detail, and molecular movies can be created augmenting static 3D

reconstructions. However, these methods are still currently undergoing rapid development, and at the moment are not capable of handling all cases of protein flexibility.

CHAPTER 3. FLEXIBILITY AND DISTRIBUTIVE SYNTHESIS REGULATE RNA PRIMING AND HANDOFF IN HUMAN DNA POLYMERASE A-PRIMASE[2]

Introduction

In eukaryotes, DNA polymerase α-primase (pol-prim) initiates DNA synthesis during replication, generating the first ~30 nucleotides of nascent strands.(1,101–104) Pol-prim is unique among replicative polymerases in its ability to perform de novo synthesis from a single-stranded DNA template; these primers are required for further synthesis by the processive polymerases ε and δ that perform the bulk of nascent strand synthesis(2,5). Pol-prim plays a particularly crucial role on the lagging strand during replication, as primers must be repeatedly synthesized due to the discontinuous nature of synthesis of this nascent strand (105) . The primers generated by pol-prim are chimeric in nature, consisting of 7-10 ribonucleotides followed by ~20 deoxyribonucleotides.(35,66) The RNA and DNA portions of this chimeric primer are generated by distinct active sites located in the primase and polymerase subunits of this tetrameric enzyme (Figure 3.1) (71).

Primase consists of the PRIM1 (also known as p48, p49, PriS, and Pri1) catalytic subunit and the PRIM2 (a.k.a. p58, PriL, and Pri2) regulatory subunit (106). PRIM2 contains PRIM2N (residues 1-245), a domain that acts as a scaffold linking the primase subunits to the pol α subunits, and the flexibly attached PRIM2C domain (residues 265-455), which is involved in regulation of both RNA and DNA synthesis (46,48,50,52,56,63,107,108). Although the catalytic site for RNA synthesis is located within PRIM1, PRIM2C greatly stimulates catalytic activity and is required for both initiation of primer synthesis and regulation of the length of the primer

[2] This work has been submitted for publication as a research article in the Journal of Molecular Biology.

(37,55,57). Polymerase α consists of the catalytic subunit POLA1 (a.k.a. p180) and the regulatory subunit POLA2 (a.k.a. p68, p70, and B subunit) (61,64). POLA1 contains a catalytic domain (POLA1cat) and a C-terminal domain (POLA1C, residues 1265-1462) that acts as a scaffold between POLA2 and PRIM2N, while also connecting to POLA1cat through a flexible linker (39,63,109). Architecturally, pol-prim is organized into three modules: the tetramer core (POLA2/POLA1C/PRIM2N/PRIM1; a.k.a. pol-prim platform), POLA1cat, and PRIM2C (43,58,110). POLA1cat and PRIM2C are tethered to the tetramer core through ~30 and ~20 residue flexible linkers, respectively. The modularity of pol-prim allows for dynamic coordination of RNA and DNA primer catalysis as well as interfacing with other replication factors (13,23,26,31,111–113).

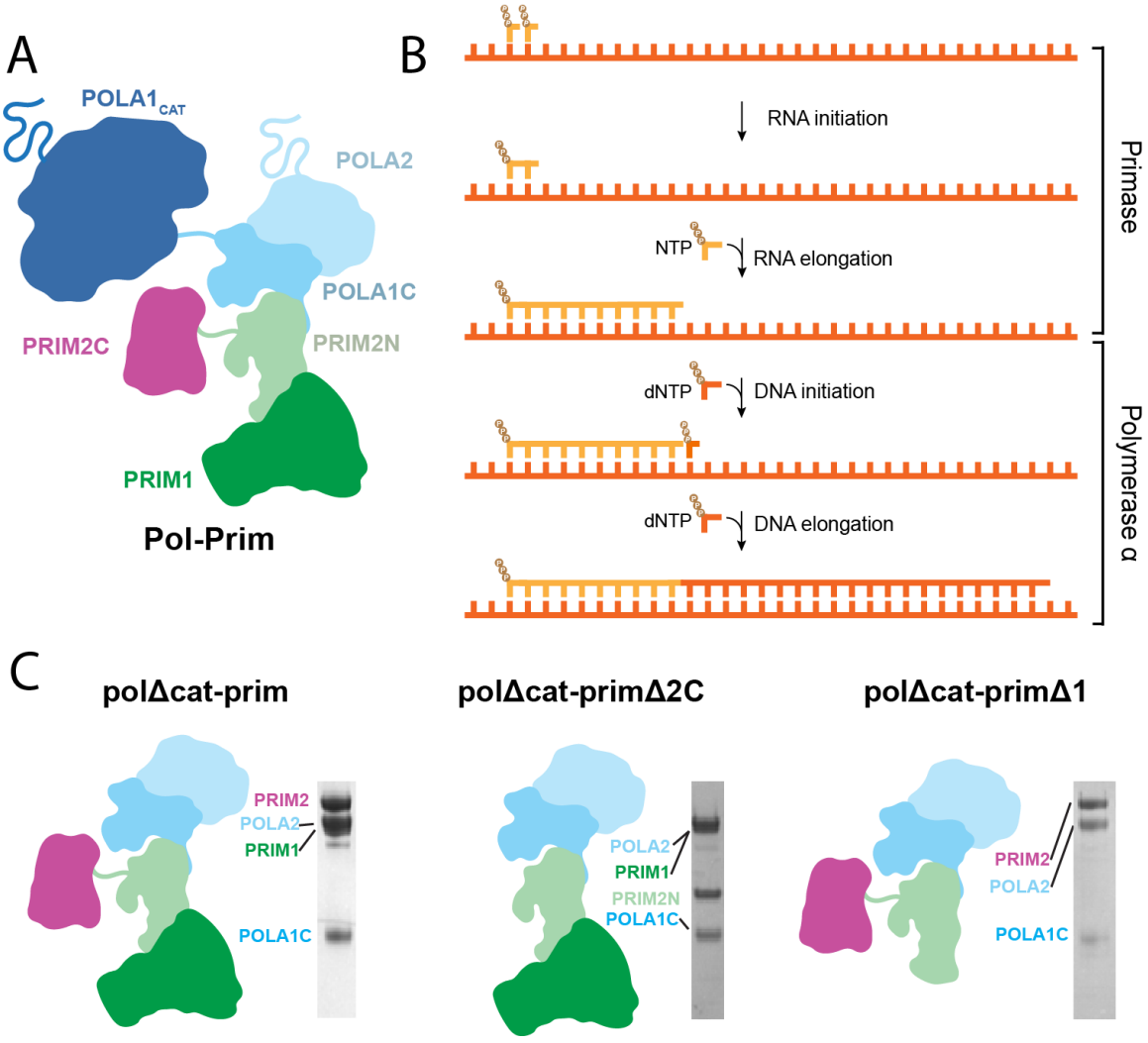

Figure 3.1 Schematic diagrams of primer synthesis and pol-prim structure.

A) The pol-prim heterotetramer. B) Steps of primer synthesis. C Diagrams of pol-prim and subcomplexes used in this study alongside SDS-PAGE images of the purified proteins.

RNA primer synthesis involves the coordinated action of PRIM1 and PRIM2C. PRIM1 alone possesses an ability to elongate pre-synthesized RNA primers in the absence of PRIM2C, but is otherwise incapable of initiating primers and has low affinity for oligonucleotide substrates on its own (66–68). PRIM2C has an inherently higher affinity for substrates, in particular for structures mimicking RNA primed templates that contain a 5' triphosphate and 3' template overhang.(58,67,68) It is generally understood that initiation of primer synthesis requires a

configuration in which PRIM1 and PRIM2C are aligned and together bind the ssDNA template, along with two incoming NTPs and catalytic di-cations, before synthesis of the initial dinucleotide can occur.(46,47,69,75,114) The need to simultaneously align the domains in combination with the relatively low binding affinities for substrate and co-factors makes formation of the first dinucleotide the rate-limiting step of primer synthesis.(66)

Once the dinucleotide is formed, the RNA primer is elongated by PRIM1 in a 5'-3' direction to a length of 7-10 nucleotides. However, PRIM1 lacks the ability to regulate primer length without PRIM2C.(37,68) Additionally, PRIM1 synthesis is distributive in nature, as synthesis of primers less than the unit length of 7 is common and is increased in the absence of PRIM2C.(66,68,69) PRIM2C not only has higher affinity than PRIM1 for the substrate, but is also remains bound to the 5' end of the primer, which plays an essential role in primer length regulation.(68,70,115) From this it can be inferred that PRIM2C and PRIM1 become spatially separated as the primer length grows. Recent studies by our groups and others have also demonstrated that PRIM2C forms an interaction with POLA1cat during DNA initiation and likewise remains bound during DNA elongation, playing a critical role in handoff and DNA primer length regulation.(42,70,76)

All current models of RNA primer synthesis invoke a critical role for PRIM2C. A recently proposed model posits that a steric clash between PRIM2C and PRIM2N results when the RNA primer reaches a length of 9 or 10 nucleotides, providing an upper limit for primer length.20,28,30 In addition, our recent work has shown that steric clash with PRIM2C prevents POLA1cat from binding an RNA primer of <7 nucleotides, providing the lower bound of unit primer length as 7.(42)

Despite advances in understanding that PRIM2C regulates primer synthesis, direct structural characterization of the priming complex during RNA catalysis is lacking. Here we report new data and present a mechanistic model for the initiation, elongation, and handoff of the RNA primer. By comparing binding affinities for RNA-primed template substrates of a variety of pol-prim constructs, we directly show that the majority of the substrate binding affinity of primase derives from the tight binding of PRIM2C. We also characterize structural transitions that occur as the RNA primer is initiated and elongated showing there is a mixture of states in all cases and that in the presence of substrate, the majority of molecules have PRIM2C but not PRIM1 bound. Our results reveal the molecular basis for the highly distributive nature of RNA primer synthesis and lead to a revised model for RNA primer synthesis that incorporates the high affinity of PRIM2C for substrate, the high-degree of flexibility of PRIM2C, and the preferential binding of RNA-primed templates by POLA1cat.

Methods

Cloning and Expression

The gene encoding FL-PRIM1 was cloned into a pBG100 expression vector while the gene encoding FL-PRIM2 was cloned into a pETduet co-expression vector. POLA2 (residues 155-598) and POLA1C (residues 1265-1462) were cloned into a pET-duet expression vector encoding a 6xHis TEV cleavable tag. The disordered N-terminal region (residues 1-155) was truncated to facilitate expression, purification, and biophysical analysis. To generate a PRIM2ΔC construct, the N-terminal domain PRIM2N (residues 1-265) was subcloned into a modified pETduet expression vector. To generate a polΔcat-primΔ1 construct, a separate tri-gene construct

was generated which encoded POLA2 (residues 155-598), POLA1C (residues 1265-1462), and FL-PRIM2, which was then cloned into a pRSF-DUET vector.

PRIM1/PRIM2 and POLA1C/POLA2 were expressed separately in Escherichia coli BL21 RIL cells. PRIM1/PRIM2 cultures were grown at 37°C in LB medium supplemented with 25 mg/L chloramphenicol, 100 mg/L ampilicin, and 30mg/L kanamycin while POLA1C/POLA2 cultures were supplemented with chloramphenicol and ampicillin only. Upon reaching mid-log phase, cultures were cooled to 16 °C and PRIM1/PRIM2 cultures were supplemented with 100 mg/L ammonium iron (III) citrate and 100 mg/L iron (II) sulfate. Overexpression was then induced by addition of IPTG to a final concentration of 0.5 mM. After overnight incubation, cells were harvested by centrifugation and pellets stored at -80 °C.

Protein Purification

PRIM1/PRIM2 and POLA1C/POLA2 cell pellets were thawed in room temperature water and resuspended in Lysis Buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, 10% [v/v] glycerol, 1 mM TCEP, 0.1 g/50mL DNAse, 1 g/50mL lysoszyme, and 1 complete EDTA-free protease inhibitor cocktail tablet [Roche] per 50 mL). The pellets were co-lysed using a Dounce homegenizer followed by gentle sonication. The lysate was then centrifuged at 50,000 g for 1 hr to pellet cell debris. The supernatant was then filtered using a 0.45 micron filter before application to a Ni-NTA column on a FPLC (Biorad). The column was washed with 10 column volumes (CV) of 10 % Nickel Buffer A (Lysis buffer without DNAse, lysozyme, and protease inhibitor) 90% Nickel Buffer B (Nickel Buffer A + 1M imidazole), followed by protein elution at 60% Nickel Buffer B. Fractions containing the protein of interest were pooled, H3C and TEV protease were added for His-tag cleavage, and dialyzed overnight at 4 °C in dialysis buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol, 1 mM TCEP) using 10 kDA MWCO dialysis

tubing. The following morning, the dialyzed solution was diluted 1:1 using Dilution buffer (20 mM HEPES pH 7.5, 5% glycerol, 1 mM TCEP) and filtered using a 0.45 micron filter before application to a Heparin column pre-equilibrated in a mixture of 5% Heparin Buffer 1 (20 mM HEPES pH 7.5, 5% glycerol, 1 mM TCEP) and 95% Heparin Buffer 2 (20 mM HEPES pH 7.5, 1M NaCl, 5% glycerol, 1 mM TCEP). The column was then washed with 5 CV of this mixture of Heparin Buffer 1 and 2 followed by a gradient over 12 CV from 5% to 100% Heparin Buffer 2. Fractions containing polΔcat-prim were then pooled and concentrated using a 10 kDa MWCO spin filter, centrifuged, and applied to a S200 Increase column (Cytiva) equilibrated in SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, and 1 mM TCEP). Protein was eluted in 1.2 CV of this buffer, and fractions containing PolΔcat-prim were pooled and concentrated. Aliquots were then used immediately or flash frozen in liquid nitrogen and stored at -80 °C. PolΔcat-primΔ2C was purified in the same manner while the polΔcat-primΔ1 construct lacking PRIM1 was purified from a single cell pellet.

Sample Preparation and Crosslinking

5'-triphosphorylated RNA substrates were synthesized using as described previously.(42) DNA oligonucleotides were purchased from Integrated DNA technologies and resuspended in MES buffer (10 mM MES pH 6.5, 40 mM NaCl). To generate RNA-primed templates, 66 uM RNA primer and 60 uM DNA templates were annealed in MES buffer by heating to 85 °C then slowly cooled to 4 °C. Protein substrate complexes for SAXS and EM experiments were prepared by incubating protein and primed substrate at a 1:1.1 ratio in binding buffer for 30 min. For crosslinking, bis(sulfosuccinimidyl)suberate (BS3) was purchased from Thermo fisher. A 10 mM stock solution was prepared immediately before use. Crosslinking reactions were performed using 2 mM BS3, 5 uM protein, and 5.5 uM substrate in SEC buffer, incubating at room

temperature in the dark for 30 minutes. The reaction was quenched using 50 mM TRIS buffer. The reaction mixture was then subjected to Size-Exclusion Chromatography to remove higher order oligomeric species. Production of crosslinked products was monitored by SDS-PAGE.

Table 3.1 Oligonucleotides.

Oligouncleotide	**Sequence**
Primer – 8mer	pppGGAUACUG
Primer – 9mer	pppGGAUACUGC
Template – 29	GTATGTATGTCAGTATCCTGTATGTATGA
Template - 30	GTATGTATGTGCAGTATCCTGTATGTATGA
Template - 21	GTATGTATGTCCTGTATGTATG

Microscale Thermophoresis

RNA/DNA binding reactions were conducted in the dark in binding buffer [20 mM HEPES•NaOH (pH 7.5), 150 mM NaCl, 5 mM MgCl2, 1 mM TCEP, 1mM AMPCPP and 0.05% (vol/vol) Tween-20] with 10 nM Cy5-labeled TriP 9mer primer- 31mer template substrates. A 9mer RNA primer was used in lieu of 8mer to maximize retention of the primer on the template at the low concentration and near room temperature conditions in which the experiments were performed. Serial dilutions (1:1) of protein (2.5 μM to 0.61 nM) were prepared in binding buffer and combined with an equal volume of 20 nM Cy5-labeled primer-template. Samples were mixed and incubated on ice for 10 min, then at 17 °C for 10 min, and centrifuged at 20k ×g at 4 °C for 5 min prior to loading into standard-coated capillary tubes (NanoTemper, GmbH, Munich, Germany). Microscale thermophoresis data were collected on a Monolith NT.115 (NanoTemper) instrument. All experiments were conducted using the red fluorescence laser at 60% power

($\lambda_{excitation}$: 600 – 640 nm; $\lambda_{emission}$: 660 – 720 nm) and the infrared laser at medium power (λ = 1480 nm). Temperature-related intensity changes (TRIC) were performed at 17 °C with a pre-IR phase of 3 sec, followed by an IR-on phase of 20 sec, and a post-IR phase of 1 sec. Data quality was tracked by monitoring fluorescence quenching, photobleaching, protein aggregation, and adsorption for each experiment. Experiments were performed in triplicate. For each protein-DNA binding titration, the normalized fluorescence values were tracked as parts per thousand using the equation:

$$F_{norm} = (\frac{F_{hot}}{F_{cold}}) \times 1000$$

where F_{hot} corresponds to the mean fluorescence intensity from 0.5-1.5 seconds and F_{cold} to the mean fluorescence intensity from -1.0 – 0.0 seconds before irradiation and converted to ΔFnorm:

$$\Delta F_{norm} = F_{norm}(n) - F_{norm}(initial)$$

Dissociation constants were derived from fits to the data using a Law of Mass Action binding model:

$$f(C) = U + \frac{(B-U)\times(C-C_{target}+K_d-\sqrt{(C+C_{target}+K_d)^2-4(C\times C_{target})}}{2\times C_{target}}$$

where f(C) is the fraction bound at a given protein concentration C, C_{target} is the concentration of fluorescently labeled oligonucleotide, B is the F_{norm} signal of the complex, U is the F_{norm} signal of the protein alone, and K_d is the dissociation constant.

Small Angle X-ray Scattering

Scattering data were collected at the ALS beamline 12.3.1 LBNL Berkeley, California. The X-ray wavelength λ was 1.03 Å and the sample-to-detector distance was set to 1.5 m

resulting in scattering vectors, q, ranging from 0.01 to 0.5 $Å^{-1}$.62 The scattering vector is defined as $q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle. For SEC-SAXS experiments, 60 µl of sample (25 uM) was loaded onto a Shodex 802.5 SEC column connected to a multi-angle light scattering system and the protein complex was eluted in SEC buffer.63 The eluent was split between SEC-MALS and SAXS channels. 3.0-s X-ray exposures were collected continuously during a ∼35 min elution. The SAXS frames recorded prior to the protein elution peak were used as buffer blanks to subtract from all other frames. The subtracted frames were examined by radius of gyration (Rg) and scattering intensity at q = 0 Å−1 (I(0)), derived using the Guinier approximation $I(q) = I(0)e^{-qRg2/3}$ with the limits q*Rg <1.5. I(0) and Rg values were compared for each collected SAXS curve across the entire elution peak. The elution peak was mapped by plotting the scattering intensity at q = 0Å−1 (I(0)), relative to the recorded frame. Uniform Rg values across an elution peak represent a homogenous assembly. Data quality was assessed by: (i) fitting of the data in the Guinier region; (ii) good agreement between Rg calculated by the Guinier method and that based on P(r) distribution; (iii) molecular weight estimates derived from the data (Porod volumes) that were consistent with the expected size of the protein.64 The SAXS data were processed using Scatter 4.0 and PRIMUS.65,66 Back-calculated scattering data from coordinates extracted from structures available in the PDB were generated using FoXS.52,67 High-Throughput SAXS was performed using the same buffer and protein concentrations as SEC-SAXS.68 An array of ssDNA (0.2-2 mM) and GMPCPP (0.1-10 mM) concentrations were screened and monitored for major changes in the Mid-Q range of the Kratky plot that is highly sensitive to changes in shape. The concentration with the highest magnitude shift in the Mid-Q region without evidence of aggregation was further processed and compared to equivalent data collected for the free protein and a back-calculated scattering curve

determined from a theoretical model of the configuration of PRIM2C that is primed for initiation (see Modeling section).

EM sample preparation and data collection

Negative stain EM samples were prepared by diluting freshly purified polΔcat-prim to a final concentration of 50 nM in SEC buffer. 400 mesh carbon film copper grids (Electron Microscopy Sciences) were glow discharged for 2 minutes for non-crosslinked samples and 3 minutes for crosslinked samples. 2.5 uL of sample was applied to the grid and incubated for 1 minute, then blotted using filter paper, washed for 5 seconds in DI water, blotted, washed for 5 seconds in Uranyl formate stain, blotted, then stained for 90 seconds in Uranyl Formate solution. Grids were screened using a Morgagni 100 kV screening microscope. Micrographs were collected using a FEI 200 kV TF20 microscope.

EM data processing

Initial data processing was performed in Sphire 1.9. Particle picking was performed with Cryolo yielding 80,000 – 100,000 particles per dataset.69 2D classification was then performed in sphire using Iterative stable alignment clustering (ISAC).70 Particles were imported into cryosparc for further processing.71 Multi-class ab initio reconstruction and heterogeneous refinement was performed, and particles from classes displaying incomplete or irregular density were removed before submitting to further rounds of multi-class ab initio reconstruction and heterogeneous refinement (Figure S6).

Modeling

Models of the tetramer core were generated by extracting coordinates from the PDB entries of eukaryotic pol-prim (5EXR, 8DOB, 7UY8, 7U5C, 8D9D, 8G9F) and rigid-body

docked models from EMD entries (29888, -29889, -29891).(42,58,75–77) Only residues from POLA2, POLA1C, PRIM2N, and PRIM1 were kept, and structures were aligned to POLA2/POLA1C for visualization. Two-state SAXS models were generated in Multi-FoXS.(87)

Models of polΔcat-prim were generated by extracting coordinates from the PDB entries of pol-prim in which PRIM2C was visible (PDB:5EXR, 7U5C, 8G9F, 8FOC, 8FOD, 8FOE, 8FOH, and 8D0K).28,43,45,46,47 Only residues from POLA2, POLA1C, PRIM2, and PRIM1 were kept and structures were aligned to PRIM1. For direct comparison to SAXS data, missing residues were added using MODELLER.72,73

BILBOMD was used to generate additional single and multiple configuration models of polΔcat-prim that are consistent with the experimental SAXS data.74 For these calculations, residues 245-265 and 455-509 in PRIM2 were defined as flexible.

Models of the RNA initiation and elongation configurations of polΔcat-prim were prepared by manually aligning PRIM1 from PDB:6R5D to PRIM1 to the tetramer core of PDB:5EXR.54,75 The structure of PRIM2C bound to substrate from PDB:8D9D was manually aligned such that the 3’ terminus of the RNA primer was aligned with the ATP molecule from PDB:6R5D for nucleophilic attack.(76) All but 8 nucleotides for elongation or all but 1 nucleotide for initiation were retained from the structure of the RNA-primed template bound to PRIM2C. 75

Models were initially aligned to EM density using the Fit in Map tool of ChimeraX to fit the teteramer core. Subsequently, atomic models were manually adjusted by treating the PRIM2C:PRIM1:substrate as an independent module and PRIM2N:POLA1C:POLA2 as another

independent module on the basis of the large degree of conformational heterogeneity observed between the two in our SAXS and modeling studies.

Results

Analysis of RNA priming by pol-prim requires specific construct design

The primase subunits of pol-prim (PRIM1, PRIM2) are responsible for the synthesis of the initial 7-10 nucleotide RNA primer. Moreover, the large POLA1cat domain, flexibly linked to the central tetramer core, is not directly involved in RNA primer synthesis. To focus our analysis of RNA priming we therefore utilized a truncated pol-prim subcomplex that lacks POLA1cat (PRIM1/PRIM2/POLA1C/POLA2, polΔcat-prim, Figure 3.1C), similar to that used for previous studies.(42,58,63,68) The truncation of POLA1cat is advantageous for several reasons. Among these, the flexible nature and large size of POLA1cat would complicate measurements probing the conformational dynamics of PRIM1 and PRIM2C, which are the primary focus of this study. Deletion of POLA1cat is also valuable because it binds RNA-primed templates with higher affinity than PRIM1, so its absence ensures efficient generation of complexes relevant to RNA primer catalysis upon addition of relevant substrates.(42,59,60,67,68) An additional important factor is that in the absence of nucleotide substrates, full length pol-prim exists primarily in a compact configuration (termed the Autoinhibitory conformation) in which POLA1cat forms physical interactions with and restricts the range of motion of both PRIM2C and the tetramer core.(42,58,70,77) Pol-prim must drastically rearrange from this configuration before primer synthesis can occur as both the substrate binding site of PRIM2C and the DNA catalytic site of POLA1cat are occluded. Thus, removal of POLA1cat allows the analysis of the inherent conformational flexibility within the tetramer core as well as flexibility of PRIM2C relative to PRIM1.

Recent structures of pol-prim in complex with the CST complex critical for telomere processing have established that CST acts as a scaffold for pol-prim that also restricts the flexibility and range of motion of the tetramer core and PRIM2C.(75,77,78) In this study, we did not use either CST or other binding partners of pol-prim such as RPA that may influence the structure and dynamics of pol-prim because our goal was to establish the inherent degree of flexibility in polΔcat-prim. Our data also inform on the extent to which scaffolding may alter flexibility and how this may contribute to the mechanism of primer synthesis by pol-prim.

The pol-prim tetramer core is conformationally heterogeneous

The flexible linkers connecting POLA1cat and PRIM2C to the tetramer core are critical to the structural changes in pol-prim that are required to initiate and synthesize the chimeric RNA-DNA primers.(43,58) For RNA priming, the positioning of PRIM2C with respect to the tetramer core is key. The tetramer core has traditionally been viewed as a stable globular platform on which the primer is assembled.(58,110) However, multiple structures of pol-prim that have emerged in the past few years reveal a significant degree of conformational heterogeneity within the core.(42,58,70,75–78) Superposition of the tetramer core from these structures suggests there are hinges at the POLA1C/PRIM2N interface and within PRIM2N itself. Upon alignment to POLA1C and POLA2, which together form a rigid body,(61) PRIM1 is seen to occupy a range of orientations that differ from each other by up to 35°, corresponding to a shift of ~40 Å with respect to the center of PRIM1 (Fig 3.2A). Hence, in order to characterize the structural dynamics of PRIM2C relative to PRIM1, it is important to understand the conformational heterogeneity within the tetramer core.

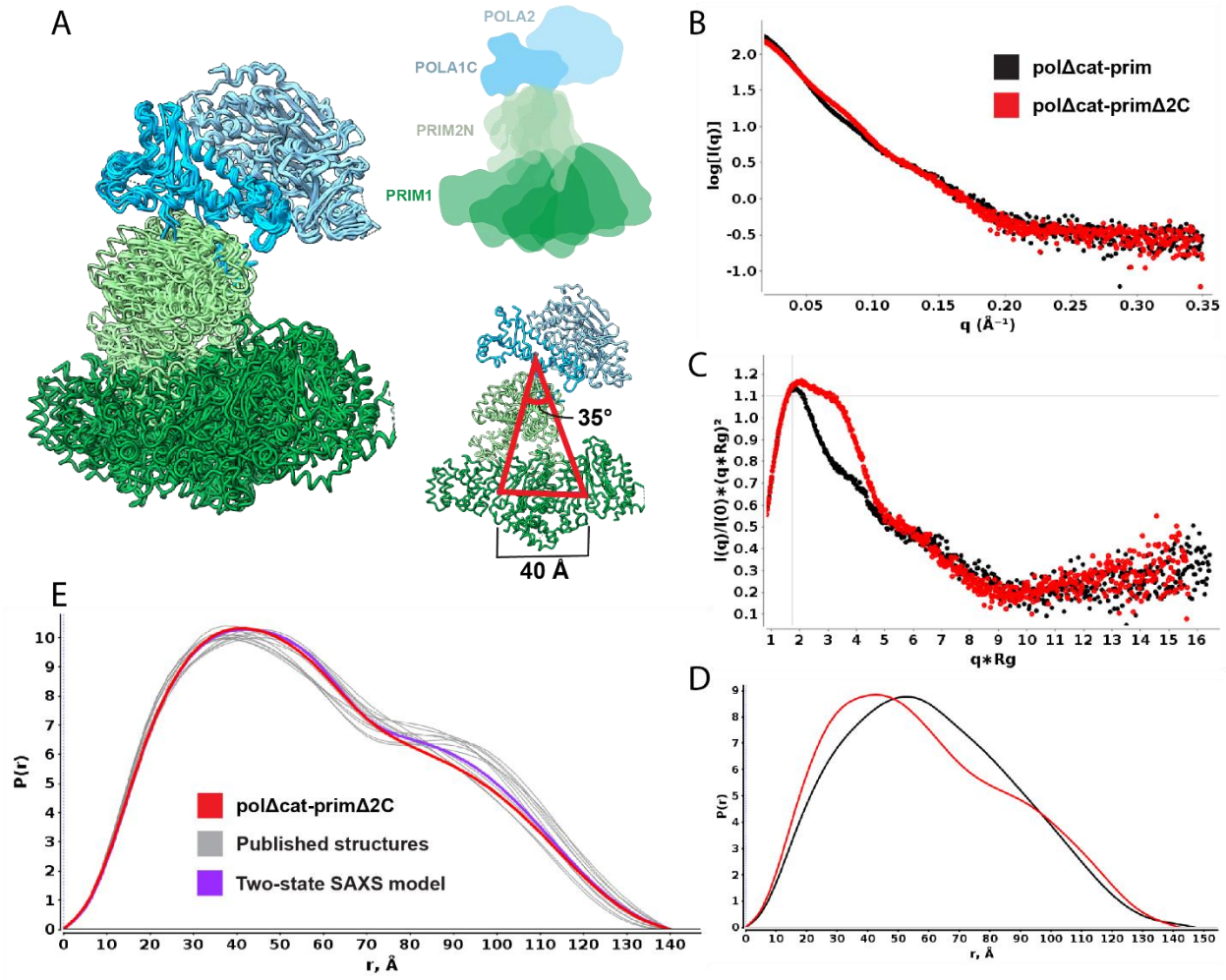

Figure 3.2 Flexibility of tetramer core in solution.

A) Extracts from structures containing the tetramer core aligned to POLA1C/POLA2/PRIM2N (light blue), showing range of motion in PRIM2N/PRIM1 (green). Models are aligned to POLA1C/ POLA2. Comparison of SAXS data for free polΔcat-prim (black) and polΔcat-primΔ2C (red): B) log10 intensity plot; C) Normalized Kratky plot; D) Distance distribution. E) P(r) distribution of polΔcat-primΔ2C (red) overlaid with back-calculated distance distributions from coordinates extracted from available structures (gray) and two-state multi-FoXS model (purple).

To directly characterize the degree of flexibility within the tetramer core, we turned to small angle X-ray scattering (SAXS) analysis of a construct in which both POLA1cat and PRIM2C were removed so that just the tetramer core remains (polΔcat-primΔ2C, Figure 3.1C). This construct proved to be highly amenable to SAXS analysis and high quality, aggregation-free scattering data were acquired (Figure 2, Table 1). The Kratky plot of the SAXS data (Figure 3.2C) is characteristic of a multi-domain globular species and the shape of the corresponding P(r) distribution (Figure 3.2D) indicates the particle is dumbbell shaped, as expected based on the available pol-prim structures. The estimated Porod exponent (P_X) was ~3.9, which indicates polΔcat-primΔ2C forms a globular particle with limited flexible regions.

To determine if there is a correspondence between the extent of heterogeneity in the tetramer core in published pol-prim structures and the apparent lack of flexibility of polΔcat-primΔ2C in solution, we extracted the coordinates for the tetramer core from the ensemble of structures and back-calculated the corresponding SAXS curves (Figure 3.2A, E). The fits to the experimental data of single back-calculated scattering curves produced chi-squared values ranging from 4.9 to 11.4, indicating that each of these structures individually are poor representations of the structure in solution. However, multi-state modeling of the SAXS data using Multi-FoXS with the ensemble of structures produced much better fits.(87) Optimization by using weighted mixtures of multiple input structures produced excellent fits even with the simplest two-state model. Interestingly, the top scoring pair (chi-squared fit of 1.8) (Figure 3.2E) corresponded to the two extremes of the ensemble of structures. This analysis suggests there is substantial conformational heterogeneity within the teteramer core and that it likely samples a continuous mixture of conformations in solution, consistent with our previous cryoEM analysis of pol-prim.(42)

If there is substantial conformational heterogeneity within the tetramer core, how can it be reconciled with SAXS analysis of polΔcat-primΔ2C indicating there is only limited flexibility? The explanation for this apparent discrepancy can be found by considering the physical basis of the P_x and other SAXS parameters reflecting flexibility. For example, the P_x, which ranges from 2.0 (completely disordered) to 4.0 (rigid), varies as a function of the surface area to volume ratio versus scattering intensity.(83) Importantly, the conformational heterogeneity within the tetramer core does not significantly alter the globularity of the protein or the surface area to volume ratio. Thus, the high P_x value for polΔcat-primΔ2C provides a valuable insight, revealing that the ensemble of tetramer core conformations observed in the various pol-prim structures represents well the conformational heterogeneity of the tetramer core in solution.

PRIM2C is a primary contributor to pol-prim configurational flexibility

Although it has long been established that PRIM2C is attached to the tetramer core of pol-prim by a flexible linker,(46,63,110) the nature and degree of variation in its relative orientation with respect to the core has not been fully characterized. Importantly, variation in PRIM2C orientation has been inferred to play a critical role in facilitating primer synthesis and regulating the length of both the RNA and DNA portions of the primer.(37,42,70,75,76,125,126) However, among the eleven available pol-prim structures in which PRIM2C is visible, eight occupy a configuration roughly similar to the autoinhibited state, in which PRIM2C is located near the PRIM2N and POLA1C interface and constrained by direct contacts with POLA1cat.(42,58,76,77) Three additional orientations are observed in recent structures of pol-prim in the presence of single-stranded DNA or an RNA-primed template.(70,75)

To obtain insight into the extent of configurational flexibility of PRIM2C, we collected SAXS data for the tetramer core with PRIM2C attached (polΔcat-prim) and compared the results

to those described above for the tetramer core construct lacking PRIM2C (polΔcat-primΔ2C) (Figure 3.2B-D, Table 3.2). As for polΔcat-primΔ2C, polΔcat-prim was highly amenable to SAXS analysis and high quality, aggregation-free scattering data were acquired. The molecular weight estimate from the SAXS analysis corresponded well to the mass of the protein and accurately reflected the ~25 kDa difference with respect to polΔcat-primΔ2C due to the presence of PRIM2C. Interestingly, despite the difference in molecular weight, the Rg and Dmax values for the two constructs were similar, from which it can be inferred that interatomic distances for both proteins are dominated by those within the tetramer core. Nevertheless, a clear difference in shape can be observed in comparing the Mid-Q region of both the intensity and Kratky plots (Fig 3.2B, C). Moreover, the P(r) distributions (Figure 3.2D) indicate a change in shape to overall cylindrical from a more dumbbell shape with PRIM2C present, which is possible if, on average, PRIM2C is located somewhat centrally between the POLA1C/POLA2 lobe and the PRIM1 lobe of the tetramer core.

The comparison of the SAXS results clearly shows that PRIM2C is flexibly attached to the tetramer core. The Kratky plot indicates polΔcat-prim is a mostly globular particle but has significant interdomain flexibility compared to polΔcat-primΔ2C (Figure 3.2C). Thus, the attachment through a ~20-residue flexible linker enables PRIM2C to occupy a wide range of relative orientations with respect to the tetramer core. This is also reflected in the substantial reduction in the P_x for polΔcat-prim to ~3.3 from the value of 3.9 for the tetramer core alone. Unlike the conformational heterogeneity *within* the tetramer core, the substantial changes in the orientation of PRIM2C relative to the tetramer core have a large impact on the surface-area to volume ratio and correspondingly result in the lower P_x.

PRIM2C occupies a similar orientation in all available structures of pol-prim alone, corresponding to the compact Auto-inhibitory state. None of the corresponding models of polΔcat-prim extracted from these structures fit well to our SAXS data, and even multi-state modeling was insufficient to obtain satisfactory fits (Figure 3.3). In an effort to better understand the range and degree of configurational freedom of PRIM2C in solution we performed a limited MD simulation restrained by the SAXS data, starting from polΔcat-prim coordinates extracted from the pol-prim crystal structure in the Auto-inhibitory state. From these simulations, the best-fitting single models place PRIM2C in a central position between the PRIM1 and POLA1C/POLA2 lobes, differing significantly from the Auto-inhibitory configuration of PRIM2C, which we note is influenced by contacts with POLA1cat. [28,45,46,50] Single models from these simulations provided only adequate fits. However, the ensemble of SAXS data strongly suggest that PRIM2C occupies a range of configurations in solution that is not well represented by single structures. While even better fits to the SAXS data could be obtained with multi-state modeling, the simplistic nature of our MD simulations limited the accuracy of insights that could be obtained from analysis with multi-state modeling from predicted structures and it was decided

that a more direct approach was needed.

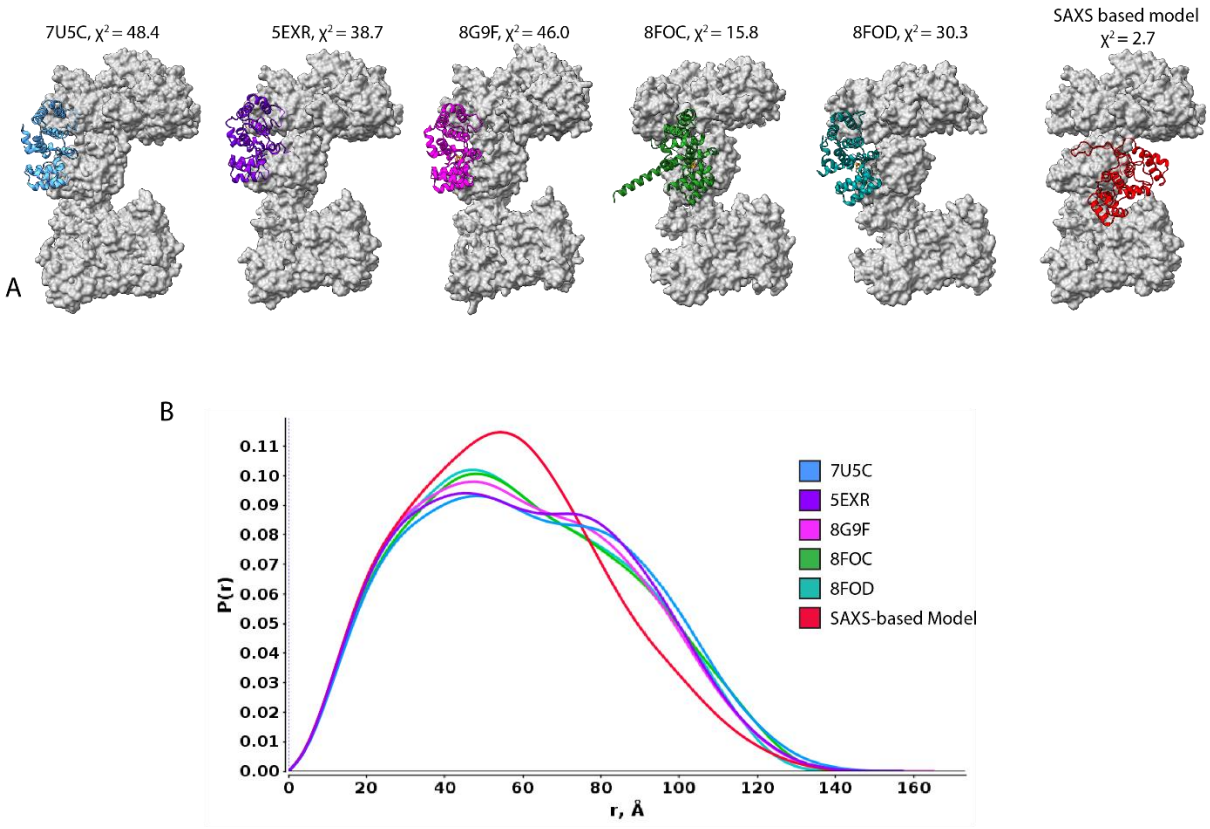

Figure 3.3 Modeling Configurational Flexibility of PRIM2C in the absence of substrate.

A) Models of polΔcat-prim extracted from published structures of eukaryotic pol-prim in the absence of substrate and a SAXS-based limited MD simulation model. In models of published structures, only atoms corresponding to polΔcat-prim were kept. Structures in which CST was present are marked with an asterisk. Tetramer core is depicted in gray surface representation while PRIM2C is highlighted in colored ribbon. Models are aligned to PRIM1. χ^2 values for each model are in comparison to polΔcat-prim scattering data. B) back calculated distance distributions from the models in A.

To more directly characterize variability in the orientation of PRIM2C with respect to the tetramer core, we analyzed the structure of polΔcat-prim by electron microscopy (Figure 3.4A). Negative stain EM images showed clear particles corresponding to the expected shape of polΔcat-prim based on the SAXS data and the available structures of pol-prim. Excellent density was observed in 2D class averages for the bulk of the globular tetramer core, albeit with blurred

density for PRIM1. In addition, the density for PRIM2C was highly blurred. 3D reconstructions lacked consistent density for PRIM2C, with only a small, disconnected segment of density visible that could not confidently be assigned to PRIM2C. The observation of blurred density for PRIM1 corresponds well to the flexibility of PRIM1 with respect to the rest of the tetramer core arising from the hinge within PRIM2N described above. Likewise, the lack of density for PRIM2C is consistent with our SAXS analysis that PRIM2C occupies a continuous range of configurations in solution rather than a limited number of discrete states.

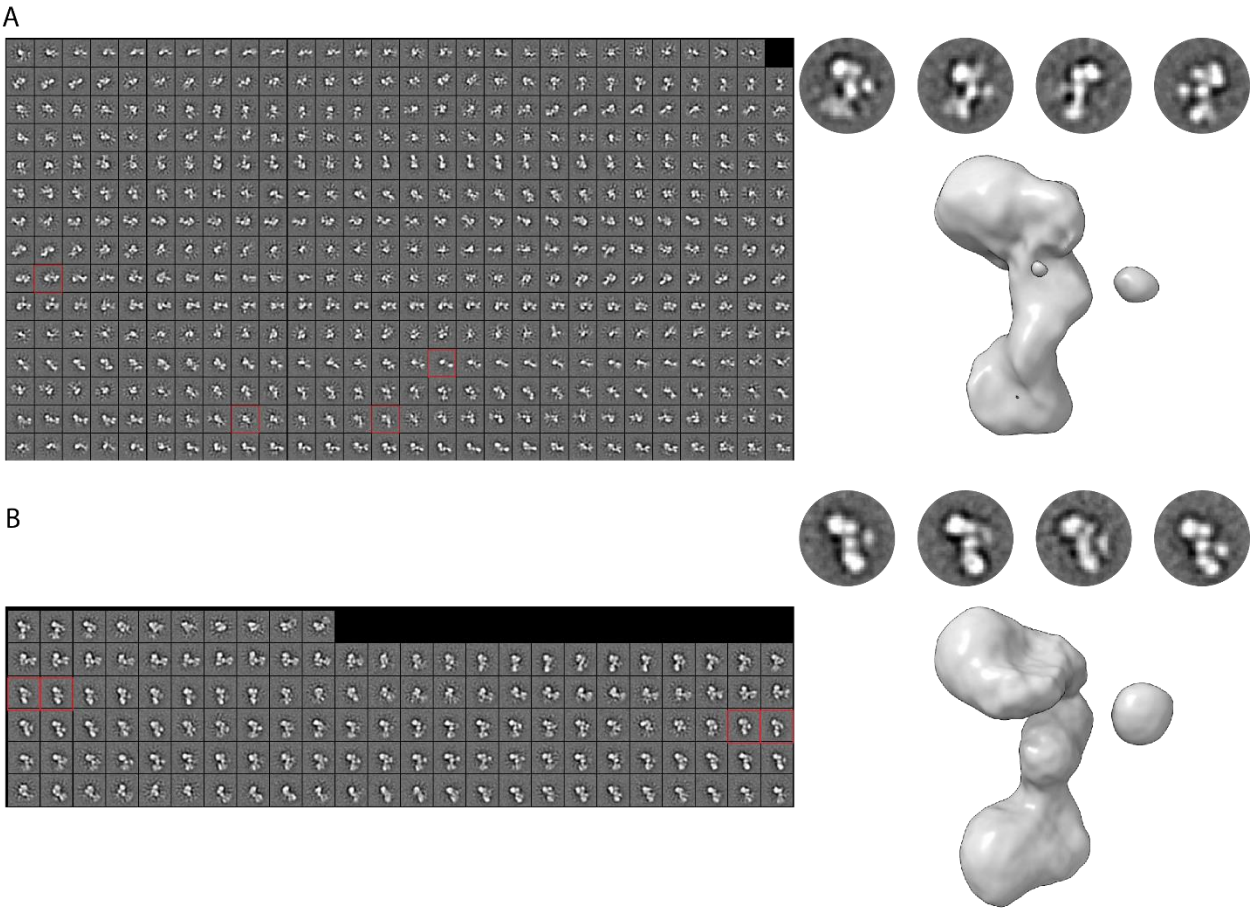

Figure 3.4 Negative Stain EM without and with an RNA elongation substrate.

2D class averages with selected 2D classes boxed in red and magnified at right, with consensus 3D reconstruction for polΔcat-prim (A) without and (B) with the RNA elongation substrate. 2D classes were created using ISAC in sphire while 3D reconstruction was performed in cryoSPARC.

To improve the resolution of our negative stain EM images, BS3 crosslinking was applied, which as anticipated greatly increased homogeneity. In 2D class averages of crosslinked polΔcat-prim, strong density was observed for the tetramer core (including PRIM1) from a variety of viewing angles (Figure 3.5-3.7). In both 2D class averages and 3D reconstructions, the conformation of the tetramer core was noticeably more homogenous than in the non-crosslinked sample and the characteristic S-shaped bilobal architecture was clearly evident. Importantly, PRIM2C was highly visible in 2D class averages of the complex in a continuous range of configurations spanning from near PRIM1, to a location similar to SAXS-based models, to the orientation observed in the Autoinhibitory state (i.e. near the interface of PRIM2N and POLA1C).(42,58,70,76,77) 3D reconstruction and classification yielded two primary classes, one in which PRIM2C was centrally located, as in the SAXS based models, and the other in which PRIM2C was located near POLA1C, as in structures of the Autoinhibitory state (Figure 4, S3.6).(42,58,70,77) These are presumably two well-populated orientations from within the ensemble of configurations accessible to PRIM2C that have been selected for by crosslinking.

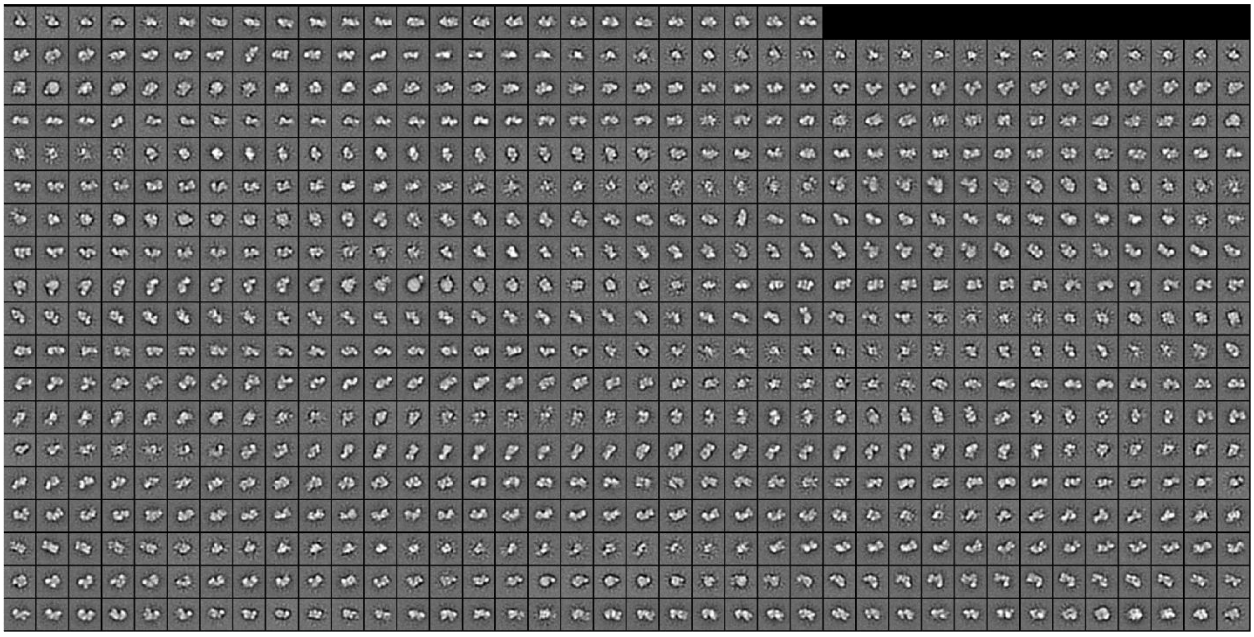

Figure 3.5 2D class averages of crosslinked no substrate polΔcat-prim.

2D classes were created using ISAC in sphire. Red boxes indicate selected class averages depicted in Figure 4.

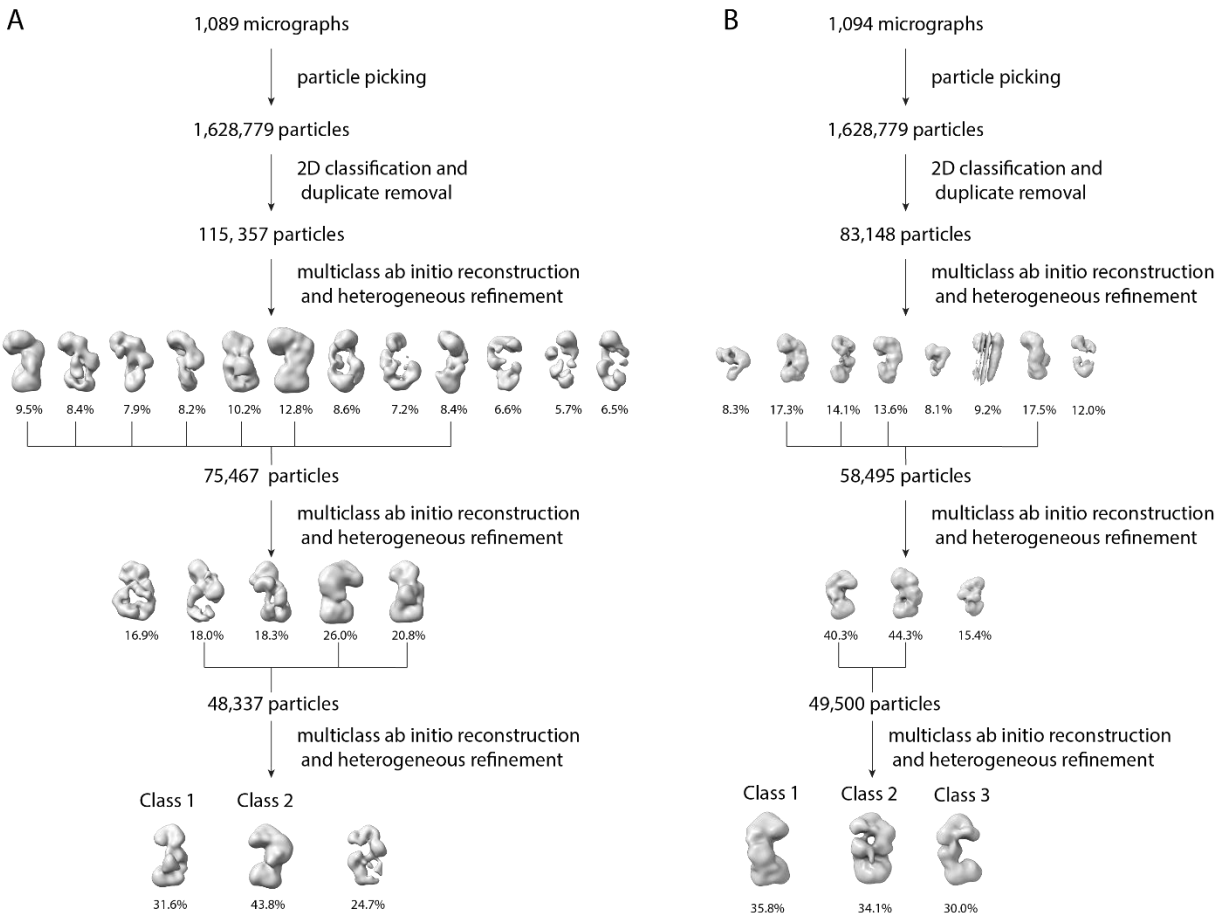

Figure 3.6 Negative Stain EM data processing approach.

Initial particle picking was performed using crYOLO in sphire and box-particle picking in cryoSPARC. Subsequent steps were performed in cryoSPARC. A) No substrate, crosslinked. B) Elongation substrate, crosslinked.

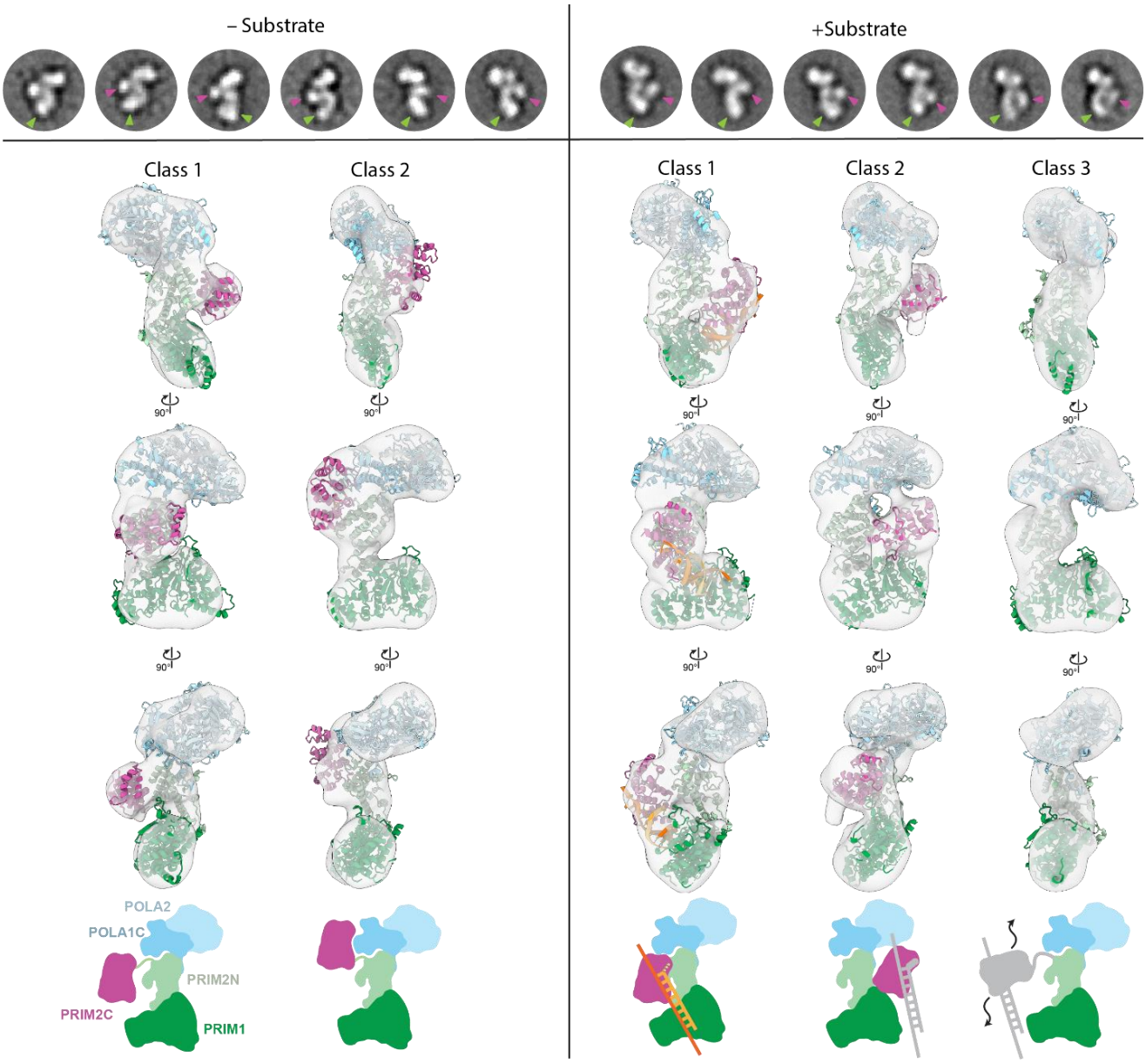

Figure 3.7 PRIM2C orientation and heterogeneity visualized by Negative Stain EM.

Top row- selected 2D class averages of prim +/- elongation substrate with PRIM2C (purple) and PRIM1 (green) density highlighted. Comparison of 3D reconstructions without (left) and with (right) elongation substrate. Atomic models rigid-body docked into the density is shown for polΔcat-prim. The +substrate class 1 model corresponds to the elongation model shown in Figure S5. Bottom row- schematic diagrams drawn to correspond to atomic models to which they align.

PRIM2C does not frequently sample the catalytically active RNA initiation configuration

The configurational flexibility of PRIM2C leading to changes in orientation of the domain is key to the progression through the multiple stages of primer synthesis. To better understand these changes, we set out to characterize the structural flexibility of polΔcat-prim under conditions required for RNA primer initiation and with an RNA-primed template. Ideally, SAXS could be used for this analysis, although we deemed it important to first determine if this approach would be sufficiently sensitive to detect the anticipated differences in the structure. For these simulations, we had to first generate structural models of polΔcat-prim: one in the configuration required for initiating RNA primer synthesis and the other in the configuration when bound to an 8mer RNA primed template elongation substrate (Figure 3.8A,E). Models for initiation require that PRIM2C is located over the active site of PRIM1 and engaged in binding nucleotide; interestingly, the configuration of PRIM2C in our model corresponds closely to that observed in a recently published structure of pol-prim bound to template DNA and scaffolded by CST,(75) but differs significantly from the structure of yeast pol-prim alone bound to template DNA in which PRIM2C and PRIM1 are separated from each other.(70) In generating our model for the complex with the elongation substrate, PRIM2C had to be moved away from PRIM1 as PRIM2C remains bound to the 5' triphosphate moiety of the nascent RNA primer as PRIM1 extends the 3' end.(58,68) Our model was generated for polΔcat-prim in an actively elongating configuration, which differs significantly from a recent structure of yeast pol-prim in the presence of an RNA-primed template in which PRIM2C is associated with the primer but the 3' end of the growing primer is ~27 Å away from the PRIM1 active site in a catalytically inactive configuration (Figure 3.8G).(70)

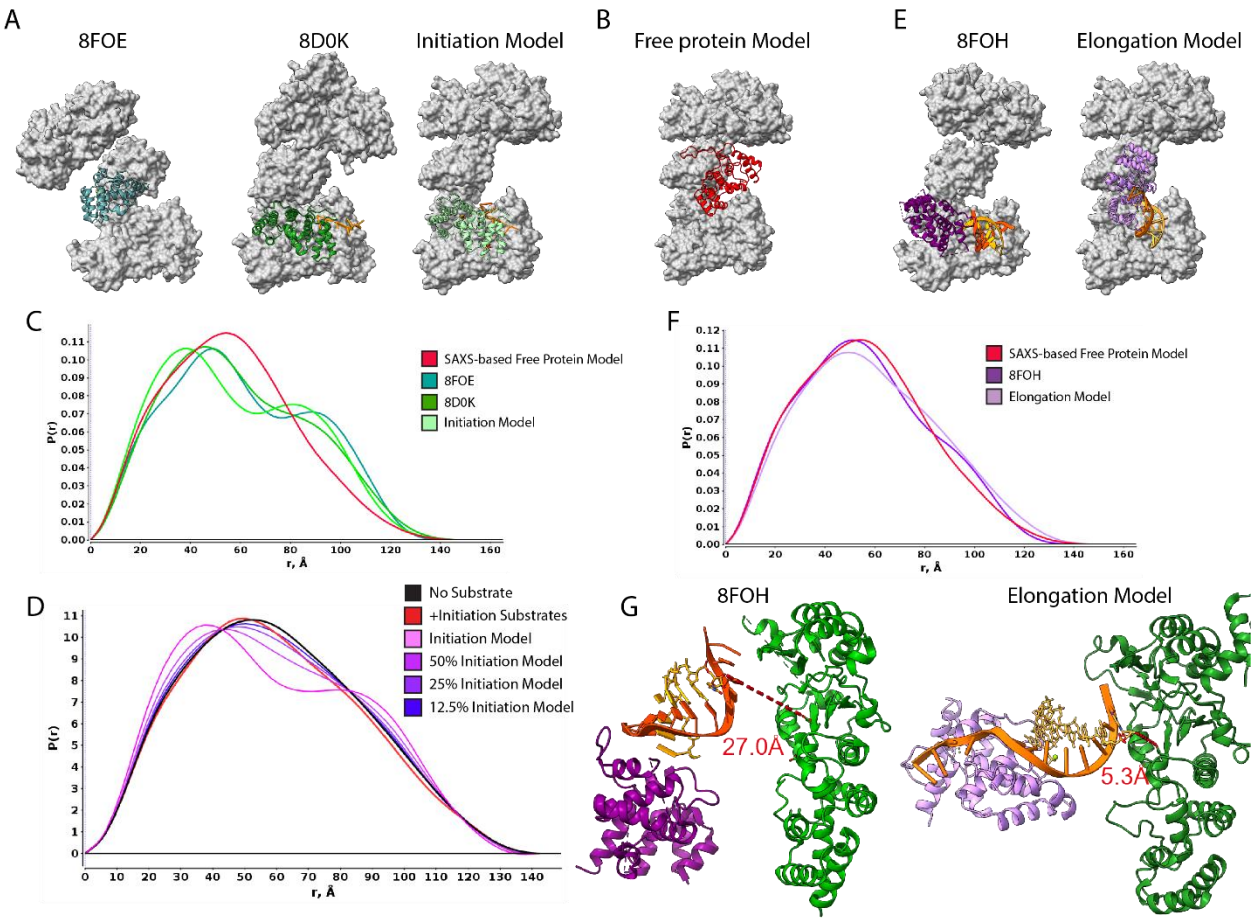

Figure 3.8 Modeling configurational change of PRIM2C during RNA priming.

A) Models of polΔcat-prim in the presence of initiation substrate. Tetramer core is depicted in gray surface representation while PRIM2C is highlighted in colored ribbon. Models are aligned to PRIM1. B) No-substrate SAXS-based model of polΔcat-prim as shown in Figure S1 for comparison. C) back-calculated distance distribution from above atomic models. D) Comparison of atomic model of initiation complex (Fig S1, pink) with experimental data (red) and mixtures of free protein (black) and initiation model (purple to blue). E) Models of polΔcat-prim in the presence of RNA elongation substrate. Tetramer core is depicted in gray surface representation while PRIM2C is highlighted in colored ribbon. Models are aligned to PRIM1. F) back-calculated distance distribution from above atomic models of polΔcat-prim bound to RNA-primed template. G) Comparison of orientation of PRIM2C and PRIM1 in a recently published structure (PDB: 8FOH)(70) and a manual model of RNA elongation. A large distance is observed between the catalytic ASP of PRIM1 and the 3' Oxygen of the RNA primer in the published structure while in our manual PRIM1 engages the substrate for catalysis.

SAXS data for the polΔcat-prim models for initiation and elongation of RNA synthesis were then back-calculated and, as a point of reference, were compared to the experimental scattering data of the free protein (Figure 3.8). The back-calculated scattering profiles and distance distributions for the models confirm that the changing orientation of PRIM2C with respect to the tetramer core during primer initiation and elongation has a significant impact on the corresponding SAXS parameters. Based on these simulations we hypothesized that the magnitude of shape change in the presence of substrate would be detectable by SAXS and designed two experiments to acquire SAXS data corresponding to these two key points in the trajectory of RNA primer synthesis.

Table 3.2 SAXS parameters.

Sample	Rg – Guinier (Å)	Rg - Real Space (Å)	Dmax (Å)	Molecular Weight (kDa)	Px
polΔcat-primΔ2C	45.3 ± 0.4	46.5 ± 0.1	145	160	3.9
polΔcat-prim	46.6 ± 2.4	46.9 ± 0.1	151	180	3.3
polΔcat-prim with Elongation substrates	49.7 ± 0.6	48.7 ± 0.1	153	200	3.3
polΔcat-prim with Initiation Substrates	45.3 ± 0.9	46.7 ± 0.5	145	180	3.3

The first experiment involved acquiring SAXS data for polΔcat-prim in the presence of a large excess of all components required for initiation: template single-stranded DNA, NTPs (in the form of the non-hydrolyzable nucleotide analogue GMPCPP), and the catalytic metal Mn^{2+} (Table 3.2). Our initial expectation was that these conditions would be sufficient to drive polΔcat-prim to occupy the configuration required for catalysis with PRIM2C aligned to the active site of PRIM1. However, the shifts in the Mid-Q region of the log10 intensity and in the Kratky plots were very modest (Figure 3.9). Moreover, the P(r) distribution was mostly unchanged relative to the free protein (Figure 3.9). These observations suggest that the presence

of the substrate and co-factors needed for initiating RNA priming does not drive pol-prim to exclusively occupy the configuration required for catalysis. As an aide to interpreting these data, we performed a series of additional simulations to estimate to what extent a shift in population would be detectable using two models: one of polΔcat-prim in an 'open' configuration with PRIM2C away from PRIM1 and the other our model of the initiation configuration with PRIM2C aligned to PRIM1 (Figure 3.8D). These involved back-calculating the scattering profiles for different mixtures of the two configurations and comparing to the experimental data. From these simulations we concluded that a precision of ±20% could be obtained in estimating the ratio of one configuration relative to the other. From this analysis of the experimental data we determined that even in the presence of excess template and co-factors only a small proportion of particles (≤10%) adopt the configuration required for initiating RNA synthesis, i.e. the presence of initiation substrates does not drive a major configurational shift on its own. These results imply that for RNA priming to initiate, pol-prim must sample a rare configuration in which both PRIM2C and PRIM1 are aligned (and simultaneously have the initiation substrates and co-factors bound), providing a molecular mechanism for the previous proposal based on biochemical studies that RNA initiation is the rate limiting step of primer synthesis.(66)

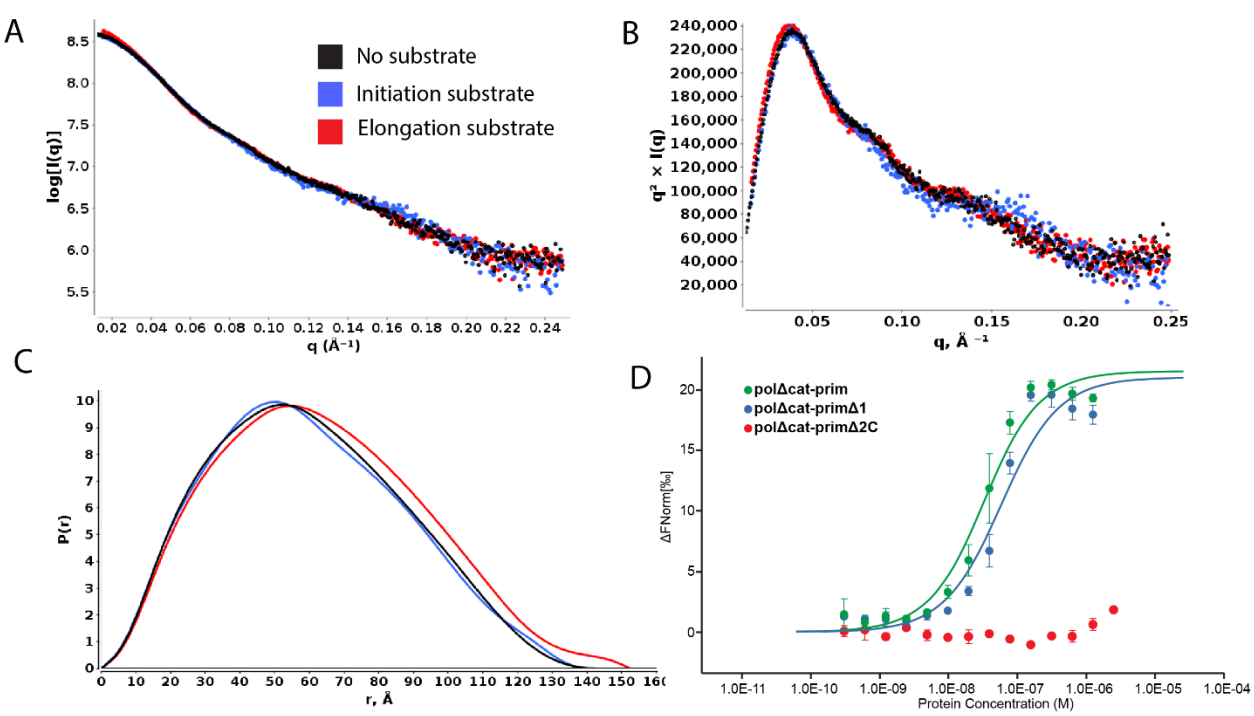

Figure 3.9 RNA elongation substrate engagement in solution.

Comparison of SAXS data for free polΔcat-prim (black) and polΔcat-prim in the presence of an RNA elongation substrate (red), and substrate and co-factors required for initiation (blue): A) log10 intensity plot; B) normalized Kratky plot; C) distance distribution. D) MST dose-response curve measuring RNA elongation substrate binding of polΔcat-prim (green), polΔcat-primΔ1 (blue), and polΔcat-primΔ2C (red).

PRIM2C flexibility is retained during RNA elongation

To investigate PRIM2C configurational flexibility during RNA elongation of the primer strand, we performed a similar set of analysis on polΔcat-prim with an 8mer RNA elongation substrate (5' tri-phosphorylated 8mer RNA annealed to single-stranded DNA). First, SAXS experiments were performed and again, the system provided high quality data. Surprisingly, the SAXS intensity and Kratky plots were quite similar to those observed for the free protein (Figure 3.9A,B), even thought there was a distinct change in size; the molecular weight estimate, Rg, and Dmax all indicated that the ~14 kDa RNA/DNA duplex substrate was bound (Table 3.2).

Moreover, the shift to the right in the P(r) distribution was consistent with an increase in the overall size of the complex, but not a change in shape (Figure 3.9C).

To confirm that the substrate was indeed bound and determine if both PRIM1 and PRIM2C were engaged, we carried out a series of substrate binding experiments using microscale thermophoresis (MST). For these experiments, we determined the relative contributions to substrate binding by using polΔcat-prim, the tetramer core construct (polΔcat-primΔ2C), and a third construct that deletes PRIM1 but retains PRIM2C (polΔcat-primΔ1) (Figure 3.1C). The substrate for these experiments was a 9mer RNA-primed template, designed specifically to maximize the stability of the duplex (Figure 3.9D, Table 3.3). As anticipated based on previous studies,(68) we observed very tight binding of this substrate by polΔcat-prim (K_d = 28 ± 7 nM). The deletion of PRIM1 resulted in a two-fold decrease in binding affinity (K_d = 53 ± 18 nM), consistent with previously reported affinities for PRIM2C alone.(68) In contrast, truncation of PRIM2C resulted in a significant decrease in affinity, below the limit of detection of the assay. Together, these data show that PRIM2C is the primary contributor to the substrate binding affinity of the primase subunits, consistent with previous studies of isolated PRIM2C and isolated primase (PRIM1/PRIM2), which concluded that PRIM2C binds substrates with higher affinity than PRIM1.(46,67,68)

Table 3.3 Estimated Dissociation Constants.

Protein	**K_d (nM)**
polΔcat-prim	27.9 ± 6.9
polΔcat-primΔ1	53.3 ± 18.3
polΔcat-primΔ2C	N.D.

We next performed negative stain EM imaging of polΔcat-prim in the presence of the elongation substrate, first without BS3 crosslinking (Figure S3.4B). As anticipated, the conformational variability within the tetramer core was unchanged relative to polΔcat-prim alone. There was however slightly better density for both PRIM1 and PRIM2C in both 2D class averages and a 3D reconstruction. This observation suggests a reduction in the configurational variability of pol-prim in the elongating state, as might be anticipated when a ligand is bound. Nevertheless, in most 2D class averages, the density for PRIM2C was blurred and in a similar position to that seen in the absence of substrate. However, in a small subset of classes PRIM2C appeared to be oriented nearer to PRIM1, an orientation that was not obviously apparent in the absence of substrate. Although there was improved density in the class averages and 3D reconstruction, a high degree of variability in the positioning of both PRIM2C and PRIM1 was still evident even when the elongation substrate is bound.

Why are the differences between the free protein and elongation complex so small? The key factor is the low affinity of PRIM1 for substrates relative to the very high affinity of PRM2C, which means that for the majority of the polΔcat-prim molecules PRIM2C is engaged but PRIM1 is not. As a result, PRIM2C can move away from its alignment over PRIM1, similar to what is observed for free polΔcat-prim. A second major factor is the overall degree of flexibility in the complex due to both the conformational variability in the tetramer core and the configurational variability of PRIM2C, which complicates alignment and reconstruction by electron microscopy.

Configurational freedom of PRIM2C and substrate dissociation from PRIM1 are key features during RNA priming

RNA elongation requires population of a pol-prim configuration in which the 3’ end of the growing primer is engaged in the PRIM1 active site, while PRIM2C remains tightly bound to the 5’ end of the primer. Although the EM data acquired without cross-linking provided some evidence of the orientation of PRIM2C in an active configuration, sorting of particles into discrete classes was hampered by the high degree of ambiguity due to PRIM1 and PRIM2C flexibility. To ultimately characterize the configurational flexibility of PRIM2C during RNA elongation, we added BS3 crosslinker to the complex of polΔcat-prim with the RNA elongation substrate. The conditions for crosslinking were optimized by following the extent of crosslinked products by SDS-PAGE and by comparing the results with and without substrate (Figure 3.10). A series of crosslinked products is observed on the gel in the absence of substrate corresponding to various combinations of the subunits, but with the four-subunit product being a predominant species. With the addition of substrate, these bands shifted to slightly higher molecular weight, consistent with successful crosslinking of the substrate to polΔcat-prim. To obtain a more homogeneous sample for EM, the mixture was purified by SEC.

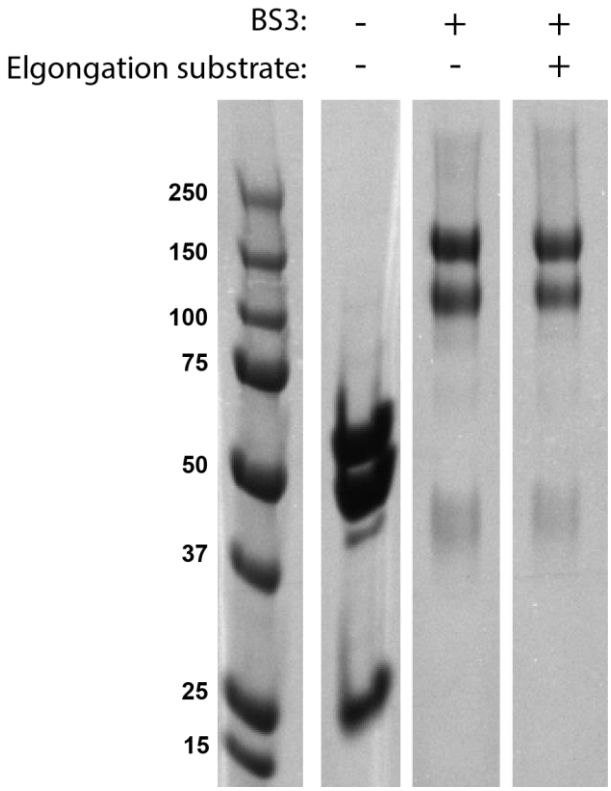

Figure 3.10 BS3 crosslinking products analyzed by SDS-PAGE.

polΔcat-prim in the presence and absence of both BS3 crosslinker and RNA elongation substrate.

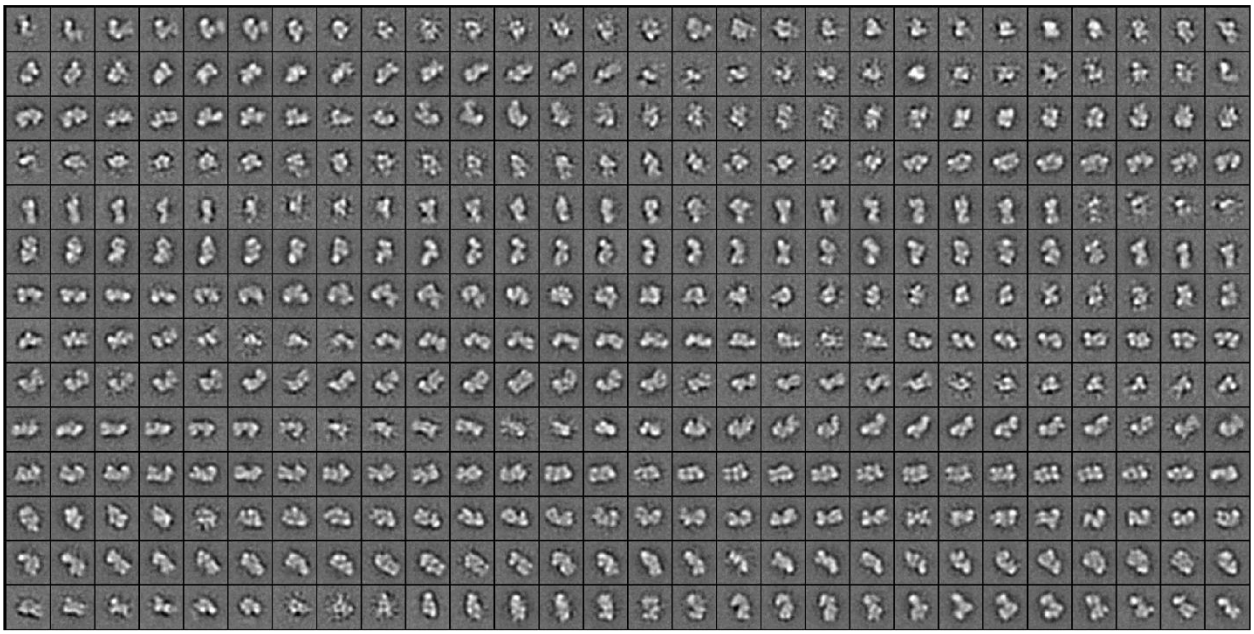

Figure 3.11 2D class averages of crosslinked polΔcat-prim in the presence of RNA elongation substrate.

Analysis of the negative stain EM data (Figure 3.6, S3.11) revealed the overall architecture of polΔcat-prim was retained when bound to the elongation substrate with the tetramer core adopting the characteristic S-shaped bilobal conformation (Figure 3.7). In 2D class averages, a shift in the proportion of classes was observed in which PRIM2C and PRIM1 were aligned and formed continuous density, indicating enrichment of particles with a configuration in which both PRIM1 and PRIM2C are aligned and bound to the elongation substrate (Figures 3.7, 3.11). However, a subset of 2D classes were present in which PRIM2C was not aligned over PRIM1 and instead occupied a central location, as expected if PRIM1 is not bound to the substrate. 3D reconstruction yielded three primary classes (Figures 3.6, 3.7). In class 1 (36%), clear density was visible for PRIM2C and formed continuous density with PRIM1, consistent with both PRIM1 and PRIM2C being bound to substrate. Interestingly, insertion of our model of polΔcat-prim bound to an RNA elongation substrate into the density revealed an excellent overlap in the position of PRIM2C. In class 2 (34%), density for PRIM2C was observed, but was shifted away from the tetramer core while still exhibiting some continuous density with PRIM1, though to a lesser extent than in class 1. This class has additional density protruding from PRIM2C, which we attribute to substrate bound to PRIM2C but not PRIM1. A recently published structure of pol-prim in complex with an RNA-primed template reveals PRIM2C bound to the substrate but positioned away from the active site of PRIM1 (Figure 3.8G).(70) While there is considerable uncertainty about the assignment of this structure as an RNA synthesis state because the primer lacks the critical 5' tri-phosphate that is key to PRIM2C interaction and required for RNA priming, we note a general similarity to class 2 of our 3D reconstructions. In class 3 (30%), density for the tetramer core was very similar to the density observed in the absence of substrate,

but density for PRIM2C was not visible, which we attribute to the configurational heterogeneity of the domain because the substrate is not bound to PRIM1.

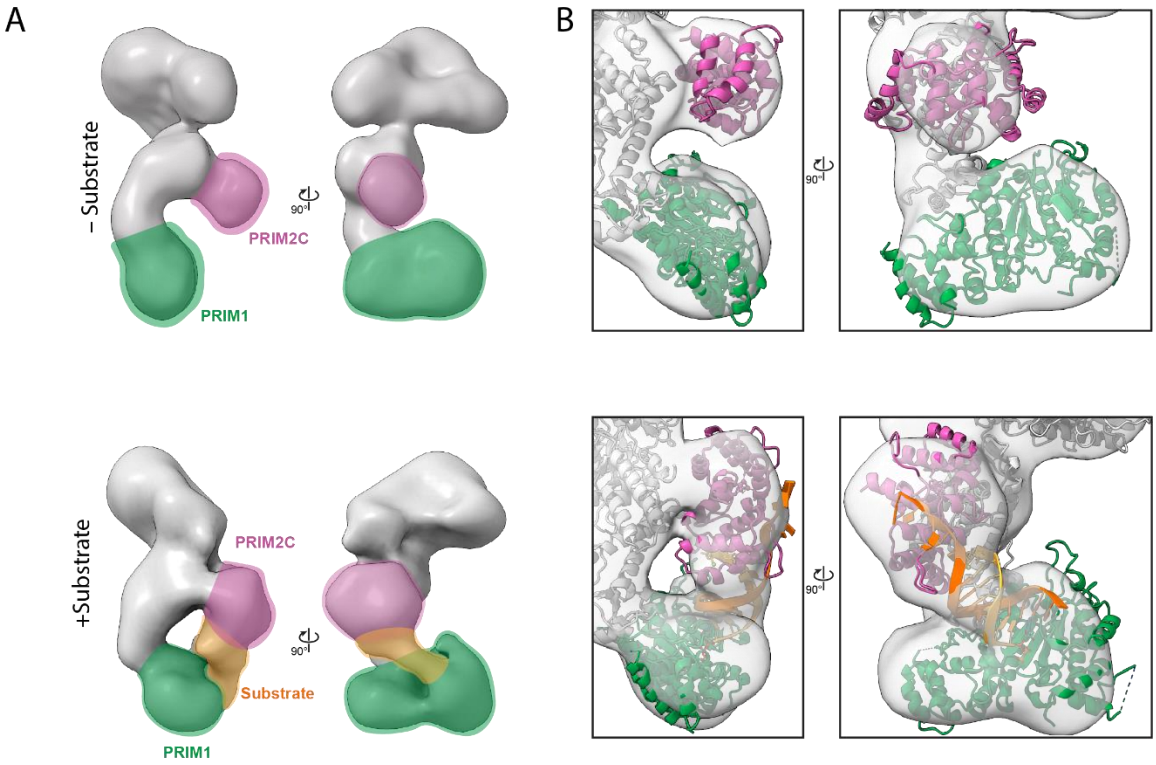

Figure 3.12 3D map of polΔcat-prim and the complex with the RNA elongation substrate.

A) 3D reconstruction of class 1 without (top) and with substrate (bottom). The density for PRIM2C (purple) PRIM1 (green), and substrate (orange) are highlighted. B) Position of PRIM2C and PRIM1 from atomic models fit into the maps obtained without and with substrate.

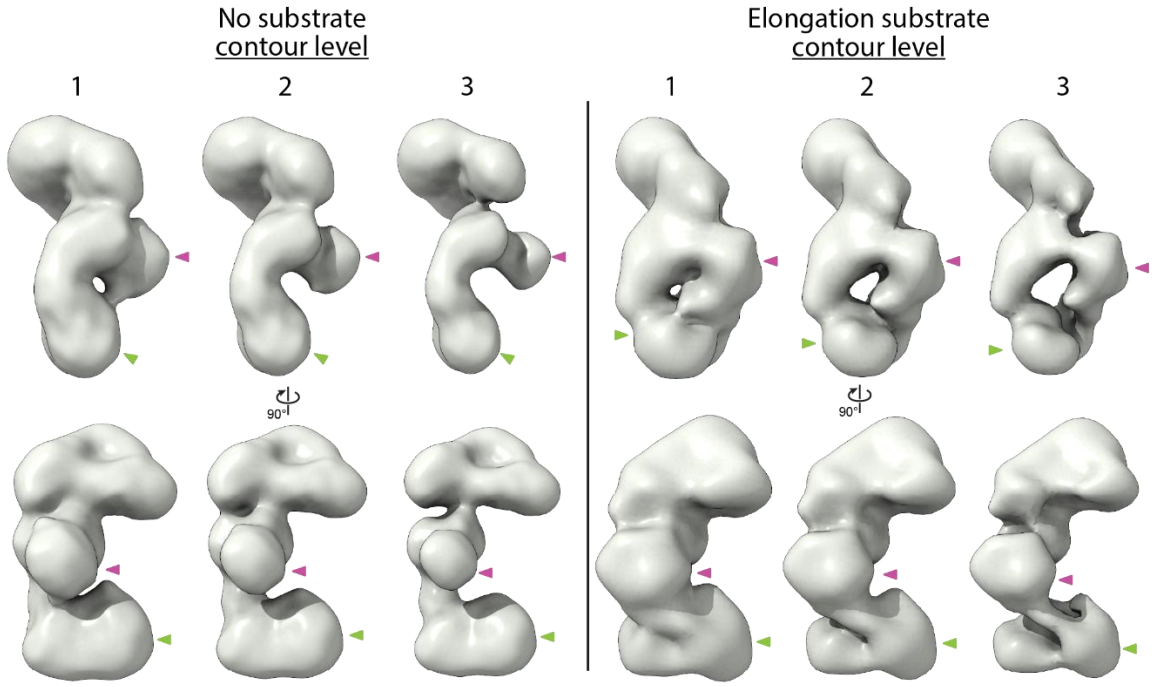

Figure 3.13 Density Connects PRIM2C and PRIM1 after addition of RNA elongation substrates.

3D reconstruction at 3 contour levels for A) class 1 no substrate and B) class 1 elongation substrate with PRIM1 (green) and PRIM2C (purple) indicated with arrows.

Interestingly, the location of PRIM2C in both class 1 with substrate and class 1 without substrate is very similar. However, comparison of the density connecting PRIM2C and PRIM1 provides a stark contrast between the two states, especially at lower contour levels (Figure 3.12, 3.13). In the absence of substrate, a clear separation between PRIM2C and PRIM1 is visible corresponding to a species in which PRIM2C is disengaged from PRIM1. In the presence of the elongation substrate, clear tubular density is visible connecting PRIM2C and PRIM1, nicely corresponding to the location of the RNA/DNA duplex in the atomic model of the elongation complex.

Overall, a significant degree of configurational variability in PRIM2C is evident even after crosslinking. In the absence of substrate, PRIM2C appears to occupy a continuous range of configurations, sampling many possible orientations. In the presence of substrate, there remains a significant range of configurations of PRIM2C, but the origin is different: it is the dissociation of the substrate from PRIM1 that allows a higher than anticipated degree of configurational freedom for PRIM2C.

Discussion

The basic steps of primer synthesis by pol-prim involves formation of the initial dinucleotide by the concerted action of PRIM1 and PRIM2C, then elongation to 7-10 nucleotides by PRIM1 before the RNA primer is handed off (transferred intramolecularly) to POLA1cat.9–11,56 PRIM2C has been shown to play roles in RNA initiation, elongation, length regulation, and handoff.(37,67) However, the structural basis for how PRIM2C functions is incomplete, having relied heavily on inference from biochemical experiments and structures of PRIM2C alone or of larger pol-prim constructs in the absence of substrates.(43,46,58,110)

Our characterization of polΔcat-prim provides key structural context for assessing proposed biochemical models of RNA primer synthesis. In particular, we established that PRIM2C has inherent configurational flexibility, visualizing a wide range of orientations relative to the tetramer core. In addition to PRIM2C flexibility, we expand upon recent studies to directly characterize the conformational heterogeneity of PRIM1 in the tetramer core. These high degrees of flexibility provide the adaptive architecture needed to accommodate the multivalent and semi-independent substrate binding by PRIM2C and PRIM1 that is characteristic of RNA primer initiation and elongation.

Our binding studies systematically tested the previous proposal that PRIM2C is the primary contributor to substrate binding affinity during RNA priming.(67,68) Our structural studies revealed the very low population of pol-prim configurations in which PRIM1 and PRIM2C are aligned for synthesis of the initial dinucleotide even when the template and co-factors are present. For the complex with an RNA elongation substrate, crosslinking enabled visualization of three primary states: one in which PRIM2C and PRIM1 are both bound to the primer in a configuration competent for RNA elongation, the second in which PRIM2C is visible and bound to substrate but has moved away from PRIM1, and the third in which PRIM2C is not visible and is structurally independent from PRIM1 and the rest of the tetramer core. This degree of configurational variability of PRIM2C even in the presence of substrate: (1) provides evidence in support of the proposal that PRIM1 dissociates from the primer during synthesis, (2) directly correlates to the distributive nature of RNA elongation by primase, and (3) further contextualizes structures of pol-prim in the presence of an RNA-primed template.(46,70)

The large difference in the affinity of PRIM1 and PRIM2C for substrates is consistent with the apparent dominance of PRIM2C-bound/PRIM1-unbound particles and has important implications for the mechanism of primer synthesis. First, the asymmetry in the interaction with the substate provides a structural basis for the distributive (rather than processive) nature of primer synthesis by PRIM1 that has been established by previous biochemical analyses. These studies showed that PRIM1 synthesizes abortive primer products (di- or trinucleotides) in abundance in both the presence and absence of PRIM2C, while the presence of PRIM2C results in an increase in unit length 7-10-nucleotide primers.(35,37,66) Thus, PRIM2C acts as a processivity factor, decreasing the dissociation of short products before synthesis of the RNA primer can be completed.(46,58) PRIM1 also elongates pre-made primers to lengths of up to 40

nucleotides in an unregulated fashion in the absence of PRIM2C, while PRIM1/PRIM2 synthesizes primers either of defined length (7-10 nucleotides) or as multiples of 7-10 nucleotides.(37,68) Hence, PRIM2C also acts as a length regulator, influencing both the maximum and minimum primer length.

Our EM reconstruction in which PRIM2C and PRIM1 both engage an RNA-primed template substrate is the first structure of pol-prim bound to substrate and in a configuration competent for RNA synthesis. It also supports models of primer synthesis in which PRIM2C remains engaged throughout. In one such model, the upper bound for primer length is attributed to a steric clash between PRIM2N and PRIM2C at a primer length of ~10 nucleotides.(58) This model is consistent with the close proximity of the PRIM2C lobe and the tetramer core in our reconstruction of the complex with the RNA elongation substrate. Our recent studies of Xenopus pol-prim provide further insight, showing that the minimal length of 7 nucleotides of RNA is due to the inability of POLA1cat to bind primers shorter than 7 nucleotides as a result of steric occlusion by PRIM2C.(42)

Our results provide the basis for a unified model of RNA priming by pol-prim, which is consistent with all available biochemical and structural data and is outlined in schematic form in Figure 3.14. The initiation of primer synthesis requires alignment of PRIM2C and PRIM1 in a catalytically competent configuration as well as binding of the two nucleotides, the catalytic metals and the ssDNA template. The sparse population of pol-prim configurations with PRIM1 and PRIM2C aligned combined with the low binding affinities for the template, nucleotides and metals makes initiation the rate-limiting step of primer synthesis. After formation of the initial dinucleotide, PRIM1 extends the RNA primer in a distributive manner due to dissociation of the nascent primer from the PRIM1 active site, the result of its low affinity for nucleotide substrates.

PRIM2C functions as a processivity factor, remaining bound to the 5' end of the primer and thereby facilitating PRIM1 re-association with the 3' end of the primer. The lower bound for PRIM1 extension is 7 nucleotides because, as we have shown elsewhere, this is the minimum primer length necessary binding by POLA1cat.(42) Importantly, POLA1cat has also been demonstrated to exhibit high affinity for substrate in the nanomolar range, several orders of magnitude higher than the affinity of PRIM1 for RNA-primed template we observed.25,46,48,57,58 Thus, after the primer has reached 7 nucleotides, competition for the primer between PRIM1 and POLA1cat occurs in which POLA1cat is highly favored. In this model, the low affinity of PRIM1 for substrate means that the primed template is inherently available for transfer to POLA1cat. An additional contribution may come from an inherent orientational preference of PolA1cat that strategically positions it to quickly bind the 3' end of the RNA primer after PRIM1 dissociation.(70) Extension of the nascent primer by PRIM1 can occur up to 10 nucleotides if PRIM1 dissociation does not occur, after which steric clash between PRIM2C and the tetramer core greatly suppresses further elongation.

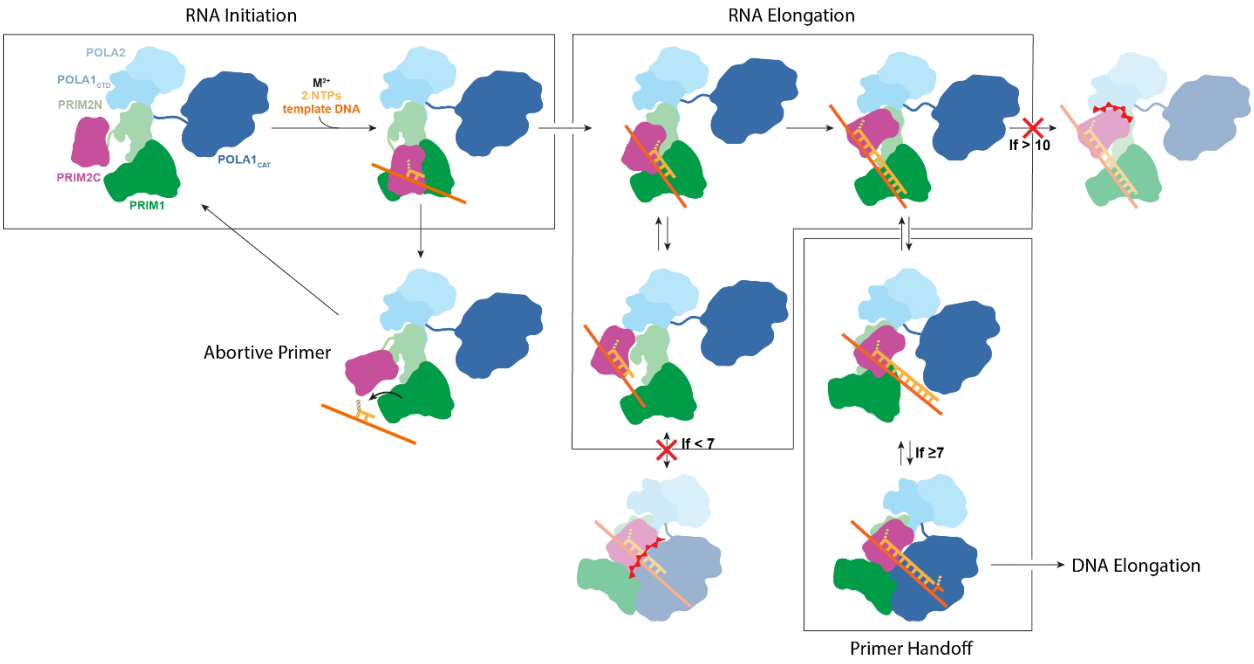

Figure 3.14 Model of RNA primer synthesis.

Initiation of RNA synthesis requires PRIM2C and PRIM1 to be aligned with template, catalytic metals and nucleotides bound to enable synthesis of the initial dinucleotide. The dinucleotide is elongated by PRIM1 while PRIM2C remains bound to the 5' triphosphate and progressively moves away from the active site of PRIM1 as the primer grows. The primed template dissociates from PRIM1 during primer synthesis due to its low affinity for the substrate. PRIM2C acts as a processivity factor by remaining bound to the primed template, which facilitates re-association with PRIM1. The primer can continue to lengthen until a steric clash between PRIM2C and PRIM2N prevents further lengthening when the primer reaches 10 nucleotides. The ready dissociation of PRIM1 facilitates handoff of the primer to POLA1cat for synthesis of the DNA portion of the primer, although handoff cannot occur until the primer reaches the minimum of 7 nucleotides required for POL1cat binding primed template with high affinity.

Although pol-prim is capable of performing all steps of primer synthesis in isolation, it is important to take into consideration the many interaction partners that recruit and modulate pol-prim activity at the replication fork.(20,26,27,31,75,77,78,128) Recent structures of pol-prim bound to CST have demonstrated that CST acts as a recruitment factor and can stabilize a configuration in which the flexible PRIM1 and PRIM2C are aligned for primer initiation while allowing for PRIM2C flexibility such that the PRIM1 and PRIM2C could spatially separate as

the primer elongates.(75,77) At replication forks, RPA may fulfill a similar function and facilitate positioning of PRIM2C and PRIM1 for primer initiation while strategically orienting POLA1cat for primer handoff. How these interactions are maintained or affected by additional components of the replisome remains poorly understood. Future studies to determine how these factors impact the structural organization and flexibility of pol-prim throughout the various steps of chimeric primer synthesis will provide deeper insight into how this dynamic molecular machine functions in its various cellular contexts.

CHAPTER 4. DISCUSSION AND FUTURE DIRECTIONS

Summary of This Work

Although it has long been known that pol-prim contains some level of interdomain flexibility because of the flexible tethers linking PRIM2C and POLA1cat to the tetramer core, the overall degree of flexibility remained uncharacterized. Underestimation of this flexibility frustrated early attempts at structural characterization by SAXS and EM, and has limited elucidation and comprehension of the mechanism of pol-prim function. Overall, this work has significantly advanced understanding of pol-prim flexibility, revealing three primary modes of flexibility and heterogeneity that are relevant to RNA priming by pol-prim (Figure 4.1). These modes are summarized below.

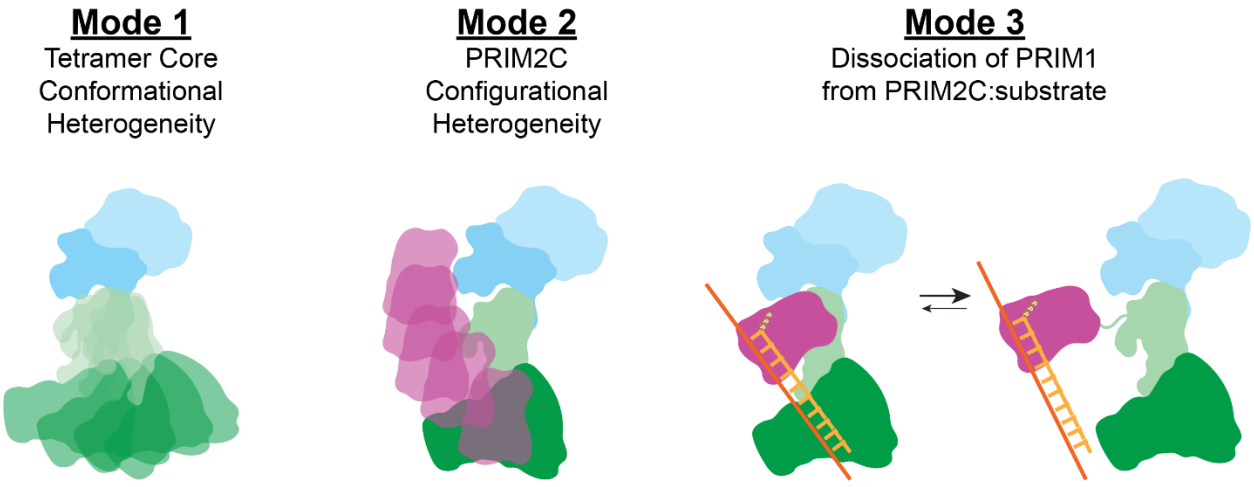

Figure 4.1 Three modes of flexibility and heterogeneity discussed in this study.

Mode 1: flexibility within the tetramer core due to a hinge in PRIM2N and at the interface between PRIM2N and POLA1C. Mode 2: configurational heterogeneity of PRIM2C as a function of its flexible linkage to the tetramer core. Mode3: dissociation of PRIM1 from the 3' end of the substrate due to its low affinity for substrate.

Pol-Prim is highly flexible

Although, advancements in cryoEM data processing software have allowed for the characterization of highly flexible proteins, our own analysis was unable to achieve an apparent resolution better than ~15 Å due to the high degree of heterogeneity within polΔcat-prim. Further SAXS, modeling, and negative stain EM shed light on the source and nature of this flexibility. In addition to the high degree of configurational flexibility of the 25 kDa PRIM2C, I observed a substantial conformational flexibility within the tetramer core of pol-prim. For EM analysis of RNA priming, this meant that the two domains of particular interest in the catalytic subunit, PRIM1 and the substrate binding domain PRIM2C, were both independently configurationally heterogenous and, due to their small size, difficult to obtain alignments on. I developed an approach involving crosslinking, negative stain EM and SAXS-based modeling to elucidate the range of configurational states present in polΔcat-prim.

Pol-prim undergoes structural transitions during RNA priming

Despite the challenge of experimentally characterizing flexibility within pol-prim, there is no doubt it is highly important for RNA priming. From modeling of the protein combined with negative stain EM, I established that PRIM2C initially occupies a range of configurations with respect to the tetramer core in the absence of substrates and co-factors. In the presence of an RNA elongation substrate and even after crosslinking, there remains significant heterogeneity in the configuration of PRIM2C. However, crosslinking enabled capture of multiple configurations in which both the PRIM2C domain and PRIM1 domain are bound to and engaged with the substrate. Moreover, my work has provided the first direct evidence of PRIM2C remaining bound to the primed template during RNA synthesis, and the first structural data in which

PRIM2C and PRIM1 are aligned in a catalytic configuration with substrate present at the PRIM1 active site.

Substrate readily dissociates from Primase active site

In addition to the heterogeneity observed relating to the flexibility of PRIM2C and within the tetramer core, several experiments revealed that simultaneous engagement of the substrate by both PRIM1 and PRIM2C is a rare event in solution during both RNA initiation and RNA elongation. In the case of RNA initiation, SAXS and modeling revealed that the configuration in which both PRIM1 and PRIM2C are aligned for initiation is rare (<10%) even in the presence of an excess of the substrate and cofactors required for initiation. For RNA elongation, binding studies directly showed PRIM2C is the primary contributor to substate binding as previously proposed based on studes of isolated PRIM2C, as removal of the PRIM2C domain from polΔcat-prim results in almost complete loss of binding for a 5' triphosphorylated RNA primed template substrate. Similarly, SAXS studies revealed that although the substrate was clearly bound, there was minimal shift in the overall shape of the protein. This suggests that in solution there is minimal shift from isolated polΔcat-prim in which PRIM2C is highly flexible to the RNA elongation complex, in which PRIM2C and PRIM1 were presumed to be both engaged on the substrate. Finally, crosslinked negative stain EM revealed three classes of particles in the presence of a primed template substrate. In addition to the species in which PRIM2C and PRIM1 are both engaged, in the second class, PRIM2C is visible along with some density corresponding to the presence of substrate, but there is no density connecting PRIM2C and PRIM1. This suggests that the substrate is disengaged from PRIM1 but still bound to PRIM2C. In the third class, density for the tetramer core is visible but not density for PRIM2C, likely due to a high degree of motion in this domain, as PRIM2C is bound to the substrate but not PRIM1, so

PRIM2C continues to occupy a wide range of configurations even in the presence of the tightly bound RNA enlongation substrate.

Implications of the Results

Comparison with recently published CryoEM Structures of Pol-prim

Recently published higher resolution cryoEM structures have offered critical insights into the structural organization of pol-prim throughout both the RNA and DNA primer synthesis cycles. Three papers in particular have shed light on the RNA priming half-reaction that is the focus of this thesis.(75,129,130) In a structure of human pol-prim scaffolded by CST in the presence of single-stranded DNA, density for the template DNA was observed that was oriented towards PRIM1 but strong density for the template was not observed bound to PRIM1.(75) PRIM2C was observed to be highly flexible in this state, with one extreme of the 3D variablitiy analysis trajectory exhibiting a configuration in which PRIM2C and PRIM1 were located in close proximity, termed a pre-initiation complex. The O'Donnell group subsequently published a series of structures with FL pol-prim in the presence of several nucleotide substrates representing key steps during RNA and DNA priming.(130) Importantly, in this study a harsh formaldehyde crosslinker was used to stabilize protein/DNA and protein/protein interactions. In the presence of single-stranded DNA meant to represent a pre-initiation state, the orientation of PRIM2C observed was close to but not directly oriented over PRIM1 for primer initiation. Density bound to PRIM1 or PRIM2C was not observed. In the presence of an RNA elongation substrate (though lacking the critical 5' triphosphate), a configuration of PRIM2C was observed in which PRIM2C was bound to the RNA primed template in a central configuration with respect to the tetramer core but the 3' end of the primer was not bound in the active site of PRIM1, consistent with our finding of ready dissociation of the primer from PRIM1. A series of structures recently published

on a preprint server by the Pellegrini group offer even further insight into several stages of RNA primer synthesis.(129) For these structures, FL pol-prim was incubated with single-stranded template and rNTPs and either immediately crosslinked with BS3 or crosslinked after a longer incubation period. In the shorter incubation sample, several pre-initiation and initiation configurations were observed in which PRIM2C occupies different configurations near to PRIM1, similar to the structures published by the Lim group in the presence of CST. For the longer incubation period, PRIM2C was observed to occupy a range of configurations apparently tracking the 5' end of the growing as PRIM1 extended from the 3' end. However, density for the primed template substrate was not observed either at the PRIM2C end or engaged with PRIM1, likely due to the inherent flexibility described in this work and the heterogenous lengths of primers synthesized and captured using this assay, as well as the complex approaches used to sort particles into discrete PRIM2C classes and determine structures. Despite these many attempts to characterize structural transitions during RNA priming by pol-prim, it is remarkable that this thesis presents the first, and to date only, conclusive structures in which density for the RNA primer is observed that is bound to both the PRIM1 catalytic site and the PRIM2C high affinity substrate binding site. Taken together, my work and these additional structures from other laboratories offer complementary pictures of the RNA priming mechanism. Two key takeaways are consistent are across all studies: 1) RNA primer initiation involves a configuration in which PRIM2C and PRIM1 are both engaged with the single-stranded template, but that this species represents only a small proportion of the orientations present in solution in the presence of initiation substrates as PRIM2C retains a high degree of configurational heterogeneity and 2) PRIM2C remains bound to the 5' end of the growing primer during RNA primer elongation while PRIM1 can readily dissociate from the 3' end of the primer.

RNA priming occurs via a dissociative mechanism facilitated by interdomain flexibility

The large difference in the affinity of PRIM1 and PRIM2C for substrates gives rise to the dominance of states in which PRIM2C is bound and PRIM1 is unbound. This in turn has important implications for the mechanism of primer synthesis. First, the weak binding of substrate by PRIM1 provides a structural basis for the distributive (rather than processive) nature of primer synthesis by PRIM1 that has been established by previous biochemical analyses. These studies showed that PRIM1 synthesizes abortive primer products (di- or trinucleotides) in abundance in both the presence and absence of PRIM2C, while the presence of PRIM2C results in an increase in unit length 7-10-nucleotide primers(35,37,66). Thus, PRIM2C acts as a processivity factor, decreasing the dissociation of short products before synthesis of the RNA primer can be completed(46,58). PRIM1 also elongates pre-made primers to lengths of up to 40 nucleotides in an unregulated fashion in the absence of PRIM2C, while PRIM1/PRIM2 synthesizes primers either of defined length (7-10 nucleotides) or as multiples of 7-10 nucleotides(37,68). Hence, PRIM2C also acts as a length regulator, influencing both the maximum and minimum primer length.

Multiple models have been proposed in which PRIM2C remains engaged throughout the three phases of RNA primer synthesis and this has fundamental implications for the length of RNA portion of the primer. In one such model, the upper bound for the length of the primer strand is attributed to a steric clash between PRIM2N and PRIM2C, which arises at a primer length of ~10 nucleotides(58). This model is consistent with the close proximity of the PRIM2C lobe to the tetramer core in our reconstruction of the complex of polΔcat-prim with the RNA elongation substrate, as the density for PRIM2C is in close proximity to the density of PRIM2N/POLA1C lobe of the tetramer core. Further extension of the primer past the 8

nucleotides present in this study would need to be accommodated by movement of PRIM2C towards PRIM2N/POLA1C, eventually resulting in a steric clash that cannot be compensated for by the flexibility within the tetramer core. Insight into the lower limit for the RNA portion of the primer was provided by recent studies of Xenopus pol-prim, which showed that the minimal length of 7 nucleotides of RNA is due to the inability of POLA1cat to bind primers shorter than 7 nucleotides as a result of steric occlusion by PRIM2C(42).

These results provide the basis for a unified model of RNA priming by pol-prim, which is consistent with all available biochemical and structural data and is outlined in schematic form in Figure 4.2. In order for the RNA primer to be initiated, pol-prim must undergo several structural transitions from an autoinhibitory conformation, to an unfolded conformation capable of binding the template DNA and incoming NTPs, then to a configuration in which PRIM2C and PRIM1 are aligned for synthesis of the dinucleotide. The sparse population of pol-prim configurations with PRIM1 and PRIM2C aligned combined with the low binding affinities for the template, nucleotides and metals makes initiation the rate-limiting step of primer synthesis. Similarly, the low affinity of PRIM1 for substrate can result in dissociation of the newly synthesized di-nucleotide primer from the active site, after which the dinucleotide can dissociate from the template strand resulting in an abortive primer. After formation of the initial dinucleotide, PRIM1 extends the RNA primer in a distributive manner due to dissociation of the nascent primer from the PRIM1 active site, the result of its low affinity for the substrate. PRIM2C functions as a processivity factor, remaining bound to the 5’ end of the primer and thereby facilitating PRIM1 re-association with the 3’ end of the primer. The lower bound for PRIM1 extension is 7 nucleotides because this is the minimum primer length necessary for binding by POLA1cat (42). Importantly, POLA1cat has also been demonstrated to exhibit high affinity for

substrate in the nanomolar range, several orders of magnitude higher than the affinity of PRIM1 for RNA-primed template we observed (40–42,60,64). Thus, after the primer has reached 7 nucleotides, competition for the primer between PRIM1 and POLA1cat occurs in which POLA1cat is highly favored. In this model, the low affinity of PRIM1 for substrate means that the primed template is available for transfer to POLA1cat. An additional contribution may come from an inherent orientational preference of POLA1cat that strategically positions it to quickly bind the 3' end of the RNA primer after PRIM1 dissociation.(70) Extension of the nascent primer by PRIM1 can occur up to 10 nucleotides if PRIM1 dissociation does not occur, after which steric clash between PRIM2C and the tetramer core greatly suppresses further elongation. Recent structures by the Pellegrini and O'Donnell group also suggest that POLA1cat may play a more direct role in primer length regulation in which it is strategically oriented to both inhibit primer growth after reaching ~10 nucleotides and is spatially close to PRIM2C to facilitate fast handoff. Once primer handoff has occurred the primer is extend with ~20 nucleotides of DNA before handoff of the completed primer to a processive replicative polymerase.

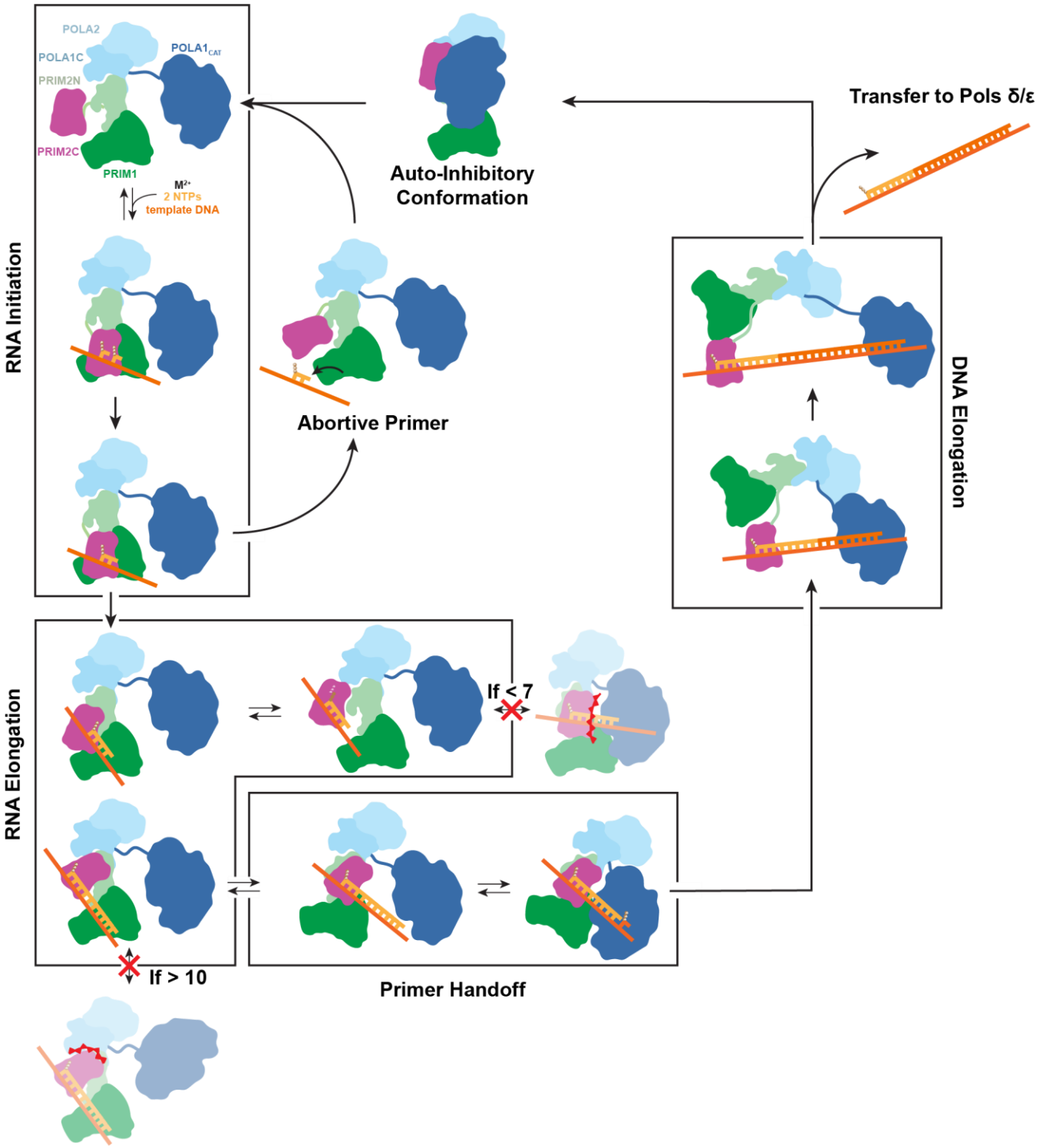

Figure 4.2 Schematic of chimeric RNA/DNA primer synthesis by pol-prim.

Future Directions

Expanding on current structural data for RNA priming

Although the data described in this thesis are of sufficient detail to gain insight into structural mechanisms, further work could be performed to improve upon these results as several key questions remain unanswered. First, what drives the transition from the autoinhibitory conformation to the unfolded conformation in which PRIM2C and PRIM1 are both capable of binding the nucleotide substrate? Second, what interactions, if any, stabilize the RNA initiation configuration, and how are the incoming nucleotides oriented in the PRIM1/PRIM2C binding pocket? Finally, direct evidence of steric hindrance between PRIM2C and the tetramer core as the RNA primer is elongated has not yet been obtained. Several improvements upon the current work could be employed to begin to address these gaps.

First, although cryoEM data have been collected for the Xenopus polΔcat-prim complex in the presence of RNA elongation and RNA initiation substrates (Chapter II), there remains significant room for improvement that could ultimately result in high-resolution reconstructions of pol-prim bound to RNA priming substrates and shed further light upon the interactions that drive structural re-arrangement during RNA priming. For both substrates, the small size and flexibility of the constituent domains of the complex were significant limitations for collecting high-contrast micrographs, particle picking and classification, and 3D flexibility analysis. Several strategies could be implemented to partially address these limitations. The first could be to develop a nanobody fiducial marker targeted to the POLA2/POLA1C lobe to 1) increase the overall size of the particle and 2) facilitate automatic alignment of the particles by providing a highly distinguishable feature. Furthermore, this could aid particle-picking by increasing the size

of and providing an identifying feature for end-view particles that were previously challenging to classify due to the elongated, bilobal shape of polΔcat-prim.

The second strategy would be to optimize sample preparation and grid plunging conditions for crosslinked protein-substrate species. Crosslinking greatly increased the homogeneity of polΔcat-prim in negative stain samples, but re-embarking upon ice optimization after crosslinking conditions were established was not necessary to obtain the information required for this study. Nevertheless, crosslinking could push polΔcat-prim into a series of discrete configurations that are easier to classify and reconstruct than the continuous flexibility observed in the non-crosslinked samples. Similarly, additional insights could be obtained from negative stain or cryoEM data of crosslinked polΔcat-prim in the presence of initiation substrates. This could further illuminate how RNA primer initiation occurs and provide more direct data as to the distribution of configurations and more clearly define the population in which PRIM2C is oriented near PRIM1 poised for initiation.

Due to the rapid development of new, powerful 3D classification and flexibility analysis algorithms, more effective analysis of the high degree of flexibility for PRIM2C and PRIM1 is to be expected and would ultimately provide higher-resolution structures and trajectories of motions. At present, 3D reconstructions from my cryoEM data are frustratingly low resolution and unable to delineate various configurations, but two factors suggest that this may be an addressable problem. The first is that higher resolution features are visible in the tetramer core from 2D class averages of the protein, suggesting that the data are of sufficient quality to achieve higher resolution and is limited only by the flexibility of the system. The second is that PRIM2C is clearly blurred in 2D class averages, and is seen in negative stain reconstructions to occupy a wide range of orientations, but all attempted 3D flexibility analysis methods have thus far been

unable to differentiate various configurations of PRIM2C. There remains the possibility that recently developed methods could overcome this limitation (98), or that more powerful algorithms that have yet to be developed could extract more information from these data.

In addition to potential for improvement in the current EM data, more direct characterization of the nature of the flexibility of PRIM2C during RNA priming could be highly informative, providing further support for the proposed dissociative mechanism of RNA primer length regulation. One approach to characterizing this flexibility could be to take advantage of the paramagnetic center in PRIM2C to perform Electron Paragmegnetic Resonance (EPR) experiments or to detect the Paramagnetic Relaxation Enhancement (PRE) effect by NMR. In an EPR experiment, one could develop a spin labeling approach to label PRIM1 with a paramagnetic center that could sense the paramagnetic center in PRIM2C. This could then be used to describe the distance distribution between these two domains. These distances could then be measured for the free protein, in the presence of RNA initiation factors, and RNA elongation substrates to determine if there are shifts in the distance distribution commensurate with a population shift from a highly flexible PRIM2C domain to a more compacted orientation in the presence of RNA initiation substrates, to a less flexible state for an RNA elongation complex. Similarly for a PRE experiment, the paramagnetic effect from the 4Fe-4S cluster in PRIM2C could be determined in all three states (free, initiation, elongation), detecting a decrease in signal in PRIM1 chemical shifts as PRIM2C shifts to a configuration that is closer to the active site of PRIM1 after substrate binding. Alternatively, a spin label could be introduced to either the PRIM1 or PRIM2C domains if the paramagnetic effect from the 4Fe-4S center is too weak, or even to RNA/DNA substrate to detect if there is a significant population of PRIM1 that is bound to the substrate in solution. A similar approach could be employed using a Fluorescence

Resonance Energy Transfer (FRET) experiment. However, in most cases, this strategy relies on the development of a labeling strategy that may prove challenging. As pol-prim is highly cysteine rich, a standard cysteine labeling strategy is not feasible. Instead, an amber codon/non-natural amino acid approach could be employed. A non-natural amino acid that incorporates a side chain for click chemistry could then be incorporated into the protein, after which a spin-label could be attached. One potential caveat is that many click chemistry reactions rely the absence of reducing agent or the presence of a copper catalyst, both of which may be problematic for pol-prim.

How does the presence of other Replication factors affect structure and mechanism of pol-prim

Perhaps the most exciting yet challenging avenue of future investigation will be to continue the investigation of coordination of priming in the context of other replication factors at the replisome. In recent publications in which structures of pol-prim in complex with the telomeric factor CST were determined, stabilization of the flexible architecture of pol-prim was observed while orienting and allowing for flexibility for the key PRIM1, PRIM2C, and POLA1cat domains(75). This scaffolding plays an important role at telomeres and results in increased priming activity by pol-prim(75,131). It has been noted that CST is evolutionarily related and structurally similar to RPA(42,75,132), and interaction between RPA and pol-prim has long been established](20,34). It remains unclear if RPA functions as merely a recruitment factor for pol-prim at the replication fork or if there is similarly intimate scaffolding and stabilization of pol-prim as in CST. Thus, an obvious next step will be to further investigate the interaction between RPA and pol-prim, specifically to determine the effect of RPA on the structural dynamics of pol-prim. Importantly, the proposed mechanism in this thesis relies on significant domain flexibility and distributive, dissociative mechanism for primer synthesis.

Though elegant in that it reconciles the currently available structural and biochemical data of pol-prim, the ready dissociation of PRIM1 during primer synthesis seems as though it could be problematic in the context of the replisome in which many factors must be carefully coordinated across the leading and lagging strand. The role of PRIM2C as a processivity factor would this be even more critical. Neverthess, the majority of the structural and biochemical data that have informed this mechanism have been collected in the absence of RPA. Could RPA have a role in orienting the domains of pol-prim to facilitate primer initiation by PRIM1? Discourage dissociation during RNA elongation by PRIM1? Facilitate primer handoff? Biochemical assays and structural investigations of pol-prim in the presence of RPA will be crucial in addressing these questions. A potential obstacle for these investigations is the highly flexible nature of both pol-prim and RPA, which could prove challenging for cryoEM investigations.

Finally, how the activity of pol-prim is coordinated within the larger replisome remains poorly understood. Although many interaction partners for pol-prim have been established, it remains unclear how these interactions are organized spatially and temporally. In addition to a possible interaction with RPA, recently published structures have indicated that pol-prim is closely associated with Ctf4 and by extension the CMG helicase in yeast (26). As the interaction between Ctf4 and yeast pol-prim is significantly different than for the human AND-1 pol-prim interaction, structural investigation of the human CMG/AND-1/pol-prim complex is needed to determine the organization of this core component of the replisome. Recently, the Yeeles group successfully reconstituted a model human replisome in which two important findings relating to pol-prim were presented (133,134). The first is that pol-prim commigrates on a gel with CMG even in the absence of AND-1, seemingly circumventing the interaction between AND-1 and pol-prim that has been established. Relatedly, the presence of pol-prim (both with and without

AND-1) resulted in significantly increased lagging strand synthesis. It will be important to further investigate the structural organization of pol-prim at such a model replisome structure. As the nature of the interaction between pol-prim and Ctf4/AND-1/CMG and pol-prim is likely quite flexible due to it involving the non-structured extension of POLA1 in yeast or the unstructured N-terminal domain of POLA2 in human, accommodation of a scaffolding partner for pol-prim such as RPA could be conceivable. In addition to the architectural organization pol-prim and factors at the replisome, it remains unclear how repeated priming of the lagging strand is spatially facilitated, in addition to priming of both the leading and lagging strand at the beginning of DNA synthesis immediately after the parent duplex has been unwound. These gaps in knowledge all require significant progress in the investigation of structural dynamics of the replisome.

Concluding Remarks

Although pol-prim was the first human polymerase described and has been known and studied for almost a half-century(6), fundamental gaps in knowledge remain in understanding its structure, mechanism, and placement within the replisome. With the development of novel cryoEM techniques, the last two years have seen a reinvigoration of investigation of pol-prim. Newly developed methods that are capable of determining high-resolution structures of large multiprotein complexes, as well as ongoing development of deep-learning based algorithms that are able to determine the structure and motion of even highly flexible proteins, have and will continue to enable the structural investigation of increasingly complex systems such as the eukaryotic replisome.

In spite of these developments, there remains a middle-ground in scale for which recently developed techniques have not proven to be sufficient for structural investigation due to the

small-size and high degree of flexibility of multiple subcomponents of pol-prim. This work has employed an integrative approach to address a series of questions relating to the coordination of RNA priming specifically. From this work and in combination with many recent structural studies, we now have a complete understanding of how pol-prim is able to tightly regulate chimeric RNA/DNA primer synthesis across multiple active sites. The stage is now set to, rather than break down the problem into separate pieces, begin to build up an understanding of the role of pol-prim in the complex multi-protein machinery that is responsible for duplication of the genome.

REFERENCES

1. Burgers PMJ, Kunkel TA. Eukaryotic DNA Replication Fork. 2017 [cited 2019 Jan 7]; Available from: https://doi.org/10.1146/annurev-biochem-

2. Lujan SA, Williams JS, Kunkel TA. DNA Polymerases Divide the Labor of Genome Replication. Trends Cell Biol [Internet]. 2016 Sep 1 [cited 2019 Jan 20];26(9):640–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27262731

3. Jain R, Aggarwal AK, Rechkoblit O. Eukaryotic DNA polymerases. Curr Opin Struct Biol [Internet]. 2018 Dec 1 [cited 2019 Apr 16];53:77–87. Available from: https://www.sciencedirect.com/science/article/pii/S0959440X18300083#fig0005

4. Pavlov YI, Shcherbakova P V. DNA polymerases at the eukaryotic fork—20 years later. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis [Internet]. 2010 Mar 1 [cited 2019 Jan 20];685(1–2):45–53. Available from: https://www.sciencedirect.com/science/article/pii/S0027510709002486?via%3Dihub

5. Garg P, Burgers PMJ. DNA polymerases that propagate the eukaryotic DNA replication fork. Vol. 40, Critical Reviews in Biochemistry and Molecular Biology. 2005. p. 115–28.

6. Spadari S, Weissbach A. RNA-primed DNA synthesis: specific catalysis by HeLa cell DNA polymerase alpha. Proceedings of the National Academy of Sciences [Internet]. 1975 Feb 1 [cited 2023 Jul 26];72(2):503–7. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.72.2.503

7. Ogawa T, Okazaki T. Discontinuous DNA Replication [Internet]. Vol. 49, Ann. Rev. Biochem. 1980 [cited 2019 Jun 12]. Available from: www.annualreviews.org

8. Garbacz MA, Lujan SA, Burkholder AB, Cox PB, Wu Q, Zhou ZX, et al. Evidence that DNA polymerase δ contributes to initiating leading strand DNA replication in Saccharomyces cerevisiae. Nat Commun [Internet]. 2018 Dec 27 [cited 2019 Jan 30];9(1):858. Available from: http://www.nature.com/articles/s41467-018-03270-4

9. Zhou ZX, Lujan SA, Burkholder AB, Garbacz MA, Kunkel TA. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. Nature Communications 2019 10:1 [Internet]. 2019 Sep 5 [cited 2023 Jul 26];10(1):1–10. Available from: https://www.nature.com/articles/s41467-019-11995-z

10. Moldovan GL, Pfander B, Jentsch S. PCNA, the Maestro of the Replication Fork. Cell. 2007 May 18;129(4):665–79.

11. Stith CM, Sterling J, Resnick MA, Gordenin DA, Burgers PM. Flexibility of Eukaryotic Okazaki Fragment Maturation through Regulated Strand Displacement Synthesis. Journal of Biological Chemistry. 2008 Dec 5;283(49):34129–40.

12. Maga G, Villani G, Tillement V, Stucki M, Locatelli GA, Frouin I, et al. Okazaki fragment processing: Modulation of the strand displacement activity of DNA polymerase δ by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. Proc Natl Acad Sci U S A [Internet]. 2001 Dec 4 [cited 2023 Jul

26];98(25):14298–303. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.251193198

13. Sun J, Shi Y, Georgescu RE, Yuan Z, Chait BT, Li H, et al. The architecture of a eukaryotic replisome. Nat Struct Mol Biol [Internet]. 2015 Dec 2 [cited 2019 Jan 20];22(12):976–82. Available from: http://www.nature.com/articles/nsmb.3113

14. Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX. Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell [Internet]. 2009 Nov 13 [cited 2019 Jan 29];139(4):719–30. Available from: https://www.sciencedirect.com/science/article/pii/S0092867409013038?via%3Dihub

15. Rzechorzek NJ, Hardwick SW, Jatikusumo VA, Chirgadze DY, Pellegrini L. CryoEM structures of human CMG-ATPS-DNA and CMG-AND-1 complexes. Nucleic Acids Res [Internet]. 2020 [cited 2023 Jul 9];48(12):6980–95. Available from: https://academic.oup.com/nar/article/48/12/6980/5846034

16. Li H, O'Donnell ME. The Eukaryotic CMG Helicase at the Replication Fork: Emerging Architecture Reveals an Unexpected Mechanism. BioEssays [Internet]. 2018 Mar 1 [cited 2019 Jan 20];40(3):1700208. Available from: http://doi.wiley.com/10.1002/bies.201700208

17. Goswami P, Abid Ali F, Douglas ME, Locke J, Purkiss A, Janska A, et al. Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. [cited 2019 Feb 26]; Available from: www.nature.com/naturecommunications

18. Pellegrini L. Dual Roles of Ctf18-RFC: Loading the Clamp and Angling for the Polymerase. Structure [Internet]. 2018 Jan 2 [cited 2018 Dec 19];26(1):1–2. Available from: https://www.sciencedirect.com/science/article/pii/S0969212617304082?via%3Dihub

19. Grabarczyk DB, Silkenat S, Kisker C. Structural Basis for the Recruitment of Ctf18-RFC to the Replisome. Structure [Internet]. 2018 Jan [cited 2019 Jan 15];26(1):137-144.e3. Available from: https://linkinghub.elsevier.com/retrieve/pii/S096921261730357X

20. Braun KA, Lao Y, He Z, Ingles CJ, Wold MS. Role of protein-protein interactions in the function of replication protein a (RPA): RPA modulates the activity of DNA polymerase α by multiple mechanisms. Biochemistry [Internet]. 1997 Jul 15 [cited 2023 May 29];36(28):8443–54. Available from: https://pubs.acs.org/doi/full/10.1021/bi970473r

21. Iftode C, Daniely Y, Borowiec JA. Replication Protein A (RPA): The Eukaryotic SSB. http://dx.doi.org/101080/10409239991209255 [Internet]. 2008 [cited 2023 Jul 26];34(3):141–80. Available from: https://www.tandfonline.com/doi/abs/10.1080/10409239991209255

22. Zou Y, Liu Y, Wu X, Shell SM. Functions of human replication protein A (RPA): From DNA replication to DNA damage and stress responses. J Cell Physiol [Internet]. 2006 Aug 1 [cited 2023 Jul 26];208(2):267–73. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.20622

23. Brosey CA, Yan C, Tsutakawa SE, Heller WT, Rambo RP, Tainer JA, et al. A new structural framework for integrating replication protein A into DNA processing machinery. Nucleic Acids Res. 2013 Feb;41(4):2313–27.

24. Kilkenny ML, Simon AC, Mainwaring J, Wirthensohn D, Holzer S, Pellegrini L. The human CTF4-orthologue AND-1 interacts with DNA polymerase a/primase via its unique C-Terminal HMG box. J R Soc Interface. 2017 Nov 1;7(136).

25. Bermudez VP, Farina A, Tappin I, Hurwitz J. Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication. J Biol Chem [Internet]. 2010 Mar 26 [cited 2019 Jan 20];285(13):9493–505. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20089864

26. Yuan Z, Georgescu R, Santos R de LA, Zhang D, Bai L, Yao NY, et al. Ctf4 organizes sister replisomes and Pol α into a replication factory. Elife. 2019;8:1–26.

27. Simon AC, Zhou JC, Perera RL, van Deursen F, Evrin C, Ivanova ME, et al. A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. Nature [Internet]. 2014 Jun 4 [cited 2019 Jan 9];510(7504):293–7. Available from: http://www.nature.com/articles/nature13234

28. Villa F, Simon AC, Ortiz Bazan MA, Kilkenny ML, Wirthensohn D, Wightman M, et al. Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. Mol Cell [Internet]. 2016 Aug 4 [cited 2019 Jan 9];63(3):385–96. Available from: https://www.sciencedirect.com/science/article/pii/S1097276516302404?via%3Dihub

29. Brosh RM, Trakselis MA. Fine-tuning of the replisome: Mcm10 regulates fork progression and regression. https://doi.org/101080/1538410120191609833 [Internet]. 2019 May 19 [cited 2023 Jul 9];18(10):1047–55. Available from: https://www.tandfonline.com/doi/abs/10.1080/15384101.2019.1609833

30. Grabarczyk DB. The Fork Protection Complex: A Regulatory Hub at the Head of the Replisome. Subcell Biochem [Internet]. 2022 [cited 2023 Jul 26];99:83–107. Available from: https://pubmed.ncbi.nlm.nih.gov/36151374/

31. Warren EM, Huang H, Fanning E, Chazin WJ, Eichman BF. Physical Interactions between Mcm10, DNA, and DNA Polymerase α. Journal of Biological Chemistry. 2009 Sep 4;284(36):24662–72.

32. Cho WH, Kang YH, An YY, Tappin I, Hurwitz J, Lee JK. Human Tim-Tipin complex affects the biochemical properties of the replicative DNA helicase and DNA polymerases. Proc Natl Acad Sci U S A [Internet]. 2013 Feb 12 [cited 2023 Jul 26];110(7):2523–7. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1222494110

33. De Falco M, Ferrari E, De Felice M, Rossi M, Hübscher U, Pisani FM. The human GINS complex binds to and specifically stimulates human DNA polymerase α-primase. EMBO Rep [Internet]. 2007 Jan 1 [cited 2023 Jul 26];8(1):99–103. Available from: https://onlinelibrary.wiley.com/doi/full/10.1038/sj.embor.7400870

34. Dornreiter I, Erdile LF, Gilbert IU, von Winkler D, Kelly TJ, Fanning E. Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen. EMBO J [Internet]. 1992 Feb 1 [cited 2019 Jan 20];11(2):769–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1311258

35. Kuchta RD, Reid B, Chang LM. DNA primase. Processivity and the primase to polymerase alpha activity switch. J Biol Chem [Internet]. 1990 Sep 25 [cited 2019 May 29];265(27):16158–65. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2398049

36. Plevani P, Foiani M, Falconi MM, Pizzagalli A, Santocanale C, Francesconi S, et al. The yeast DNA polymerase-primase complex: Genes and proteins. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 1988 Dec 20;951(2–3):268–73.

37. Zerbe LK, Kuchta RD. The p58 Subunit of Human DNA Primase Is Important for Primer Initiation, Elongation, and Counting †. 2002 [cited 2019 Jan 27]; Available from: https://pubs.acs.org/sharingguidelines

38. Grosse F, Krauss G. The primase activity of DNA polymerase alpha from calf thymus. Journal of Biological Chemistry. 1985 Feb 10;260(3):1881–8.

39. Zhang Y, Baranovskiy AG, Tahirov TH, Pavlov YI. The C-terminal domain of the DNA polymerase catalytic subunit regulates the primase and polymerase activities of the human DNA polymerase α-primase complex. J Biol Chem [Internet]. 2014 Aug 8 [cited 2019 Jan 20];289(32):22021–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24962573

40. Perera RL, Torella R, Klinge S, Kilkenny ML, Maman JD, Pellegrini L. Mechanism for priming DNA synthesis by yeast DNA Polymerase α. Elife [Internet]. 2013 Apr 2 [cited 2019 Jan 28];2013(2). Available from: https://elifesciences.org/articles/00482

41. Zhang Y, Baranovskiy AG, Tahirov ET, Tahirov TH, Pavlov YI. Divalent ions attenuate DNA synthesis by human DNA polymerase α by changing the structure of the template/primer or by perturbing the polymerase reaction. DNA Repair (Amst) [Internet]. 2016 Jul 1 [cited 2018 Dec 19];43:24–33. Available from: https://www.sciencedirect.com/science/article/pii/S1568786416300611?via%3Dihub

42. Mullins EA, Salay LE, Durie CL, Jackman JE, Chazin WJ, Eichman BF. A mechanistic model of primer synthesis from catalytic structures of DNA polymerase α–primase. bioRxiv. 2023;

43. Baranovskiy AG, Lisova AE, Morstadt LM, Babayeva ND, Tahirov TH. Insight into RNA–DNA primer length counting by human primosome. Nucleic Acids Res [Internet]. 2022 Jun 24 [cited 2023 May 27];50(11):6264–70. Available from: https://academic.oup.com/nar/article/50/11/6264/6605887

44. Voitenleitner C, Fanning E, Nasheuer HP. Phosphorylation of DNA polymerase α-primase by Cyclin A-dependent kinases regulates initiation of DNA replication in vitro. Oncogene 1997 14:13 [Internet]. 1997 [cited 2023 Jul 26];14(13):1611–5. Available from: https://www.nature.com/articles/1200975

45. Voitenleitner C, Rehfuess C, Hilmes M, O'Rear L, Liao PC, Gage DA, et al. Cell Cycle-Dependent Regulation of Human DNA Polymerase α-Primase Activity by Phosphorylation. https://doi.org/101128/MCB191646 [Internet]. 2023 Jan 1 [cited 2023 Jul 26];19(1):646–56. Available from: https://www.tandfonline.com/doi/abs/10.1128/MCB.19.1.646

46. Baranovskiy AG, Zhang Y, Suwa Y, Babayeva ND, Gu J, Pavlov YI, et al. Crystal structure of the human primase. J Biol Chem [Internet]. 2015 Feb 27 [cited 2018 Dec 19];290(9):5635–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25550159

47. Kuchta RD, Stengel G. Mechanism and evolution of DNA primases. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics [Internet]. 2010 May 1 [cited 2019 May 29];1804(5):1180–9. Available from: https://www.sciencedirect.com/science/article/pii/S1570963909001496?via%3Dihub

48. Lao-Sirieix SH, Nookala RK, Roversi P, Bell SD, Pellegrini L. Structure of the heterodimeric core primase. 2005 [cited 2019 Jan 27]; Available from: http://www.consurf.tau.ac.il

49. Iyer LM, Koonin E V, Leipe DD, Aravind L. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res [Internet]. 2005 [cited 2019 Jul 4];33(12):3875–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16027112

50. Vaithiyalingam S, Arnett DR, Aggarwal A, Eichman BF, Fanning E, Chazin WJ. Insights into Eukaryotic Primer Synthesis from Structures of the p48 Subunit of Human DNA Primase. J Mol Biol [Internet]. 2014 Feb 6 [cited 2018 Dec 19];426(3):558–69. Available from: https://www.sciencedirect.com/science/article/pii/S0022283613007183?via%3Dihub

51. Holzer S, Rzechorzek NJ, Short IR, Jenkyn-Bedford M, Pellegrini L, Kilkenny ML. Structural Basis for Inhibition of Human Primase by Arabinofuranosyl Nucleoside Analogues Fludarabine and Vidarabine. ACS Chem Biol [Internet]. 2019 Sep 20 [cited 2023 Jun 20];14(9):1904–12. Available from: https://pubs.acs.org/doi/full/10.1021/acschembio.9b00367

52. Kilkenny ML, Longo MA, Perera RL, Pellegrini L. Structures of human primase reveal design of nucleotide elongation site and mode of Pol α tethering. Proc Natl Acad Sci U S A [Internet]. 2013 Oct 1 [cited 2019 Jan 20];110(40):15961–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24043831

53. Makarova KS, Krupovic M, Koonin E V. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. Front Microbiol [Internet]. 2014 Jul 21 [cited 2019 Jul 4];5:354. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2014.00354/abstract

54. Holzer S, Yan J, Kilkenny ML, Bell SD, Pellegrini L. Primer synthesis by a eukaryotic-like archaeal primase is independent of its Fe-S cluster. [cited 2019 Feb 13]; Available from: www.nature.com/naturecommunications

55. Weiner BE, Huang H, Dattilo BM, Nilges MJ, Fanning E, Chazin WJ. An iron-sulfur cluster in the C-terminal domain of the p58 subunit of human DNA primase. J Biol Chem [Internet]. 2007 Nov 16 [cited 2018 Dec 19];282(46):33444–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17893144

56. Agarkar VB, Babayeva ND, Pavlov YI, Tahirov TH. Crystal structure of the C-terminal domain of human DNA primase large subunit Implications for the mechanism of the primase - Polymerase α switch. Cell Cycle. 2011 Mar 15;10(6):926–31.

57. Klinge S, Hirst J, Maman JD, Krude T, Pellegrini L. An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. Nat Struct Mol Biol [Internet].

2007 Sep 19 [cited 2019 Jan 27];14(9):875–7. Available from: http://www.nature.com/articles/nsmb1288

58. Baranovskiy AG, Babayeva ND, Zhang Y, Gu J, Suwa Y, Pavlov YI, et al. Mechanism of Concerted RNA-DNA Primer Synthesis by the Human Primosome. J Biol Chem [Internet]. 2016 May 6 [cited 2019 Jan 27];291(19):10006–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26975377

59. Coloma J, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Human DNA polymerase α in binary complex with a DNA:DNA template-primer. Nature Publishing Group [Internet]. 2016 [cited 2019 Jun 6]; Available from: www.nature.com/scientificreports/

60. Baranovskiy AG, Duong VN, Babayeva ND, Zhang Y, Pavlov YI, Anderson KS, et al. Activity and fidelity of human DNA polymerase α depend on primer structure. J Biol Chem [Internet]. 2018 May 4 [cited 2018 Dec 19];293(18):6824–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29555682

61. Suwa Y, Gu J, Baranovskiy AG, Babayeva ND, Pavlov YI, Tahirov TH. Crystal Structure of the Human Pol α B Subunit in Complex with the C-terminal Domain of the Catalytic Subunit. J Biol Chem [Internet]. 2015 Jun 5 [cited 2018 Dec 19];290(23):14328–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25847248

62. Baranovskiy AG, Gu J, Babayeva ND, Kurinov I, Pavlov YI, Tahirov TH. Crystal structure of the human Polє B-subunit in complex with the C-terminal domain of the catalytic subunit. J Biol Chem [Internet]. 2017 Sep 22 [cited 2018 Dec 19];292(38):15717–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28747437

63. Núñez-Ramírez R, Klinge S, Sauguet L, Melero R, Recuero-Checa MA, Kilkenny M, et al. Flexible tethering of primase and DNA Pol α in the eukaryotic primosome. Nucleic Acids Res [Internet]. 2011 Oct 1 [cited 2019 Jan 7];39(18):8187–99. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr534

64. Klinge S, Núñez-Ramírez R, Llorca O, Pellegrini L. 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases. EMBO J [Internet]. 2009 Jul 8 [cited 2019 Jan 20];28(13):1978–87. Available from: http://emboj.embopress.org/content/28/13/1978.long

65. Copeland WC, Wang TS. Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. J Biol Chem [Internet]. 1993 Dec 15 [cited 2019 May 29];268(35):26179–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8253737

66. Sheaff RJ, Kuchta RD. Mechanism of Calf Thymus DNA Primase: Slow Initiation, Rapid Polymerization, and Intelligent Termination* * [Internet]. Vol. 32, Biochemistry. 1993 [cited 2019 Jan 27]. Available from: https://pubs.acs.org/sharingguidelines

67. Arezi B, Kirk BW, Copeland WC, Kuchta RD. Interactions of DNA with Human DNA Primase Monitored with Photoactivatable Cross-Linking Agents: Implications for the Role of the p58 Subunit †. Biochemistry [Internet]. 1999 [cited 2019 Jan 27];38:12899–907. Available from: https://pubs.acs.org/sharingguidelines

68. Baranovskiy AG, Zhang Y, Suwa Y, Gu J, Babayeva ND, Pavlov YI, et al. Insight into the Human DNA Primase Interaction with Template-Primer. J Biol Chem [Internet]. 2016 Feb 26 [cited 2018 Dec 19];291(9):4793–802. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26710848

69. Kirk BW, Kuchta RD. Arg304 of Human DNA Primase Is a Key Contributor to Catalysis and NTP Binding: Primase and the Family X Polymerases Share Significant Sequence Homology †. 1999 [cited 2019 Jan 27]; Available from: http://www.ncbi.nlm.nih.gov/

70. Yuan Z, Georgescu R, Li H, O'Donnell ME. Molecular choreography of primer synthesis by the eukaryotic Pol α-primase. Nature Communications 2023 14:1 [Internet]. 2023 Jun 21 [cited 2023 Jun 27];14(1):1–14. Available from: https://www.nature.com/articles/s41467-023-39441-1

71. Sheaff RJ, Kuchta RD, Ilsley D. Calf Thymus DNA Polymerase a-Primase: Communication and Primer-Template Movement between the Two Active Sites* [Internet]. Vol. 33, Biochemistry. 1994 [cited 2019 Jan 27]. Available from: https://pubs.acs.org/sharingguidelines

72. O'Brien E, Holt ME, Thompson MK, Salay LE, Ehlinger AC, Chazin WJ, et al. Response to Comments on "The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport". Science [Internet]. 2017 Jul 21 [cited 2018 Dec 12];357(6348):eaan2762. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28729485

73. O'Brien E, Holt ME, Thompson MK, Salay LE, Ehlinger AC, Chazin WJ, et al. The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport. Science (1979) [Internet]. 2017 Feb 24 [cited 2018 Dec 12];355(6327):eaag1789. Available from: http://science.sciencemag.org/content/355/6327/eaag1789.long

74. Teo RD, Rousseau BJG, Smithwick ER, Di Felice R, Beratan DN, Migliore A. Charge Transfer between [4Fe4S] Proteins and DNA Is Unidirectional: Implications for Biomolecular Signaling. Chem [Internet]. 2019 Jan [cited 2019 Jan 20];5(1):122–37. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2451929418304467

75. He Q, Lin X, Chavez BL, Agrawal S, Lusk BL, Lim CJ. Structures of the human CST-Polα–primase complex bound to telomere templates. Nature. 2022 Aug 25;608(7924):826–32.

76. He Q, Baranovskiy AG, Morstadt LM, Lisova AE, Babayeva ND, Lusk BL, et al. Structures of human primosome elongation complexes. Nature Structural & Molecular Biology 2023 30:5 [Internet]. 2023 Apr 17 [cited 2023 May 27];30(5):579–83. Available from: https://www.nature.com/articles/s41594-023-00971-3

77. Cai SW, Zinder JC, Svetlov V, Bush MW, Nudler E, Walz T, et al. Cryo-EM structure of the human CST–Polα/primase complex in a recruitment state. Nat Struct Mol Biol. 2022 Aug 1;29(8):813–9.

78. He Y, Song H, Chan H, Liu B, Wang Y, Sušac L, et al. Structure of Tetrahymena telomerase-bound CST with polymerase α-primase. Nature. 2022 Aug 25;608(7924):813–8.

79. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys [Internet]. 2007 Aug [cited 2023 Jun 24];40(3):191–285. Available from: https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/abs/xray-solution-scattering-saxs-combined-with-crystallography-and-computation-defining-accurate-macromolecular-structures-conformations-and-assemblies-in-solution/2B17B59A3997E519892223F78D51BA0E

80. Rosenberg DJ, Hura GL, Hammel M. Size exclusion chromatography coupled small angle X-ray scattering with tandem multiangle light scattering at the SIBYLS beamline. Methods Enzymol. 2022 Jan 1;677:191–219.

81. Dyer KN, Hammel M, Rambo RP, Tsutakawa SE, Rodic I, Classen S, et al. High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. Methods Mol Biol [Internet]. 2014 [cited 2019 Jun 12];1091:245–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24203338

82. Jacques DA, Trewhella J. Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. Protein Science [Internet]. 2010 Apr 1 [cited 2023 Jul 17];19(4):642–57. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/pro.351

83. Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers [Internet]. 2011 Aug [cited 2019 Jun 12];95(8):559–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21509745

84. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL, Tsutakawa SE, et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods. 2009;6(8):606–12.

85. Svergun DI. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. Biophys J. 1999 Jun 1;76(6):2879–86.

86. Grant TD. Ab initio electron density determination directly from solution scattering data. Nat Methods [Internet]. 2018 Mar 29 [cited 2019 Jun 12];15(3):191–3. Available from: http://www.nature.com/articles/nmeth.4581

87. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res [Internet]. 2016 Jul 8 [cited 2023 Jun 2];44(W1):W424–9. Available from: https://academic.oup.com/nar/article/44/W1/W424/2499365

88. Manalastas-Cantos K, Konarev P V., Hajizadeh NR, Kikhney AG, Petoukhov M V., Molodenskiy DS, et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. urn:issn:1600-5767 [Internet]. 2021 Feb 1 [cited 2023 Jun 2];54(1):343–55. Available from: https://scripts.iucr.org/cgi-bin/paper?ge5081

89. Cheng P, Peng J, Zhang Z. SAXS-Oriented Ensemble Refinement of Flexible Biomolecules. Biophys J. 2017 Apr 11;112(7):1295–301.

90. Kourkoutis LF, Plitzko JM, Baumeister W. Electron Microscopy of Biological Materials at the Nanometer Scale. https://doi.org/101146/annurev-matsci-070511-155004 [Internet]. 2012 Jul 13 [cited 2023 Jul 26];42:33–58. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-matsci-070511-155004

91. Franken LE, Grünewald K, Boekema EJ, Stuart MCA, Franken LE, Grünewald K, et al. A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. Small [Internet]. 2020 Apr 1 [cited 2023 Jul 26];16(14):1906198. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/smll.201906198

92. Cong Y, Ludtke SJ. Single Particle Analysis at High Resolution. Methods Enzymol. 2010 Jan 1;482(C):211–35.

93. Benjin X, Ling L. Developments, applications, and prospects of cryo-electron microscopy. Protein Science [Internet]. 2020 Apr 1 [cited 2023 Jul 26];29(4):872–82. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/pro.3805

94. Chua EYD, Mendez JH, Rapp M, Ilca SL, Tan YZ, Maruthi K, et al. Annual Review of Biochemistry Better, Faster, Cheaper: Recent Advances in Cryo-Electron Microscopy. 2022 [cited 2023 Jul 26]; Available from: https://doi.org/10.1146/annurev-biochem-032620-

95. Scarff CA, Fuller MJG, Thompson RF, Iadaza MG. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. JoVE (Journal of Visualized Experiments) [Internet]. 2018 Feb 6 [cited 2023 Jul 26];2018(132):e57199. Available from: https://www.jove.com/v/57199/variations-on-negative-stain-electron-microscopy-methods-tools-for

96. Burgess SA, Walker ML, Thirumurugan K, Trinick J, Knight PJ. Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. J Struct Biol. 2004 Sep 1;147(3):247–58.

97. Nakane T, Kimanius D, Lindahl E, Scheres SH. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. Elife [Internet]. 2018 Jun 1 [cited 2019 Feb 12];7. Available from: https://elifesciences.org/articles/36861

98. Punjani A, Fleet DJ. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. Nature Methods 2023 20:6 [Internet]. 2023 May 11 [cited 2023 Jul 26];20(6):860–70. Available from: https://www.nature.com/articles/s41592-023-01853-8

99. Punjani A, Fleet DJ. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. J Struct Biol. 2021 Jun 1;213(2):107702.

100. Zhong ED, Bepler T, Berger B, Davis JH. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. Nature Methods 2021 18:2 [Internet]. 2021 Feb 4 [cited 2023 Jul 26];18(2):176–85. Available from: https://www.nature.com/articles/s41592-020-01049-4

101. Nethanel T, Reisfeld S, Dinter-Gottlieb G, Kaufmann G. An Okazaki Piece of Simian Virus 40 May Be Synthesized by Ligation of Shorter Precursor Chains it was proposed that the leading DNA chains are initiated at ori by DNA primase-pola and that their [Internet]. JOURNAL OF VIROLOGY. 1988. Available from: https://journals.asm.org/journal/jvi

102. Bullock PA, Soo Seo Y, Hurwitz J. Initiation of simian virus 40 DNA synthesis in vitro. Mol Cell Biol [Internet]. 1991 May 1 [cited 2023 Apr 23];11(5):2350–61. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=tmcb20

103. Eki T, Matsumoto T, Murakami Y, Hurwitz J. The replication of DNA containing the simian virus 40 origin by the monopolymerase and dipolymerase systems. Journal of Biological Chemistry. 1992;267(11):7284–94.

104. Arezi B, Kuchta RD. Eukaryotic DNA primase [Internet]. 2000 [cited 2019 Jan 27]. Available from: https://ac.els-cdn.com/S0968000400016807/1-s2.0-S0968000400016807-main.pdf?_tid=3d423034-ff63-45a6-808e-67afc9997c8e&acdnat=1548715121_9b44376af2dab5f1f49d7d14eb99ba70

105. Stodola JL, Burgers PM. Mechanism of Lagging-Strand DNA Replication in Eukaryotes. In Springer, Singapore; 2017 [cited 2018 Dec 12]. p. 117–33. Available from: http://link.springer.com/10.1007/978-981-10-6955-0_6

106. Pellegrini L. The Pol α-Primase Complex. In Springer, Dordrecht; 2012 [cited 2019 Jan 20]. p. 157–69. Available from: http://www.springerlink.com/index/10.1007/978-94-007-4572-8_9

107. Vaithiyalingam S, Warren EM, Eichman BF, Chazin WJ. Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase. Proc Natl Acad Sci U S A [Internet]. 2010 Aug 3 [cited 2018 Dec 19];107(31):13684–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20643958

108. Sauguet L, Klinge S, Perera RL, Maman JD, Pellegrini L. Shared active site architecture between the large subunit of eukaryotic primase and DNA photolyase. PLoS One. 2010;

109. Kilkenny ML, De Piccoli G, Perera RL, Labib K, Pellegrini L. A conserved motif in the C-terminal tail of DNA polymerase α tethers primase to the eukaryotic replisome. J Biol Chem [Internet]. 2012 Jul 6 [cited 2019 Jan 20];287(28):23740–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22593576

110. Baranovskiy A, Tahirov T, Baranovskiy AG, Tahirov TH. Elaborated Action of the Human Primosome. Genes (Basel) [Internet]. 2017 Feb 8 [cited 2018 Dec 19];8(2):62. Available from: http://www.mdpi.com/2073-4425/8/2/62

111. Kilkenny ML, Simon AC, Mainwaring J, Wirthensohn D, Holzer S, Pellegrini L. The human CTF4-orthologue AND-1 interacts with DNA polymerase a/primase via its unique C-Terminal HMG box. J R Soc Interface. 2017 Nov 1;14(136).

112. Uchiyama M, Wang TSF. The B-subunit of DNA polymerase alpha-primase associates with the origin recognition complex for initiation of DNA replication. Mol Cell Biol [Internet]. 2004 Sep 1 [cited 2019 Jan 20];24(17):7419–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15314153

113. Zhou B, Arnett DR, Yu X, Brewster A, Sowd GA, Xie CL, et al. Structural basis for the interaction of a hexameric replicative helicase with the regulatory subunit of human DNA polymerase α-primase. J Biol Chem [Internet]. 2012 Aug 3 [cited 2019 Jan 20];287(32):26854–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22700977

114. Frick DN, Richardson CC. DNA PRIMASES [Internet]. 2001 [cited 2019 Jan 27]. Available from: www.annualreviews.org

115. Thompson HC, Sheaff RJ, Kuchta RD. Interactions of calf thymus DNA polymerase α with primer/templates. Nucleic Acids Res [Internet]. 1995 Oct 25 [cited 2023 Apr 23];23(20):4109–15. Available from: https://academic.oup.com/nar/article/23/20/4109/2385463

116. Classen S, Hura GL, Holton JM, Rambo RP, Rodic I, McGuire PJ, et al. Implementation and performance of SIBYLS: A dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. J Appl Crystallogr [Internet]. 2013 Feb 17 [cited 2023 Jun 24];46(1):1–13. Available from: http://scripts.iucr.org/cgi-bin/paper?he5572

117. Förster S, Apostol L, Bras W. Scatter: Software for the analysis of nano-and mesoscale small-angle scattering. J Appl Crystallogr [Internet]. 2010 Apr 15 [cited 2023 Jun 2];43(3):639–46. Available from: //scripts.iucr.org/cgi-bin/paper?ce5075

118. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J [Internet]. 2013 Aug 20 [cited 2023 Jun 2];105(4):962–74. Available from: https://pubmed.ncbi.nlm.nih.gov/23972848/

119. Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Communications Biology 2019 2:1 [Internet]. 2019 Jun 19 [cited 2023 Jun 2];2(1):1–13. Available from: https://www.nature.com/articles/s42003-019-0437-z

120. Yang Z, Fang J, Chittuluru J, Asturias FJ, Penczek PA. Iterative Stable Alignment and Clustering of 2D Transmission Electron Microscope Images. Structure. 2012 Feb 8;20(2):237–47.

121. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods 2017 14:3 [Internet]. 2017 Feb 6 [cited 2023 Jun 2];14(3):290–6. Available from: https://www.nature.com/articles/nmeth.4169

122. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics [Internet]. 2016 Jun 1 [cited 2023 Jun 20];54(1):5.6.1-5.6.37. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cpbi.3

123. Fiser A, Do RKG, Šali A. Modeling of loops in protein structures. Protein Science. 2000 Jan;9(9):1753–73.

124. Schneidman-Duhovny D, Hammel M. Modeling structure and dynamics of protein complexes with SAXS profiles. Methods in Molecular Biology [Internet]. 2018 [cited 2023 Jun 20];1764:449–73. Available from: https://link.springer.com/protocol/10.1007/978-1-4939-7759-8_29

125. Baranovskiy AG, Babayeva ND, Zhang Y, Gu J, Suwa Y, Pavlov YI, et al. Mechanism of Concerted RNA-DNA Primer Synthesis by the Human Primosome. J Biol Chem [Internet].

2016 May 6 [cited 2018 Dec 19];291(19):10006–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26975377

126. Kilkenny ML, Longo MA, Perera RL, Pellegrini L. Structures of human primase reveal design of nucleotide elongation site and mode of Pol α tethering. [cited 2019 Jan 28]; Available from: www.pnas.org/cgi/doi/10.1073/pnas.1311185110

127. Arezi B, Kuchta RD. Eukaryotic DNA primase. Trends Biochem Sci [Internet]. 2000 Nov 1 [cited 2018 Dec 12];25(11):572–6. Available from: https://www.sciencedirect.com/science/article/pii/S0968000400016807?via%3Dihub

128. Kilkenny ML, Simon AC, Mainwaring J, Wirthensohn D, Holzer S, Pellegrini L. The human CTF4-orthologue AND-1 interacts with DNA polymerase α/primase via its unique C-terminal HMG box. Open Biol [Internet]. 2017 Nov 22 [cited 2018 Dec 12];7(11):170217. Available from: http://rsob.royalsocietypublishing.org/lookup/doi/10.1098/rsob.170217

129. Yin Z, Kilkenny ML, Ker DS, Pellegrini L. CryoEM insights into RNA primer synthesis by the human primosome. bioRxiv [Internet]. 2023 Jul 20 [cited 2023 Aug 20];2023.07.20.549859. Available from: https://www.biorxiv.org/content/10.1101/2023.07.20.549859v1

130. Yuan Z, Georgescu R, Li H, O'Donnell ME. Molecular choreography of primer synthesis by the eukaryotic Pol α-primase. Nature Communications 2023 14:1 [Internet]. 2023 Jun 21 [cited 2023 Aug 20];14(1):1–14. Available from: https://www.nature.com/articles/s41467-023-39441-1

131. Zaug AJ, Goodrich KJ, Song JJ, Sullivan AE, Cech TR. Reconstitution of a telomeric replicon organized by CST. Nature 2022 608:7924 [Internet]. 2022 Jul 13 [cited 2023 Jul 26];608(7924):819–25. Available from: https://www.nature.com/articles/s41586-022-04930-8

132. Lim CJ, Barbour AT, Zaug AJ, Goodrich KJ, McKay AE, Wuttke DS, et al. The structure of human CST reveals a decameric assembly bound to telomeric DNA. Science (1979) [Internet]. 2020 Jun 5 [cited 2023 Jul 26];368(6495):1081–5. Available from: https://www.science.org/doi/10.1126/science.aaz9649

133. Jones ML, Baris Y, Taylor MRG, Yeeles JTP. Structure of a human replisome shows the organisation and interactions of a DNA replication machine. EMBO J. 2021 Dec;40(23).

134. Baris Y, Taylor MRG, Aria V, Yeeles JTP. Fast and efficient DNA replication with purified human proteins. Nature 2022 606:7912 [Internet]. 2022 May 18 [cited 2023 Jul 26];606(7912):204–10. Available from: https://www.nature.com/articles/s41586-022-04759-1

CHAPTER 5. APPENDIX

BINDING BY CALMODULIN IS COUPLED TO TRANSIENT UNFOLDING OF THE THIRD FF DOMAIN OF PRP40A[3]

Summary

Human Pre-mRNA processing protein 40 homolog A (hPrp40A) is a splicing factor that interacts with the Huntington's disease protein huntingtin (Htt). Evidence has accumulated that both Htt and hPrp40A are modulated by the intracellular Ca^{2+} sensor calmodulin (CaM). Here we report characterization of the interaction of human CaM with the third FF domain (FF_3) of hPrp40A using calorimetric, fluorescence and structural approaches. Homology modeling, differential scanning calorimetry and small angle X-ray scattering (SAXS) data show FF_3 forms a folded globular domain. CaM was found to bind FF_3 in a Ca^{2+}-dependent manner with a 1:1 stoichiometry and a dissociation constant (K_d) of 25 ± 3 μM at 25 °C. NMR studies showed that both domains of CaM are engaged in binding and SAXS analysis of the FF_3-CaM complex revealed CaM occupies an extended configuration. Analysis of the FF_3 sequence showed that the anchors for CaM binding must be buried in its hydrophobic core, suggesting that binding to CaM requires unfolding of FF_3. Trp anchors were proposed based on sequence analysis and confirmed by intrinsic Trp fluorescence of FF_3 upon binding of CaM and substantial reductions in affinity for Trp-Ala FF_3 mutants. The consensus model of the complex showed that binding to CaM binding occurs to an extended, non-globular state of the FF_3, consistent with coupling to transient unfolding of the domain. The implications of these results are discussed in the context of the complex interplay of Ca^{2+} signaling and Ca^{2+} sensor proteins in modulating Prp40A-Htt function.

Introduction

[3] This chapter has been adapted from a co-first author publication in the journal *Protein Science* (2022) under the same title and is in partial fulfillment of the requirements of the Chemical and Physical Biology Program.

Pre-mRNA processing protein 40 (Prp40) is an essential RNA splicing factor with two homologs in humans, A and B (hPrp40A and hPrp40B) (1, 2). Prp40 proteins are composed of two WW domains (WW_1-WW_2) followed by six FF domains (FF_1-FF_6), with a flexible linker between the WW_2 and FF_1 (3). The tandem WW domains have been implicated in severe neurological diseases, including Huntington's disease (HD) (4-6) and Rett syndrome (7) on the basis of interactions with both huntingtin (Htt) and a disease-associated mutant of methyl-CpG-binding protein 2 (MeCP2), respectively. Besides the WW domains, Prp40 has been reported to interact with other proteins via its FF domains, including RNA splicing factors. The N-terminal crooked neck-like tetratricopeptide repeat (crn-TRP) of the splicing factor Clf1 interacts with yeast Prp40 (yPrp40) FF_1 during the early steps of the spliceosome assembly (8-10). Luc7 and Snu71 interact with yPrp40 FF domains simultaneously as components of the U1 small nuclear ribonucleoprotein (snRNP) (3, 11, 12). Our group showed that the C-terminal domain of the EF-hand Ca^{2+} sensor human centrin 2 (hCen2) interacts in a Ca^{2+}-dependent manner with a 33-residue truncated hPrp40A FF_3 construct containing the centrin-binding consensus motif $W_1xxL_4xxxL_8$ (13). This observation suggests that Prp40 function may be modulated by intracellular Ca^{2+} signaling.

Calmodulin (CaM), a ubiquitous EF-hand Ca^{2+} sensor closely related to centrin, has also been found to interact with Prp40; yeast CaM bound to yPrp40 in protein microarray experiments (14). Interestingly, human CaM has been shown to interact with Htt and has higher affinity for a disease-associated mutant than the wild-type (WT) protein (15-19). Moreover, CaM is known to play a role in regulating cross-linking of transglutaminase 2 to mutant Htt and these proteins co-localize in the brain of HD patients (16). On the other hand, centrin has also been found to colocalize with Htt at the centrosome and in the photoreceptor cilium (20, 21). These results

suggest that hPrp40A and Htt are part of a coordinated network regulated by Ca^{2+} signaling via centrin 2 and CaM.

Here we report the first biophysical and structural analysis of the interaction of CaM and hPrp40A, involving the same FF_3 domain that interacts with hCen2. The structure and stability of the isolated FF_3 was characterized by small angle X-ray scattering (SAXS) and differential scanning calorimetry (DSC), confirming the validity of a homology model. Isothermal titration calorimetry (ITC) and fluorescence spectroscopy were used to investigate the binding of the FF_3 to CaM, and the structure of the complex was probed by NMR and SAXS. These results are discussed in the context of the coupled unfolding and binding of the FF domain and the potential roles of Ca^{2+} modulation of the Htt-Prp40-Cen-CaM network in RNA splicing.

Materials and methods

Cloning, mutagenesis, expression and purification of hPrp40A FF3 domain

A dsDNA fragment of the hPrp40A FF3 domain (residues Ser516-Glu593) with an additional N-terminal Met for translation was synthesized (Strings DNA Fragments, Thermo Fisher) and subcloned in-frame into pBG100 (in-house pET-27 derived vector) using seamless cloning (In-Fusion HD Cloning Kit, Takara). To generate the Trp to Ala mutants, non-overlapping site-specific primers were designed to mutate the codons for Trp531 and Trp550 to Ala residues using site-directed mutagenesis (Q5 SDM Kit, NEB). Non-overlapping single-stranded DNA (ssDNA) primers were designed to mutate these sites using site-directed mutagenesis. Primers were obtained from IDT. Primers used included the following:

W531A Forward Primer: 5′-CAAACGTAACGCGGAAGCATTAAAAAATATTCTG-3′

W531A Reverse Primer: 5′-CGGAGTTGCTTTGCTTGT-3′

W550A Forward Primer: 5′-TAGCACCACCGCATCAGAAGCACAGC-3′

W550A Reverse Primer: 5′-TAGGTAACATTGGCCATATTATC-3′

To obtain the double mutant, pBG100-FF_3 W531A was used as a template and the W550A substitution was generated using the corresponding primers. Successful constructs were validated by Sanger sequencing (Genewiz).

Recombinant hPrp40A FF_3, comprised of the sequence from Ser_{516} to Glu_{593}, and 4FF3 W531A, W550A, and W531A/W550A variants were expressed in BL21 (DE3) cells with minor modifications to the previously reported protocol (22). The harvested pellet was lysed using cold buffer solution containing 20 mM Tris (pH 7.4), 100 mM NaCl, and 0.1% Nonidet P-40. To minimize proteolytic cleavage, one tablet of protease inhibitor cocktail (Roche) was added. The crude lysate was subjected to centrifugation using a JA 25.50 rotor at 20,000 rpm for one hour at 4 °C. Then, the supernatant was filtered using a 0.45 μm syringe filter.

For purification, $(NH_4)_2SO_4$ in fine powder form was slowly added to the sample with thorough mixing at room temperature until the sample reaches a 70% of saturation. At this percentage of saturation, hPrp40A FF_3 remains in solution while most of the proteins of the bacterial host cell were precipitated. The precipitant was removed from the soluble sample by centrifugation using a JA 25.50 rotor at 20,000 rpm for 30 minutes at 4 °C. The soluble sample was then extensively dialyzed against the buffer 20 mM Tris (pH 7.4) and 100 mM NaCl to remove the $(NH_4)_2SO_4$ from the sample. After dialysis, the sample was purified by DEAE Sepharose anion exchange chromatography. The column was equilibrated using 20 mM Tris (pH 7.4). A gradient of NaCl was used from 0 to 1 M with an isocratic step at 200 mM NaCl for the elution of the protein. Then, the pooled sample was filtered and concentrated using Amicon 30 kDa and 3 kDa centrifugal filters (Merck Millipore). The sample was then filtered using a 0.45 μm syringe filter

and further purified by Superdex S200 GL 10/300 SEC. The fractions were eluted in 50 mM HEPES (pH 7.4), 150 mM NaCl, 2 mM $CaCl_2$, and 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl).

Expression and purification of CaM

The expression and purification of CaM has been described elsewhere (35). *Escherichia coli* BL21 (DE3) cells were transformed using a pET-15b human CaM vector containing *CALM3* inserted between 5′-NcoI and 3′-BamHI restriction sites. The culture was grown in 6 L of 2x YT medium containing 100 μg/mL ampicillin at 230 rpm and 37 °C while monitoring its growth via optical density (OD) measurement at 600 nm. Once the OD measurement was 0.6-1.0, the bacterial cultures were induced with a final concentration of 0.5 mM isopropyl-β-thiogalactoside (IPTG). Cells were harvested after 4 hours of the induction by centrifugation using a JLA 8.1000 rotor at 6,500 rpm for 20 minutes at 4 °C.

For the expression of the ^{15}N-labelled CaM (^{15}N-CaM), the culture was grown in 2 L of ^{15}N-enriched M9 minimal media containing 50 μg/mL ampicillin at 230 rpm and 37 °C while monitoring its growth via OD measurement at 600 nm. Once the OD measurement was ~ 0.7, the cultures were induced with a final concentration of 0.5 mM IPTG and transferred to a room temperature incubator. Cells were harvested the next day by centrifugation using a JLA 8.1000 rotor at 6,500 rpm for 20 minutes at 4 °C.

For both CaM and ^{15}N-CaM, the harvested pellet was lysed using cold buffer solution containing 50 mM Tris (pH 7.5), 500 mM KCl, 1 mM EDTA, and 1 mM phenylmethylsulfonyl fluoride (PMSF). To minimize proteolytic cleavage, one tablet of protease inhibitor cocktail from Roche (Mannheim, Germany) was added. The crude lysate was subjected to centrifugation using a JA 25.50 rotor at 20,000 rpm for 30 minutes at 4 °C. Then, the supernatant was filtered using a

0.45 μm syringe filter. CaM was purified in one step using Phenyl Sepharose CL-4B chromatography, and the fractions containing the pure protein were pooled. In the case of ^{15}N-CaM, size exclusion chromatography (SEC) was added as a second purification step using a HiLoad 16/600 Superdex 75 pg column.

Protein characterization

The identity of all proteins was validated by SDS-PAGE and mass spectrometry. For each of the methods performed below, the protein concentrations were determined using a Jasco model Cary 100 UV/Vis spectrophotometer (Jasco Corporation, Tokyo, Japan). The calculated ε for CaM and the hPrp40A FF_3 are 2,980 $M^{-1}cm^{-1}$ and 13,980 $M^{-1}cm^{-1}$, respectively.

Isothermal Titration Calorimetry

ITC experiments were performed using a TA Instruments Affinity ITC. CaM and the hPrp40A FF_3 samples were simultaneously dialyzed in the same beaker against 50 mM HEPES (pH 7.4), 150 mM NaCl, 4 mM $CaCl_2$, and 1 mM TCEP-HCl or 100 mM β-mercaptoethanol. For analysis in the absence of Ca^{2+}, the samples were extensively dialyzed against 50 mM HEPES (pH 7.4), 150 mM NaCl, 2 mM EDTA, and 1 mM TCEP-HCl. Samples were passed through a 0.45 μM PVDF filter and degassed for at least 10 min before use. In a typical experiment with the WT proteins, 40-100 μM FF_3 in the sample cell was titrated with 300-600 μM CaM by automatic injection with volumes ranging from 2-4 μL. Titrations with the FF_3 mutants and controls with WT FF_3 were performed with an initial injection of 0.5 μL FF_3 followed by 30 x 3 μL injections at 250-400 second intervals. The data were fit with the 'One Set of Sites' model, using the instrument's NanoAnalyze software. Experiments were repeated in triplicate and the standard deviation is reported for each measurement.

Differential Scanning Calorimetry

A sample of hPrp40A FF_3 (60 μM) in 50 mM HEPES (pH 7.4), 150 mM NaCl, 4 mM $CaCl_2$, 4 $MgCl_2$, and 1 mM TCEP-HCl was analyzed using a MicroCal VP-DSC microcalorimeter. The protein sample and the reference were degassed for at least 10 min before use. Thermograms were collected at 25 psi and a scan rate of 60 °C/h over a temperature range of 10-95 °C, with a 16 s filtering period. The data analysis was performed using the instrument's Origin software. The thermogram was reference subtracted and a progressive baseline was obtained for accurate determination of the thermal denaturation mid-point (T_m).

Homology Modeling

Homology models were generated using Modeller v10.1 (23). The homology model of the CaM-FF_3 complex was generated by aligning the sequence of the FF_3 domain to the Munc13-1 peptide, then threading the sequence through the structure of the CaM-Munc13-1 complex. As a check on the model, we examined the fit to the experimental data for the CaM-FF_3 complex and found it is similar to that of the CaM-Munc13-1 complex (see Fig. 6C). This homology model also fit reasonably well into the *ab initio* electron map of the CaM-FF_3 complex generated from the SAXS data using DENSS (see Fig. 6D).

SEC-SAXS

The experiments were performed at the SIBYLS beamline using previously described protocols (24). To briefly summarize, CaM in its Ca^{2+} saturated state was combined with hPrp40A FF_3 in a 1:1.5 CaM:FF_3 ratio and concentrated using an Amicon 3 kDa centrifugal filter. All data were acquired at room temperature. The sample (10 mg/mL) was loaded onto a Shodex 802.5 SEC column connected to a multi-angle light scattering system and the CaM-FF_3 complex was eluted in 50 mM HEPES (pH 7.4), 100 mM NaCl, 2 mM $CaCl_2$, and 1 mM TCEP-HCl. Data were also collected for isolated CaM (8 mg/mL) and hPrp40A FF_3 (8 mg/mL) in the same buffer conditions.

The sample volume run through the SEC was 50 µl and the eluent was split between SEC-MALS and SAXS channels. 3.0-s X-ray exposures were collected continuously during a ~35 min elution. The SAXS frames recorded prior to the protein elution peak were used as buffer blanks to subtract from all other frames. The subtracted frames were examined by radius of gyration (R_g) and scattering intensity at $q = 0$ Å^{-1} ($I(0)$), derived using the Guinier approximation $I(q) = I(0)e^{-qR_g^2/3}$ with the limits $q*R_g < 1.5$. $I(0)$ and R_g values were compared for each collected SAXS curve across the entire elution peak. The elution peak was mapped by plotting the scattering intensity at $q = 0$ Å^{-1} ($I(0)$), relative to the recorded frame. Uniform R_g values across an elution peak represent a homogenous assembly. The merged experimental SAXS data were additionally investigated for aggregation by inspecting Guinier plots. The SAXS data were processed using Scatter 4.0. Back-calculated scattering data from previously determined structural coordinates were generated using the FoXS server (25, 26). The *ab initio* electron maps obtained from the SAXS data of CaM, hPrp40A FF_3, and the CaM-FF_3 complex were determined using DENSS (27). Similar results were obtained with other *ab initio* shape determination methods. The SAXS data and models have been deposited in the SASDB databank under accession codes 4338 (FF3), 4337 (CaM) amd 4324 (FF3-CaM).

NMR Spectroscopy

Two samples of 200 □M ^{15}N-enriched CaM in the absence and presence of 1.5 molar excess unlabeled hPrp40A FF_3 were prepared in a buffer containing 20 mM Bis-Tris (pH 6.75), 50 mM NaCl and 2 mM $CaCl_2$. 2D ^{15}N-^{1}H HSQC experiments were recorded in 5-mm tubes at 37 °C using a Bruker AVANCE 800 MHz spectrometer equipped with a TCI cryoprobe. Experiments on ^{15}N-enriched FF_3 were performed at a concentration of 50 µM in a buffer containing 50 mM HEPES pH 7.4, 150 mM NaCl, 4 mM $CaCl_2$ and 5 mM TCEP. Spectra were acquired in 3 mm

tubes at 25 °C. The pulse sequence hsqcetfpf3gpsi2 from the standard suite of Bruker pulse programs was used, and all spectra were acquired with 16 scans and 2048 in the direct dimension and 128 points in the indirect dimension. Data processing and analysis were carried out using Topspin, NMRPipe and CcpNmr version 2.4.2.

For experiments with the FF3 mutants, stock concentrations of ^{15}N-enriched CaM and FF_3 (WT or variants) were prepared by buffer exchanging into ITC Buffer [50 mM HEPES (pH 7.4), 150 mM NaCl, 4 mM $CaCl_2$, 5 mM TCEP] at 4 °C. The samples were concentrated to be $\geq$ 80 μM for ^{15}N-CaM and $\geq$ 80 μM FF_3 (WT or variants). Samples were prepared in a 1:2 with 50 μM ^{15}N-CaM, 100 μM FF_3, and 10% (v/v) D_2O in ITC Buffer [50 mM HEPES (pH 7.4), 150 mM NaCl, 4 mM $CaCl_2$, 5 mM TCEP]. Additional samples were prepared in a 1:20 ratio with 50 μM ^{15}N-CaM, 1000 μM FF_3, and 10% (v/v) D_2O. 2D ^{15}N-^{1}H HSQC experiments were recorded in 5-mm tubes at 37 °C using a Bruker AVANCE 900 MHz spectrometer equipped with a TCI cryoprobe. The spectra were acquired as above except that 32 scans were collected. Data processing and analysis were carried out using NMRPipe and CcpNmr version 2.5.1.

Fluorescence Spectroscopy

A sample of hPrp40A FF_3 (5 μM) in 50 mM HEPES (pH 7.4), 150 mM NaCl, 2 mM $CaCl_2$, and 1 mM TCEP-HCl was titrated with a solution of 100 μM CaM in the same buffer. The fluorescence emission of the two Trp residues in the FF_3 was measured at 37 °C using a Horiba Jobin Yvon (Edison, NJ) Fluoromax-3 fluorometer. The fluorescence excitation wavelength was 295 nm, and the fluorescence emission spectra were recorded between 300 and 500 nm, using quartz cuvettes.

RESULTS

CaM binds to hPrp40A FF_3 domain in a Ca^{2+}- and temperature-dependent manner

CaM and Cen2 are both EF-hand Ca^{2+} sensors and since hCen2 binds to the hPrp40A FF_3, we asked if CaM could also bind to this domain? To this end, ITC experiments were performed to examine binding of FF_3 to CaM in the absence and presence of Ca^{2+} (Fig. 1). The data show there is a Ca^{2+}-dependent interaction with a K_d value of 25 ± 3 μM and a stoichiometry of 1:1 (Fig. 1A and 1B), similar to that observed for a number of other CaM binding partners.

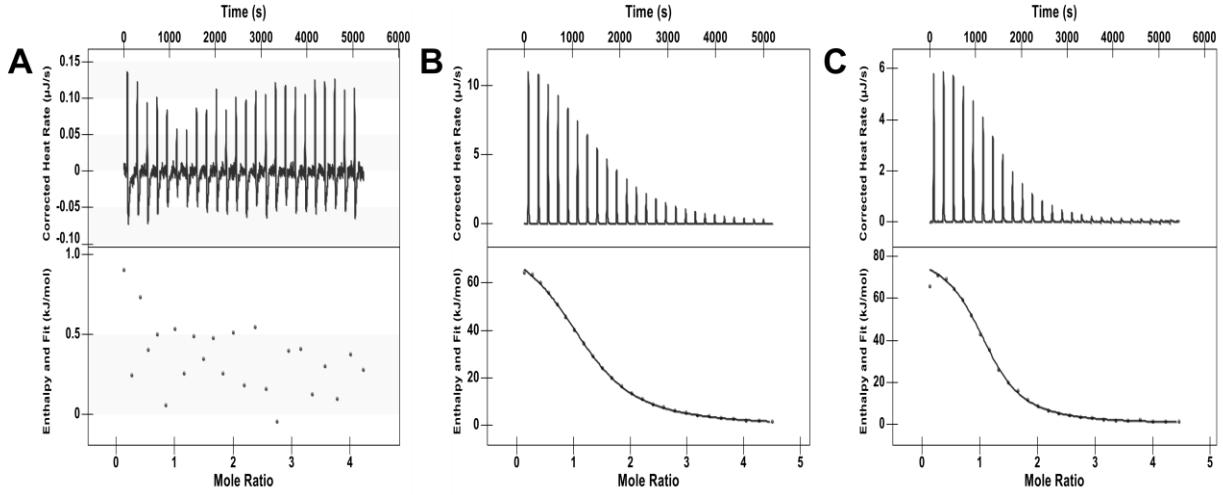

Figure 1. ITC analysis of the interaction of hPrp40A FF_3 with CaM.

(A) Titration of the hPrp40A FF_3 with (A) Ca^{2+}-free CaM, (B) Ca^{2+}-loaded CaM at 25 °C and (C) Ca^{2+}-loaded CaM at 37 °C.

Considering the relatively modest affinity at room temperature and knowing that CaM generally binds linear peptide motifs of 15-25 residues whereas FF domains are folded, we surmised that the binding by CaM of the FF_3 may require unfolding of the domain. To assess the potential for the domain to unfold, the T_m of the FF_3 was measured by DSC. A value of 50 °C was measured (Fig. S1), consistent with previously reported studies by circular dichroism and 2D IR correlation spectroscopy (22). The T_m of hPrp40A FF_3 is also similar to the T_m of other FF domains

reported in the literature, which include the p190 RhoGAP FF_1 (T_m = 52 °C) and the hTCERG1 FF_2 (T_m = 55 °C) (28) and hPrp40A FF_4 (T_m = 49 °C) domains (29). The T_m value of 50 °C for the FF_3 implies a relatively low equilibrium concentration of unfolded protein at 25 °C, consistent with the relatively weak binding measured by ITC. To obtain further insight, the ITC measurement was repeated at 37 °C (Fig. 1 C) where the equilibrium concentration of unfolded protein will be higher. As anticipated, a five-fold higher affinity (K_d of 4.9 ± 0.3 μM) was observed, consistent with binding to CaM occurring via the equilibrium population of unfolded FF_3. Further support was provided by the observation that a mixture of CaM and the FF_3 co-purify by SEC at room temperature but not at 4 °C where the equilibrium population of unfolded protein would be substantially lower (Fig. S2).

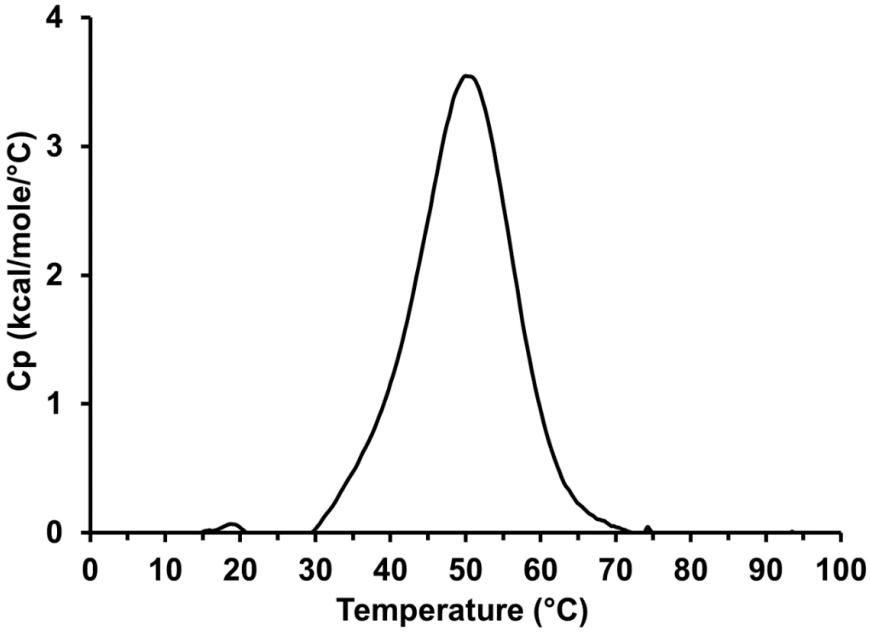

Fig. S1. DSC thermogram plot of hPrp40A FF_3. The T_m value is 50 °C.

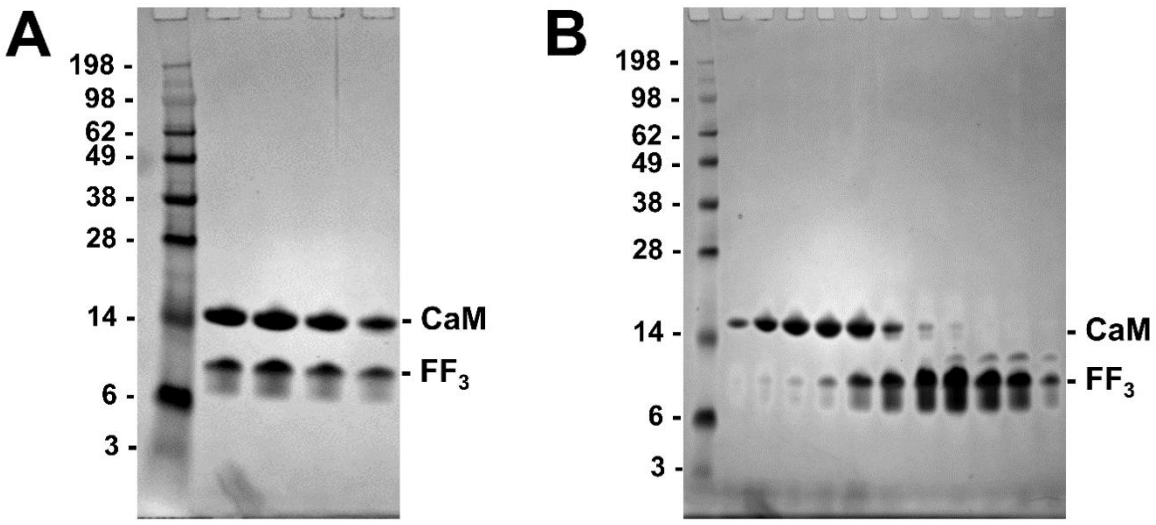

Fig. S2. Co-elution of CaM and hPrp40A FF_3 by SEC. SDS-PAGE results of eluted fractions from size exclusion chromatography at (A) room temperature and (B) 4 °C. The bands that correspond to CaM (16.7 kDa) and hPrp40A FF_3 (9.3 kDa) are identified. In each gel, the first lane (from left to right) corresponds to the SeeBlue Plus 2 Prestained molecular weight standard from Invitrogen (Carlsbad, CA)

Structure of hPrp40A FF_3

Structures have been determined for the FF_1 and FF_6 of hPrp40A and yPrp40 (10, 30-32), but despite extensive efforts, we were not able to obtain diffraction quality crystals for hPrp40A FF_3. Alignment of FF domain sequences from three different classes of proteins show there are eleven conserved residues that constitute the hydrophobic core of FF domains (Fig. 2). Of note, hPrp40A FF_3 has two non-consensus residues: the conserved Phe in α_1 is substituted by a Leu (Leu_{534}) and one of the conserved hydrophobic residues in the 3_{10}-helix is substituted by an Asp (Asp_{566}). Despite these differences, the high degree of sequence similarity overall meant that an accurate structural model could be generated using a homology modeling approach.

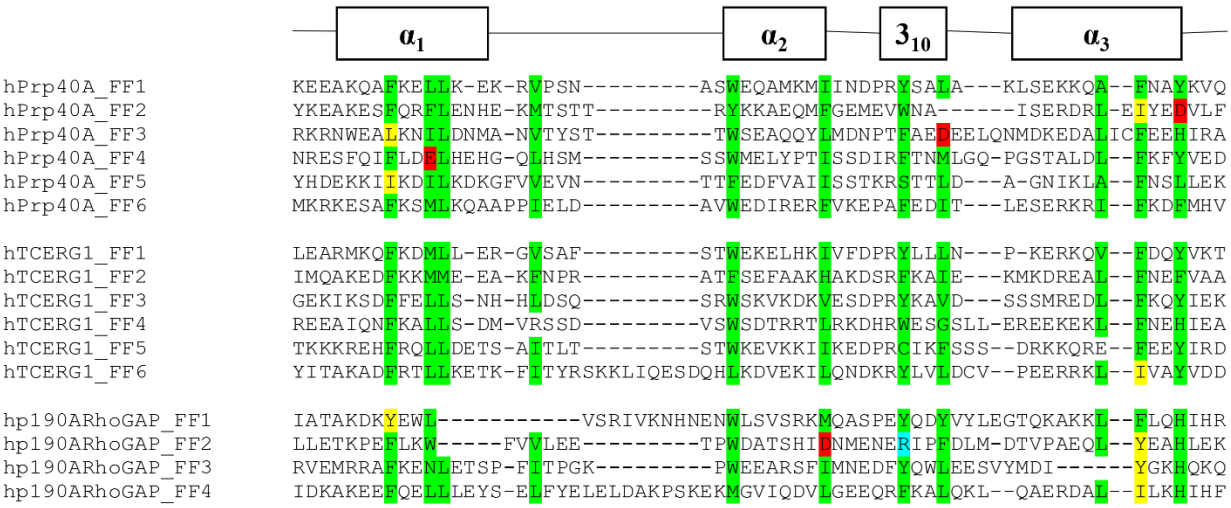

Figure 2. Sequence alignment of FF domains. FF domains from hPrp40A (Uniprot ID: O75400), transcription factor hTCERG1 (Uniprot ID: O14776), and Rho GTPase human p190A RhoGAP (hp190ARhoGAP) (Uniprot ID: Q9NRY4). Residues known to be part of the hydrophobic core of FF domains are highlighted in green. Substitutions of these conserved residues with hydrophobic, acidic, and basic residues are highlighted in yellow and red, and blue, respectively. The alignments were generated using the online tool Clustal Omega.

A homology model of hPrp40A FF_3 was generated using the QUARK server (33, 34) (Fig. 3A). Overlay of this model on the human transcription elongation regulator 1 (hTCERG1) FF_1 shows the two key 'F' residues in the FF_3 domain (Leu_{534} and Phe_{581}) are well aligned with the two conserved phenylalanine residues (Fig. 3B-D). Previous structural comparisons of FF domains revealed that the loop-3_{10}-loop region is the most variable (31). This is also the case for hPrp40A FF_3. The unique orientation of this loop is driven largely by the non-consensus residue Asp_{566}, which faces the solvent rather than inward towards the core as observed for the consensus Leu in hTCERG1 FF_1 (Fig. 3E).

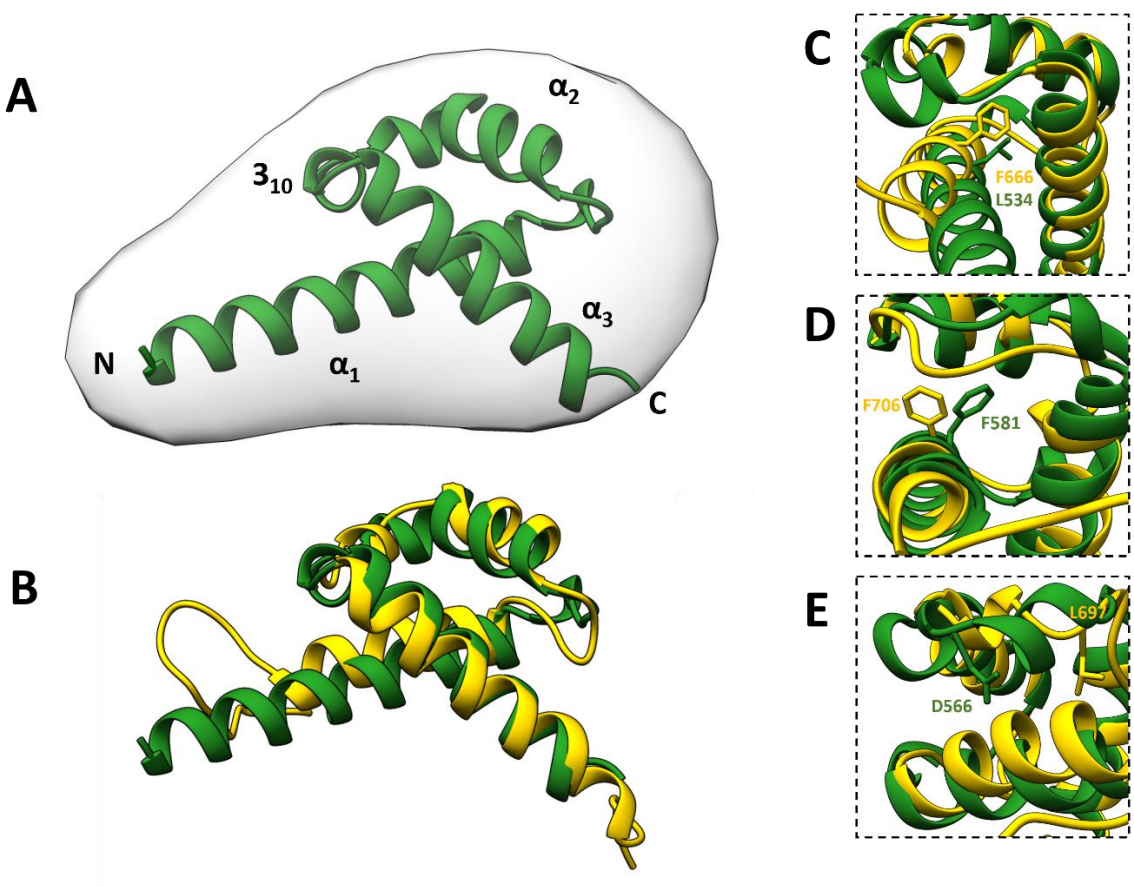

Figure 3. Structural analysis of hPrp40A FF_3. (A) Overlay of the 3D homology model (green) and the *ab initio* electron map of hPrp40A FF_3 generated by DENSS from the SAXS data. (B) Superposition of the homology model of the hPrp40A FF_3 and the NMR structure of hTERG1 FF_1 (PDB 2DOD, colored yellow). (C, D) Close-up view of the two conserved "FF domain" phenylalanine residues located in the middle of α1 and α3, respectively. (E) Close-up view of the non-conserved Asp_{566} residue in the loop-3_{10}-loop region of hPrp40A FF_3, compared to the conserved Leu_{697} in hTCERG1. Note that Asp_{566} faces the solvent whereas Leu_{697} extends into the hydrophobic core.

To assess the validity of the structural model, SAXS data were collected for isolated hPrp40A FF_3 (Fig. S3A-C; Table S1). The linearity observed in the Guinier region of the $\log_{10}$ intensity plot and Porod volumes are consistent with the sample being largely free of aggregation.

However, the ratio of the Porod volume to mass is slightly higher than expected, suggesting there may be a small amount of aggregate present. The Kratky plot and the Porod exponent (P_x) of 3.6 are indicative of a stable globular particle, but with some degree of flexibility retained, likely at the N- and/or C-terminus. The experimental probability distribution function P(r), which reflects the distribution of inter-atomic distances, was compared to the P(r) back-calculated from the structures of different FF domains. These show great similarity overall (Fig. 4), with the closest match to hTCERG1 FF_1 (PDB 2DOD; $\chi^2 = 1.97$). Thus, the SAXS analysis is consistent with the hPrp40A FF_3 possessing the common α_1-α_2-3_{10}-α_3 fold found in all FF domains. Further support is provided by the dispersion of signals observed in the ^{15}N-^{1}H NMR spectrum of the FF_3 domain (Fig. S4). The SAXS and NMR data confirm the validity of the homology model.

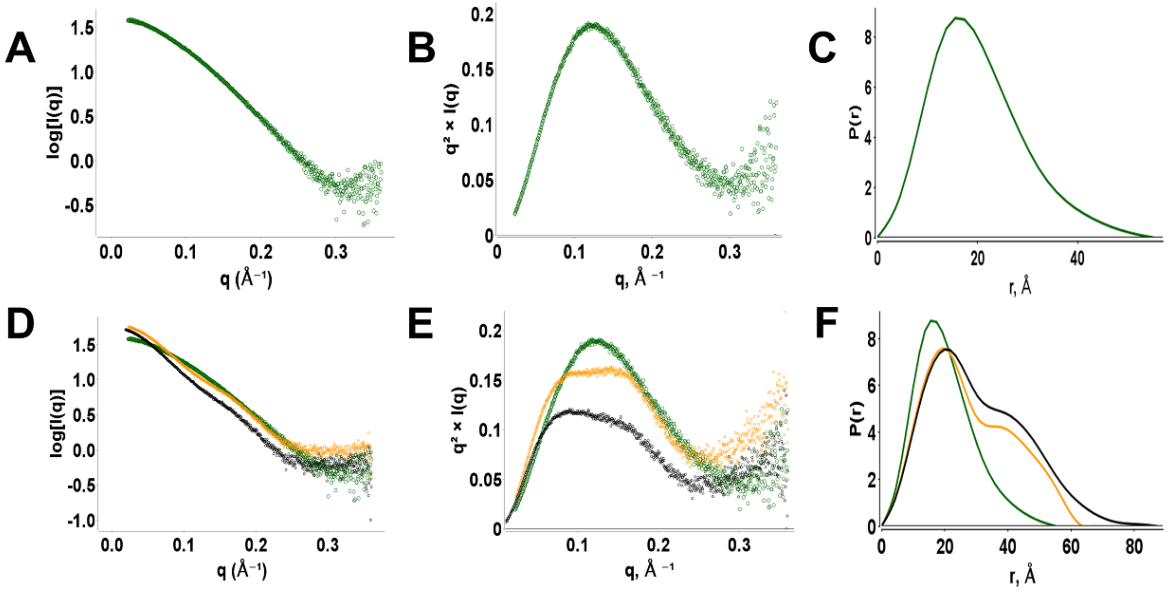

Fig. S3. SAXS data for the hPrp40A FF_3, CaM, and the CaM-FF_3 complex. (A-C) hPrp40A FF_3 $\log_{10}$ intensity, Kratky plot, and P(r), respectively. (D-F) Overlay of the $\log_{10}$ intensity, Kratky plot, and P(r) of hPrp40A FF_3 (green), CaM (orange), and the CaM-FF_3 complex (black).

Table S1. Summary of the parameters obtained from the SAXS analysis of the hPrp40A FF_3, CaM, and the CaM-FF_3 complex.

Protein	M.W. (kDa)	$R_{g\ Guinier}$ (Å)	$R_{g\ P(r)}$ (Å)	V_p (Å^3)	D_{max} (Å)	P_x
hPrp40A FF_3	9.3	**16.09**	**15.65**	**17574**	**55**	**3.7**
CaM	16.7	**21.66**	**21.99**	**28312**	**67.5**	**3.6**
CaM-FF_3 complex	26.0	**23.96**	**24.32**	**36273**	**87**	**3.7**

The values in bold were obtained from the SAXS analysis (See Fig. S3). $R_{g\ Guinier}$ and $R_{g\ P(r)}$ are the R_g values obtained from SAXS Guinier and P(r) plot analyses, respectively.

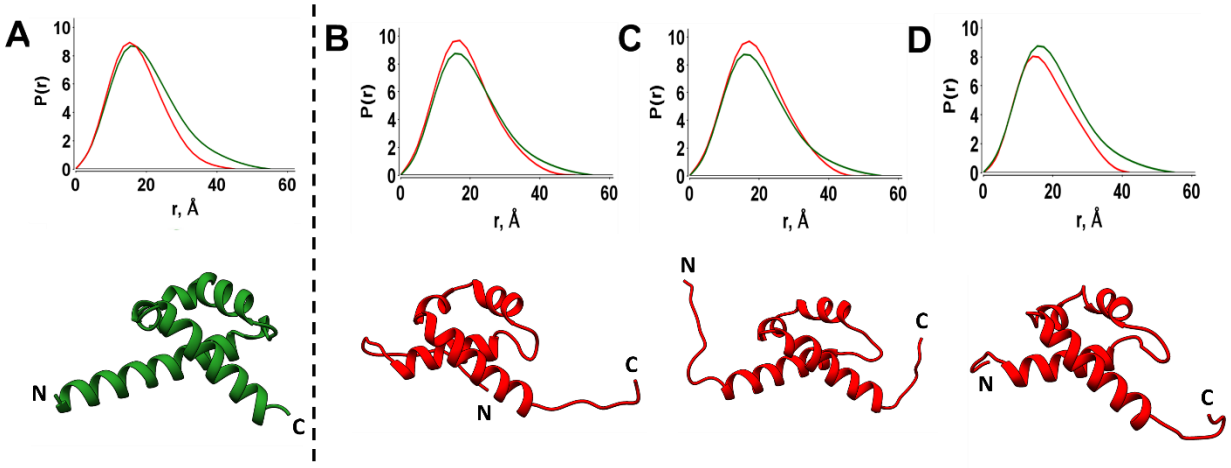

Figure 4. SAXS analysis of the structure of the hPrp40A FF_3 domain. Overlays of the I(q) vs q plots and P(r) derived from the SAXS data for FF_3 (green) versus the back-calculated P(r) (red) extracted from coordinates of (A) the homology model (χ^2 = 13.15) and NMR structures of (B) hTCERG1 FF_1 (PDB: 2DOD model 5, χ^2 = 1.97), (C) hTCERG1 FF_3 (PDB: 2DOE model 15, χ^2 = 2.12), and (D) hPrp40A FF6 (PDB. 2CQN model 5, χ^2 = 3.23). The structure of the hPrp40A FF_3

is colored green and other FF domains are colored red. Structures were generated using UCSF Chimera (44).

CaM interacts with hPrp40A FF_3 in an extended binding mode

We set out to determine the X-ray crystal structure of the CaM-FF_3 complex. However, despite the ability to isolate the complex by SEC, we were unable to find conditions that produced diffraction quality crystals. We therefore considered alternate approaches and chose to generate a structural model using an NMR and SAXS-based approach similar to that applied previously in our laboratory (24, 35). Although this approach does not directly provide complete high-resolution structures, it can accurately determine molecular shapes and topologies for globular domains and multi-domain complexes and generate accurate structural models.

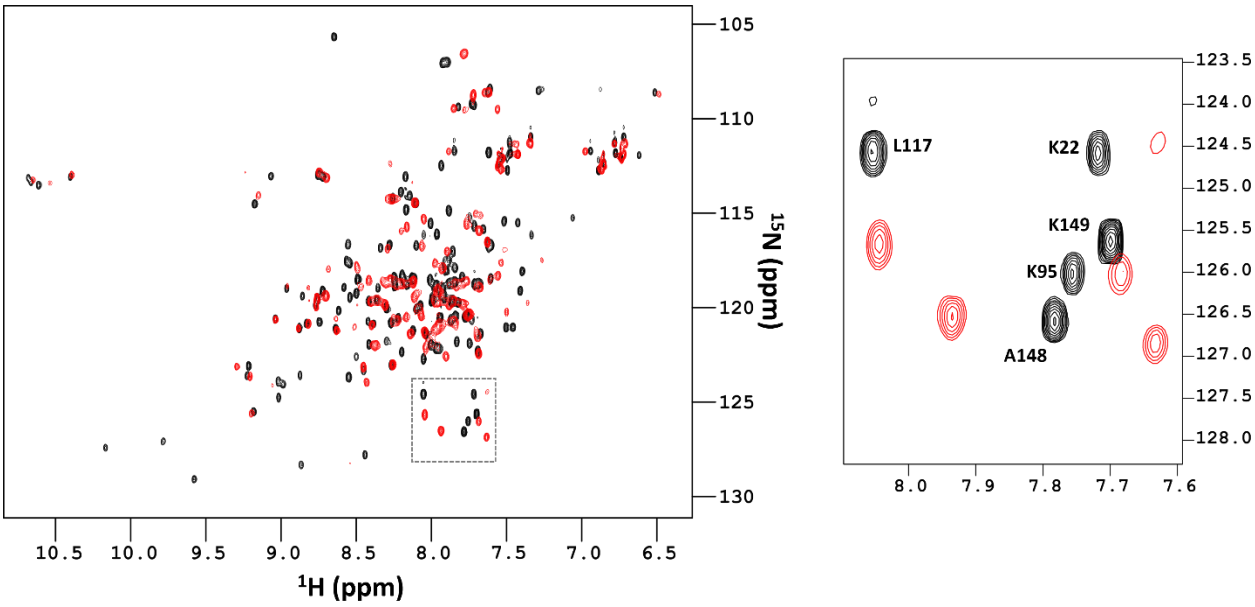

Figure 5. NMR analysis of the interaction between hPrp40A FF_3 and CaM. 800 MHz 2D ^{15}N-^{1}H HSQC NMR spectra at 37 °C of ^{15}N-enriched Ca^{2+}-loaded CaM obtained in the absence (black) and presence (red) of FF_3 (red). The sequence-specific assignment is provided for a select number

of readily identified crosspeaks from residues in both the N- and C-terminal domains of CaM to highlight both domains engage the FF_3 domain.

CaM is comprised of two EF-hand domains connected by a flexible linker, an organization that provides it with a wide range of possibilities to interact with target proteins. These different modes span the range from utilizing just one of its two domains to wrapping both domains around a short linear motif in the target to using its two domains to contact distant points in the target or even two different target molecules. Here we used NMR to determine if one or both domains of CaM are engaged in the complex with FF_3. Samples of ^{15}N-enriched CaM were prepared, and 2D ^{15}N-^{1}H HSQC NMR spectra were acquired in the absence and presence of the FF_3 (Fig. 5). Binding between ^{15}N-CaM and FF_3 resulted in several chemical shift perturbations in the N- *and* C-terminal domains of CaM, validating that both are involved in the interaction with the FF_3. We note that a select number of cross peaks in the spectrum were broadened beyond detection due to intermediate exchange between the free and bound state, which precluded complete structure determination by traditional NMR approaches.

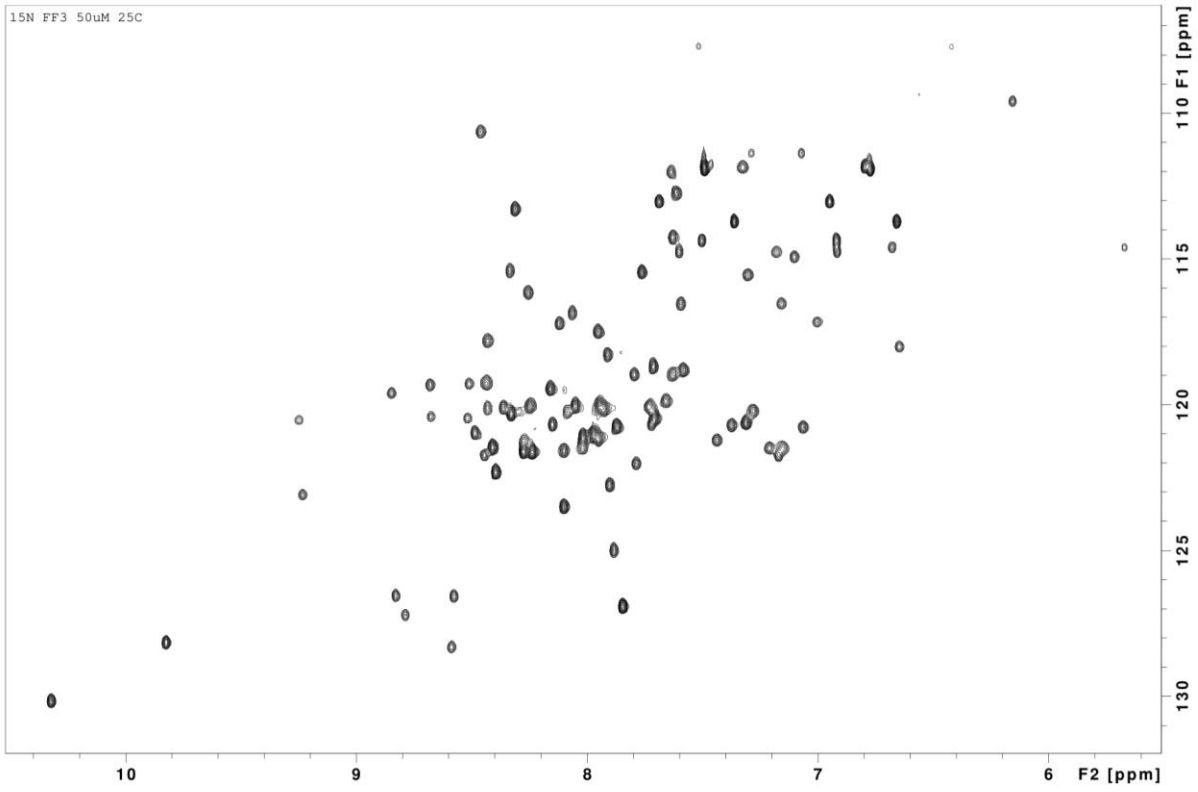

Fig. S4. 900 MHz 2D ^{15}N-^{1}H HSQC spectrum of ^{15}N-FF3. The spectrum shows chemical shift dispersion indicating a folded globular protein in the absence of CaM.

SAXS was used to determine the topology and shape of the CaM-FF_3 complex. The data were acquired for isolated FF_3 and CaM as well as for the complex (Fig. S3; Table S1). As described above for the isolated FF_3, the Guinier region of the $\log_{10}$ intensity plots and Porod volumes of isolated CaM and the CaM-FF_3 complex were consistent with the samples being largely free of aggregation (Fig. S3D). The high quality of the data is also reflected in the excellent agreement between the R_g values obtained from the Guinier and P(r) analyses (Table S1). The Kratky Plot, R_g, and D_{max} values correlate well with globular structures being present. The corresponding Porod volumes are consistent with the respective masses of hPrp40A FF_3 (9.3 kDa), CaM (16.7 kDa), and the CaM-FF_3 complex (26.0 kDa). The Porod volume for the complex was

slightly (within 20%) on the low side relative to the predicted volume to mass ratio for purely globular proteins, perhaps due to the equilibrium population of dissociated complex at room temperature. The Kratky plots and P_x values indicate a substantial degree of ordered globular structure (Fig. S3E; Table S1).

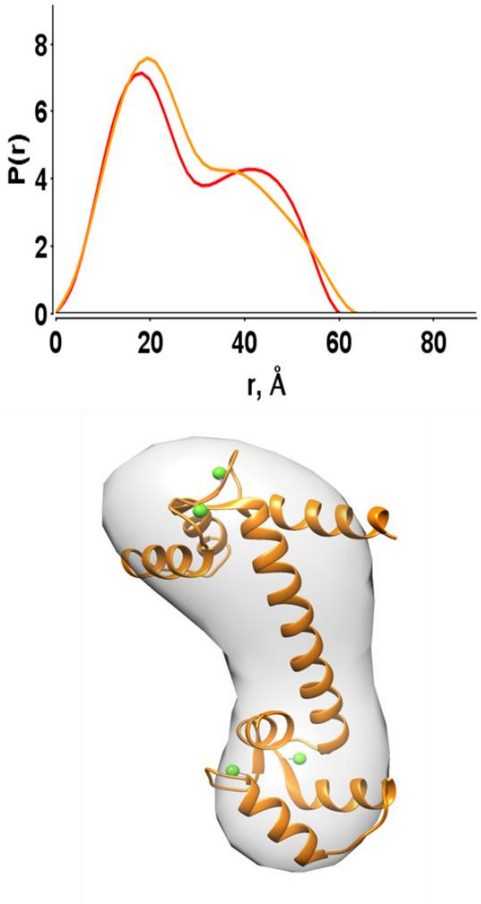

Fig. S5. Analysis of the architecture of CaM by SAXS. (A) Overlay of the experimental SAXS P(r) for Ca^{2+} saturated CaM (orange) with the back-calculated P(r) (red) extracted from the coordinates of the crystal structure of the Ca^{2+} bound CaM (PDB: 1CLL, $\chi^2 = 5.05$). The structure of CaM is superimposed with the *Ab initio* electron map generated from SAXS.

CaM has been studied extensively by SAXS and the values obtained in our study are consistent with past reports (36-38). The P(r) of free CaM contains a peak with a maximum at ~21 Å and a shoulder at ~40 Å, reflecting that on average it occupies an extended configuration, which is similar to the crystal structure of Ca^{2+}-loaded CaM (39) (Fig. S3F and Fig S5). Interestingly, the P(r) of the CaM-FF_3 complex showed the same two features (Fig. S3F), reflecting that it too occupies an extended configuration.

In order to further analyze the molecular architecture of the complex, we compared the experimental scattering of the CaM-FF_3 complex with the theoretical scattering back-calculated from structural coordinates of three CaM-target complexes with three different binding modes (Fig. 6). High values for the χ^2 fitting parameter and poor recapitulation of the shape of the curve were obtained for structures in which CaM engages its target in a "wrap-around" mode (χ^2 =147.88) or with only one CaM domain bound (χ^2 = 8.65) (Fig. 6A and 6B, respectively). The greatest similarity between the experimental and back-calculated P(r) was found for the structure of CaM bound to a Munc13-1 peptide (χ^2 =6.10) (40), in which CaM engages in an extended mode with each domain contacting separate motifs in the target that was in an extended, non-globular state (Fig. 6C). Together, the structural data indicate that CaM binds in an extended mode with both domains engaged with an extended, non-globular state of hPrp40A FF_3.

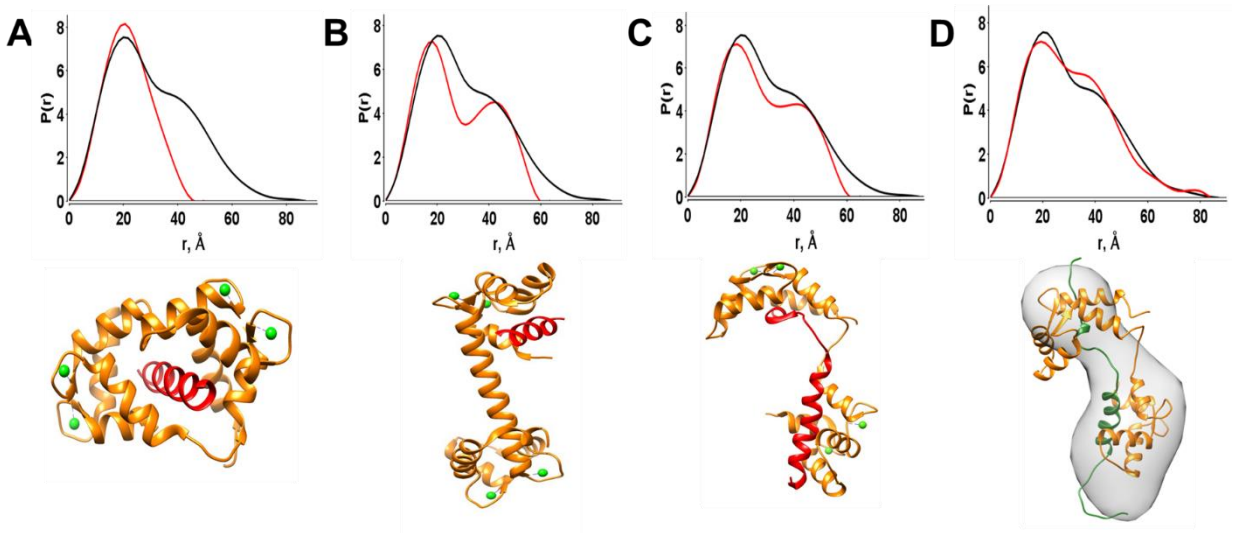

Figure 6. Comparative SAXS analysis of the CaM-hPrp40A FF_3 complex and CaM-target complexes in different binding modes. Structures along with I(q) vs q plots and P(r) functions derived from the SAXS data for the CaM-FF_3 complex (black) overlayed on the I(q) vs q and P(r) back-calculated (red) from coordinates of Ca^{2+}-bound CaM in complex with CaM-binding peptides from (A) smooth muscle light chain kinase (PDB: 2O5G, $\chi^2 = 147.88$), (B) human cardiac sodium channel ($Na_v1.5$) (PDB: 4DJC, $\chi^2 = 8.65$), (C) Munc13-1 (PDB: 2KDU model 5, $\chi^2 = 6.10$) peptides. (D) Overlay of the experimental I(q) vs q plot and P(r) with those derived from our homology model of the CaM-FF_3 complex ($\chi^2 = 7.04$). CaM is colored orange and the target peptides in red. Ca^{2+} ions are represented with green spheres. The ribbon diagrams were generated using UCSF Chimera (44).

CaM binding is anchored by two tryptophan residues in the hPrp40A FF_3 domain

In order to obtain further insights into the molecular details of the CaM-FF_3 complex, we generated a homology model of the CaM-FF_3 complex. First, the sequence of the FF_3 domain was aligned to the Munc13-1 peptide (Fig. 7A). An emphasis was placed on the key hydrophobic

residues that are known to serve as critical anchors that extend into the hydrophobic pockets of Ca^{2+}-loaded CaM. Close examination of the alignment of the sequences of hPrp40A FF_3 and the Munc13-1 peptide (Fig 7A) revealed that instead of the $W_1xxxF_5xxV_8$ motif found in Munc13-1, the first binding site predicted to mediate the interaction with CaM is comprised of the $W_1xxL_4xxI_7L_8$ motif, which overlaps with the Cen2-binding site (13). A second binding site is predicted by aligning a tryptophan in the FF_3 to the tryptophan of Munc13-1 that anchors binding to CaM N-terminal domain. Considering the possibility that both CaM-binding sites in FF_3 preserve a helical configuration upon their binding with CaM, we proceed to generate a helical wheel to visualize the spatial distribution of the $W_1xxL_4xxI_7L_8$ motif. In this projection, the key hydrophobic residues are located at the same side of the α-helix (Fig. 7B). To complete the homology model, the FF_3 sequence was threaded through the structure of the CaM-Munc13-1 complex. As anticipated, the homology model fits well to the SAXS data acquired for the complex (Figs. 6D, 8). In this structural model, the predicted hydrophobic residues in both binding sites are facing toward the hydrophobic pockets of the CaM domains (Fig. 8A), which include the two FF_3 Trp residues (Trp_{531}, Trp_{550}) primed to serve as the key hydrophobic anchors for binding.

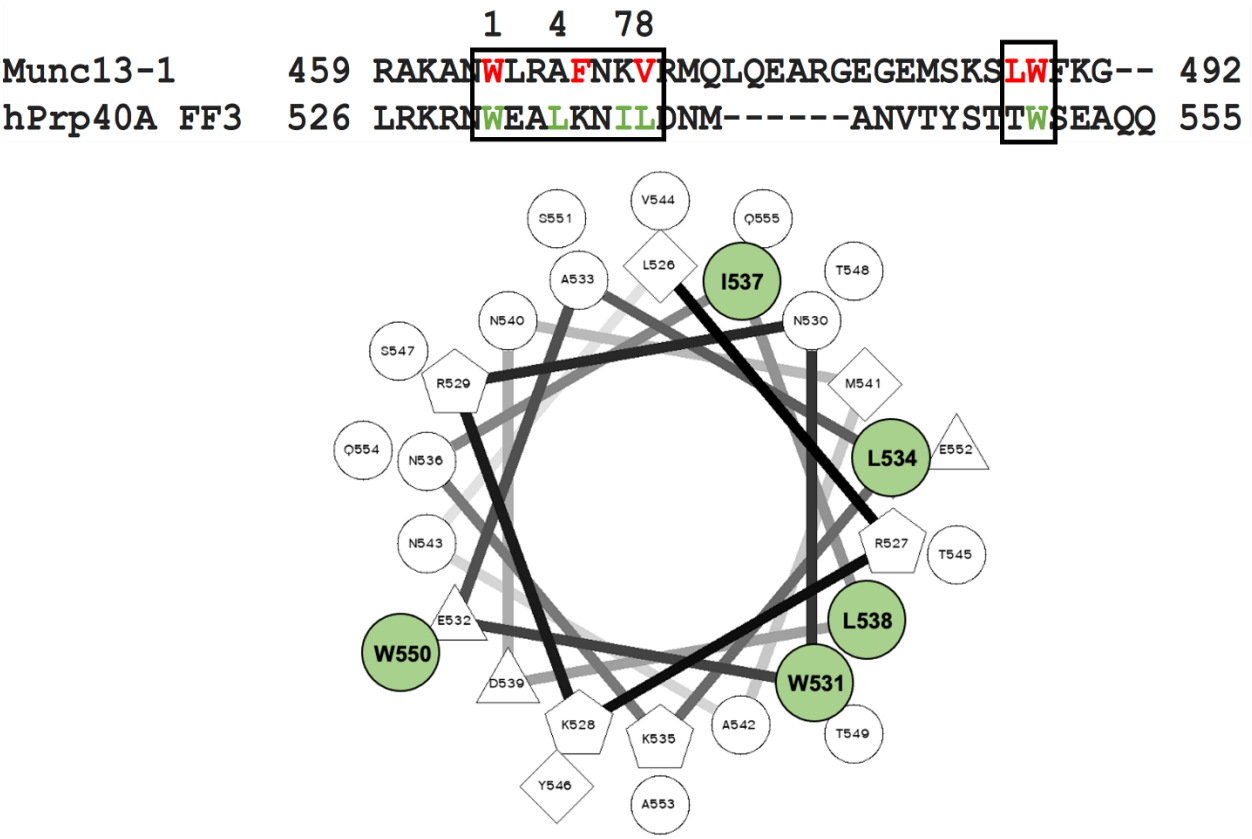

Figure 7. Identification of the two CaM-binding sites of the hPrp40A FF_3 domain. (A) Sequence alignment of the CaM-binding motifs of hPrp40A FF_3 with Munc13-1. Hydrophobic residues of Munc13-1 known to interact with CaM based on the 3D structure are highlighted in red. The two CaM-binding motifs of Munc13-1 are highlighted by boxes. The residues of the hPrp40A FF_3 predicted to interact with CaM are highlighted in green. The alignment of hPrp40A FF_3 residues 526-555 555 (Uniprot ID: O75400) with Munc13-1 residues 459-492 (Uniprot ID: Q62768) was generated using the online tool Clustal Omega. (B) Helical wheel of the FF_3 comprising the sequence from Leu_{526} to Gln_{555}. The predicted hydrophobic residues of the $W_1xxL_4xxI_7L_8$ motif are identified in green. The helical wheel representation was generated using the online resource at https://www.donarmstrong.com/cgi-bin/wheel.pl.

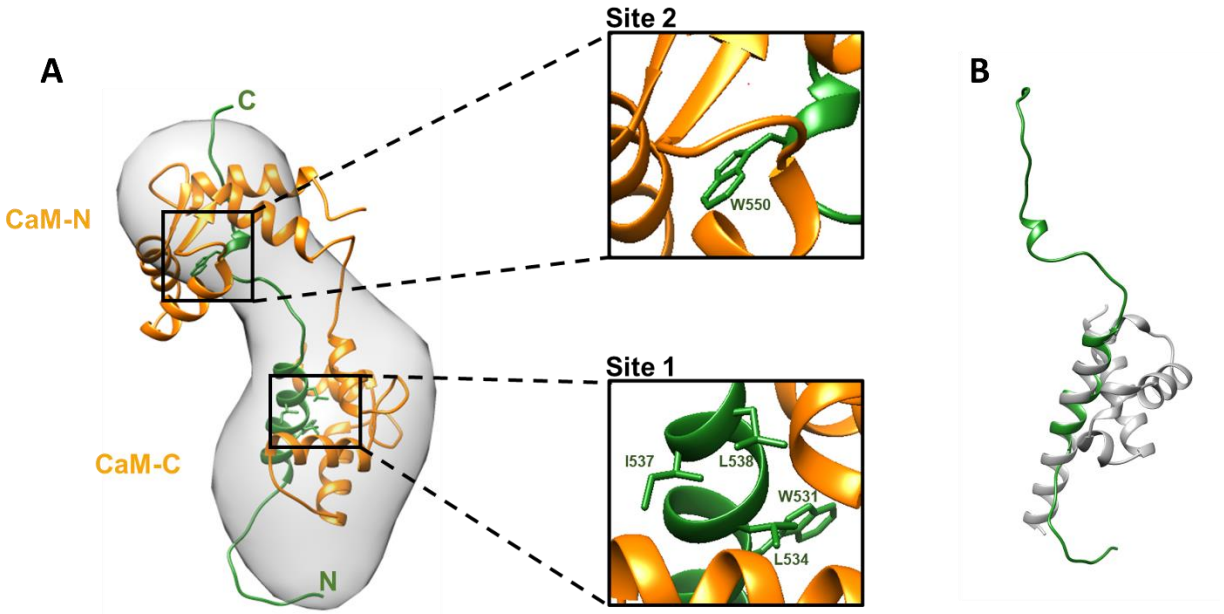

Figure 8. Structure of the CaM-FF_3 complex. (A) Overlay of the *ab initio* electron map of the CaM-FF_3 complex generated from SAXS and the 3D homology model. Black boxes present a close-up view of the two binding sites. CaM and FF_3 are colored orange and green, respectively. Ca^{2+} ions are represented with green spheres. (B) Overlay of FF_3 from CaM-FF3 complex model (green) and free FF_3 model (gray) aligned to helix α_1. The structures were generated using UCSF Chimera (44).

To test the proposal that the two Trp residues serve as anchors for binding to CaM, we performed an intrinsic Trp fluorescence experiment. This approach was feasible because the FF_3 contains only these two Trp residues and CaM has none. For this experiment, Trp fluorescence emission spectra were recorded in the absence and presence of increasing amounts of Ca^{2+}-loaded CaM (Fig. 9). Observation of maximum fluorescence emission at 354 nm for the isolated FF_3 indicates that the Trp side chains are exposed to the aqueous solvent. Additions of Ca^{2+}-loaded CaM result in smoothly increasing emission intensity and a blue shift, which reaches 17 nm in

saturating conditions. These changes in the fluorescence spectrum indicate both Trp residues are sequestered from solvent upon binding to CaM, consistent with both Trp residues binding into hydrophobic pockets. The combination of the NMR, SAXS and these intrinsic Trp fluorescence data strongly support a model in which the binding to CaM is coupled to unfolding of the FF_3 domain (Fig. 8B).

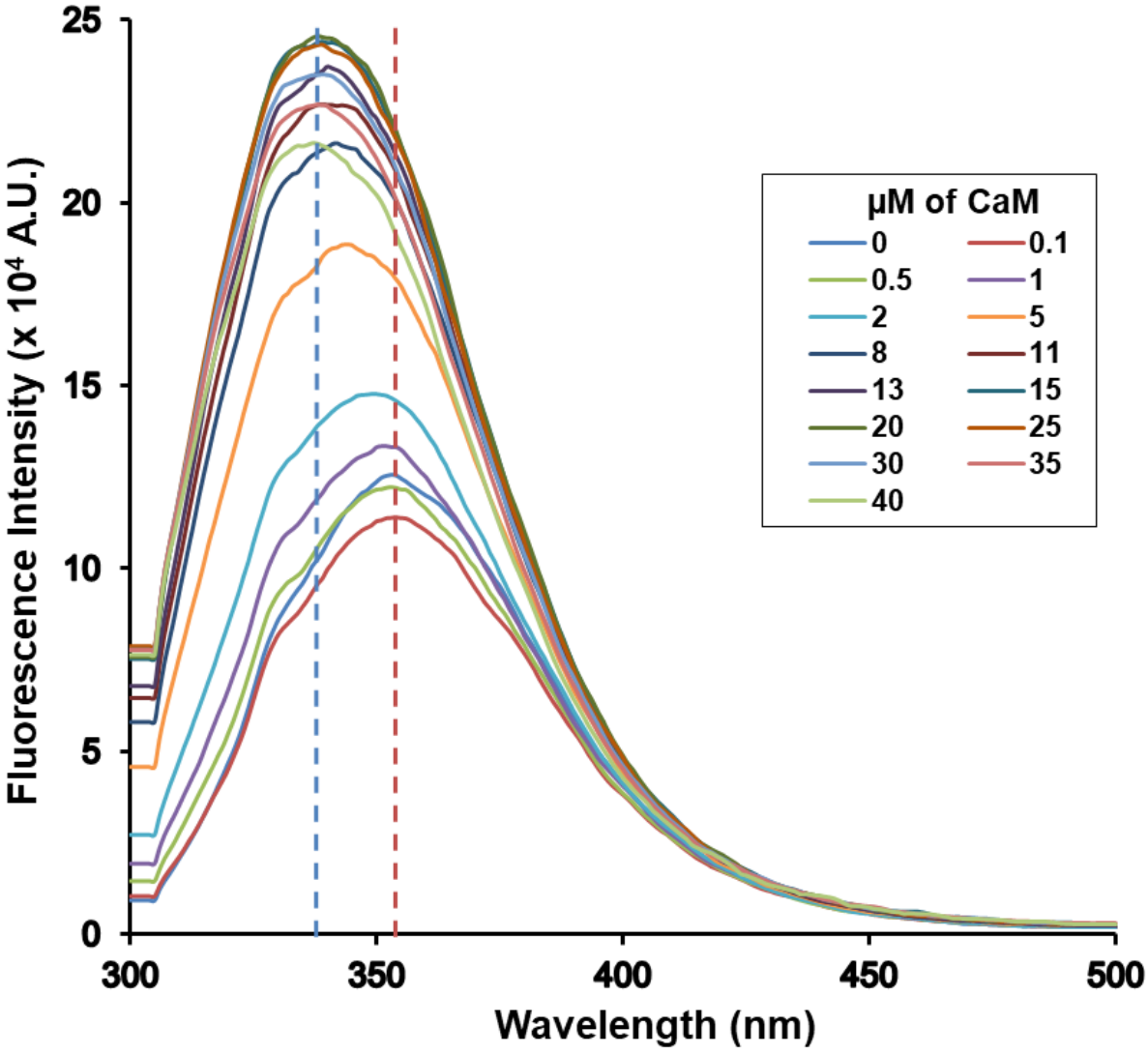

Figure 9. Titration of the hPrp40A FF_3 domain with CaM monitored by intrinsic fluorescence. Tryptophan fluorescence emission of FF_3 (5 μM) recorded at 37 °C in the absence and presence of

increasing amount of CaM. Note the shift in the maximum emission of the FF_3 from 354 nm (red lines) to 337 nm (blue lines) upon saturation with CaM.

Mutational analysis confirms the vital role of the dual tryptophan anchor for binding to CaM

To further test the hypothesis that the two tryptophan residues serve as the anchors for binding into the hydrophobic pockets of the two CaM domains, we generated three FF3 constructs with tryptophan to alanine mutations and assayed their effect on CaM binding with ITC and NMR: W531A, W550A, and the double mutant W531A,W550A. To directly monitor if the mutations result in differences in binding affinity, ITC experiments were performed under conditions optimized for the WT FF3 domain (Fig. 10). The thermograms show that the single-site mutants W531A and W550A bind more weakly to CaM than the WT, to the point that saturation is not yet reached at 4 equivalents of the mutated FF3 (Fig. 10B, C). Moreover, the binding of W531A,W550A is so weak that there is effectively no evidence of binding to CaM under these conditions (Fig. 10D).

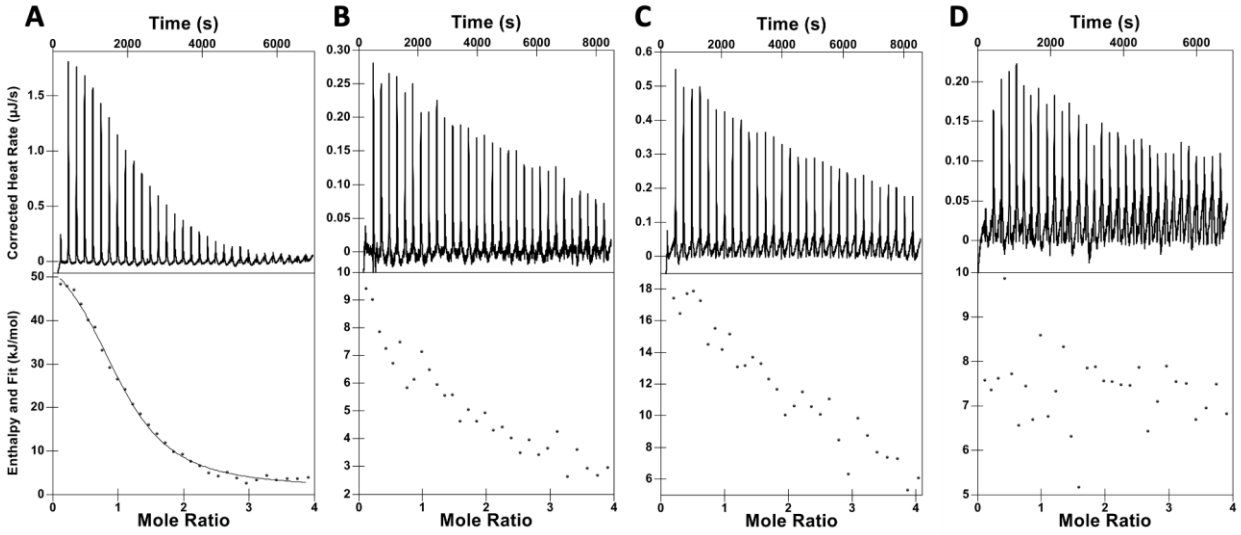

Figure 10. ITC analysis of the interaction of hPrp40A-FF3 variants and CaM. Representative thermograms are shown for the titration of (A) wt hPrp40A-FF3 with CaM, (B) hPrp40A-FF3 W531A with CaM, (C) hPrp40A-FF3 W550A with CaM, and (D) hPrp40A-FF3 W531A/W550A with CaM.

The effects of these FF_3 variants on binding to CaM was also investigated by 2D ^{15}N-^{1}H NMR. As noted above, titration of WT FF_3 into ^{15}N-enriched CaM results in general line broadening from the increase in mass and a substantial number of chemical shift perturbations in the spectrum. Figure 11 shows a comparison of a region of the CaM spectrum containing signals for 8 residues (E7, E8, F17, L33, E46, F66, E120 and E141) before and after addition of 2 molar equivalents of the WT FF3 domain or one of the mutants. In the case of the WT protein, 4 signals shift, 1 is broadened beyond detection and 4 remain unchanged, consistent with discrete, specific binding. After addition of the same 2 molar equivalents of either the W531A or W550A mutant the effects mimic those observed for the WT FF3. However, the extent of the perturbations of the spectrum is far more modest with very small chemical shift perturbations for 2 residues, and the rest remaining unchanged with no significant line broadening (Fig. 11B, C). Moreover, addition of the W531A/W550A double mutant has no effect on the spectrum of CaM indicating there is no binding (Fig. 11D). Even increasing the W531A or W550A to CaM ratio to 20:1 does not produce perturbations of the CaM spectrum comparable to the 2:1 ratio of WT FF3 to CaM (Fig. S6), fully consistent with the ITC data which show that the binding of the FF3 mutants is substantially weaker (Fig. 10). These NMR and ITC data confirm that Trp531 and Trp550 serve as the key hydrophobic anchors for the binding of the Prp40A FF3 domain to CaM.

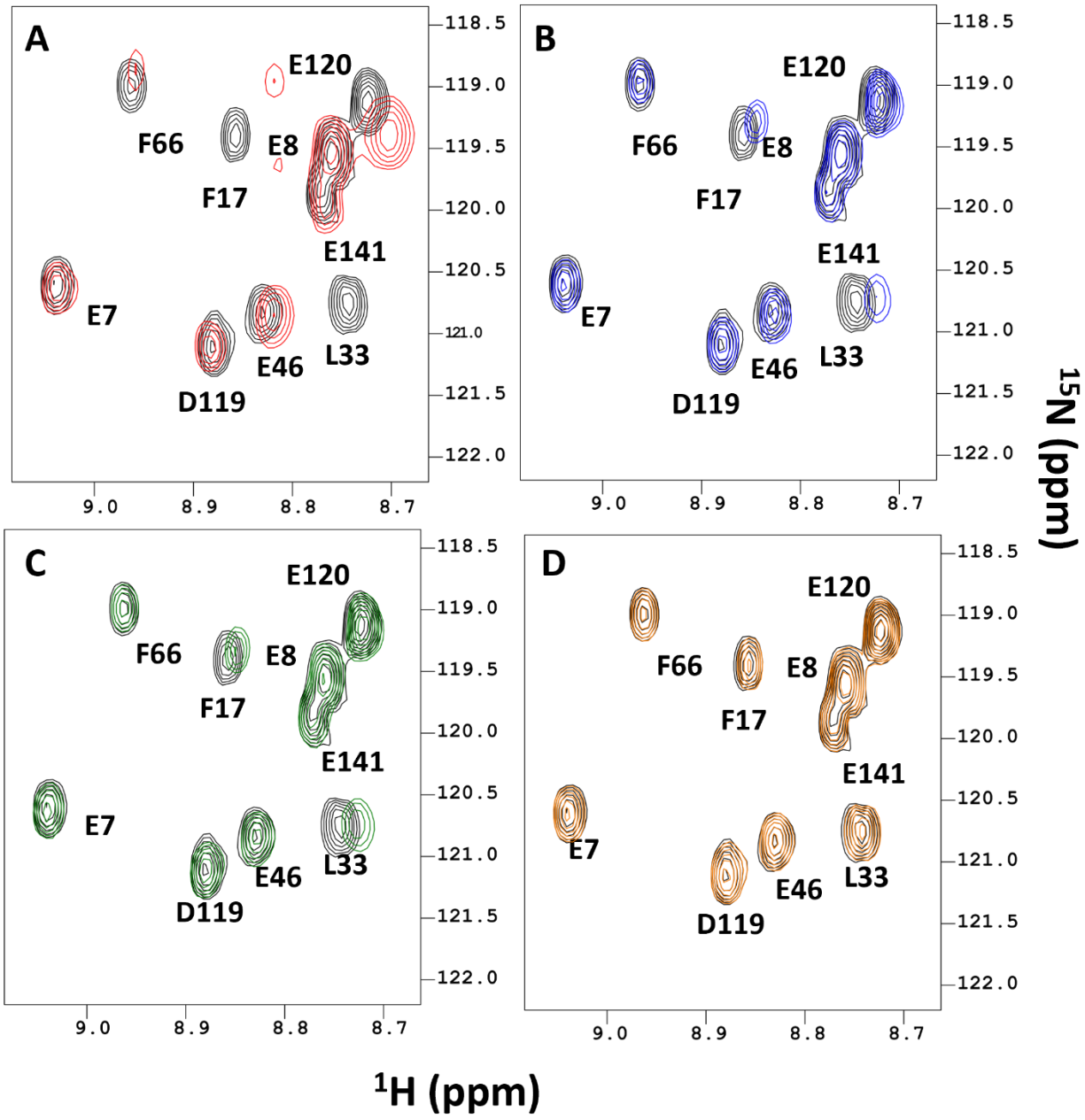

Figure 11. NMR analysis of the interaction between hPrp40A-FF3 variants and CaM. 900 MHz 2D ^{15}N-^{1}H HSQC NMR spectra at 37 °C of ^{15}N-enriched Ca^{2+}-loaded CaM alone (black) and with 2-fold excess of WT FF3 (red), FF3 W531A (blue), FF3 W550A (green), and FF3 W531A/W550A (orange). The sequence-specific assignment is provided for crosspeaks from residues in both the

N- and C-terminal domains of CaM to highlight changes in line broadening and peak shifts across different FF3 variants.

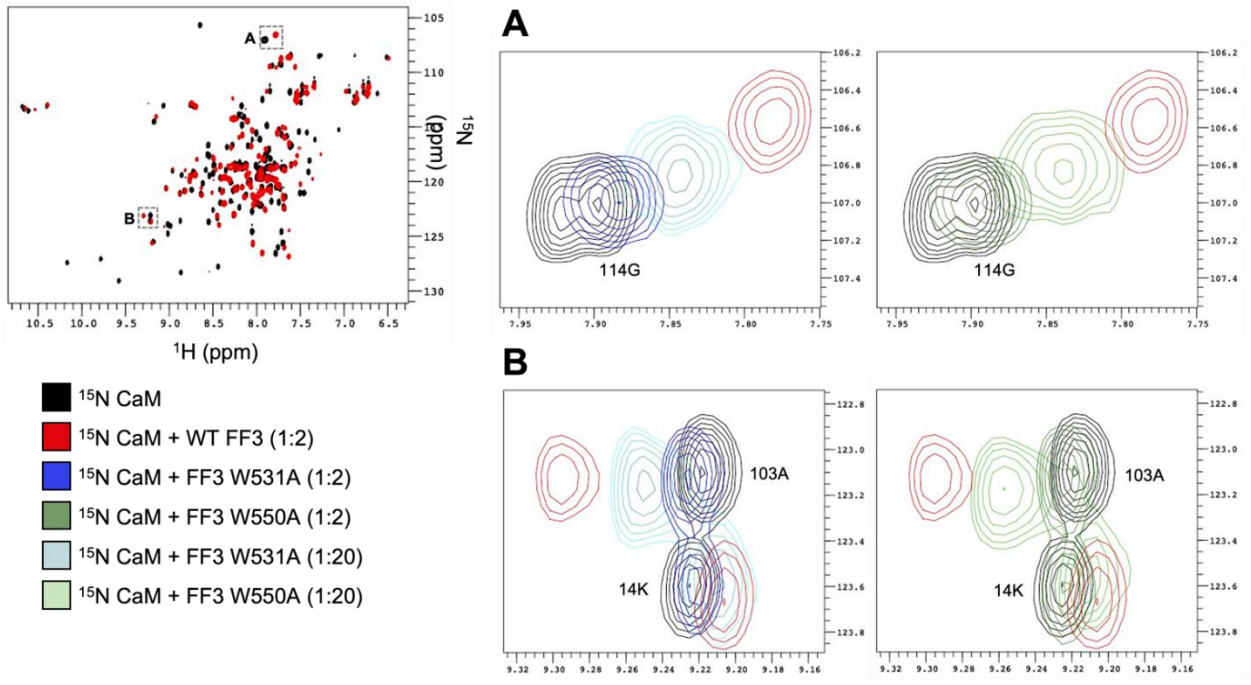

Fig. S6. Comparison of chemical shift perturbation upon the addition of 1:2 and 1:20 ratios of CaM:wtFF3 and FF3 mutants. ^{15}N-^{1}H HSQC NMR spectra at 37 °C of ^{15}N-enriched Ca^{2+}-loaded CaM alone (black) and with 2-fold excess of WT FF3 (red), FF3 W531A (blue), and FF3 W550A (green). Additional spectra were collected with 20-fold excess of FF3 W531A (cyan) and FF3 W550A (sea green). The sequence-specific assignment for two selected regions (A, B) is provided for crosspeaks from residues in both the N- and C-terminal domains of CaM to highlight changes in line broadening and peak shifts across different FF3 variants.

Discussion

Interaction between Prp40 and CaM was first reported in a *Saccharomyces cerevisiae* protein microarray (14). Our results show that like hCen2, CaM binds the hPrp40A FF_3 domain. SAXS analysis confirmed the validity of the homology model of the FF_3, which showed that the overall structure is similar to other FF domains. Comparative structural analysis of the hPrp40A FF_3 and TCERG1 FF_1 domains revealed differences only in the loop-3_{10}-loop region (Fig. 3). This can be attributed to the substitution in hPrp40A FF_3 of an aspartate residue for a conserved hydrophobic residue that ties this region into the hydrophobic core. Furthermore, this aspartate is surrounded by three glutamates residues and the strong electrostatic field associated with this cluster may contribute to the difference in this region of the structure relative to the "classical" 3_{10}-helix present in other FF domains.

NMR analysis revealed that both terminal domains of CaM are engaged in binding hPrp40A FF_3. Combined with the SAXS and intrinsic Trp fluorescence, our data strongly support a model in which the binding to CaM is coupled to unfolding of the FF_3 domain into an extended, non-globular conformation (Fig. 8B), similar to the structure of the complex of CaM with Munc13-1. Sequence alignment of the FF_3 with Munc13-1 allowed us to predict the two putative CaM-binding motifs containing key hydrophobic anchors. The FF_3 residues Trp_{531}, Leu_{534}, Ile_{537}, and Leu_{538} are the key residues in the $W_1xxL_4xxI_7L_8$ motif previously determined as the interaction site for hCen2. A second CaM-binding site is assigned to contain Trp_{550}. Based on this alignment and SAXS data showing an extended mode of binding similar to the CaM-Munc13-1 complex, we assumed hPrp40A FF_3 binds to CaM in the same antiparallel mode. The antiparallel orientation is common, having been reported for other CaM-target complexes that exhibit a bipartite binding mode with the target in an extended, non-globular conformation, including the HIV-1 MP (41) and Na^+/H^+ exchanger 1 (NHE1) (42).

As noted above, the $W_1xxL_4xxI_7L_8$ motif overlaps with a centrin-binding site within the FF_3 (13). Our previous work showed that hCen2 binds to the FF_3 in a Ca^{2+}-dependent manner solely through its C-terminal domain, which unlike the N-terminal domain has functional Ca^{2+}-binding sites (13, 43). Here we showed that CaM interacts with the FF_3 in a Ca^{2+}-dependent manner utilizing its two domains, both of which contain functional Ca^{2+}-binding sites. Thus, our studies show that these two EF-hand Ca^{2+} sensor proteins have fundamentally different binding modes to the FF_3 domain. Since both proteins bind to hPrp40A FF_3, it is reasonable to hypothesize that CaM and hCen2 modulate hPrp40A in different timeframes and their binding to hPrp40 likely has distinct biological roles.

Previous studies have implied that both hPrp40A and CaM are involved in Huntington Disease based on their interactions with WT and disease-associated mutants of Htt (4-6, 15-19). hPrp40A binds to both Htt and CaM utilizing its two WW domains, the FF_3 domain and possibly other FF domains. In addition to its interaction with Prp40A, CaM is known to bind directly to Htt. We have shown CaM uses both of its domains to engage the hPrp40A FF_3, so either multiple CaM molecules are recruited to this network, or a single CaM molecule is exchanging between different sites over the dynamic trajectory of this machinery. Regardless, the clear involvement of the two Ca^{2+}-sensors CaM and Cen2 implies regulation of this machinery by Ca^{2+} signaling. Clearly, further studies are required to clarify the action of CaM and Cen2 within this interaction network and establish if this machinery represents a viable target for HD.

Conclusion

The combination of the NMR, SAXS and intrinsic Trp fluorescence data show that CaM binds to an unfolded, non-globular state of hPrp40A FF_3 (Fig. 8B). Both CaM and the hPrp40A

FF_3 have an acidic pI, which suggest that their interaction is not driven by electrostatic but rather hydrophobic interactions. Unfolding of the FF_3 allows the exposure of hydrophobic residues, including two key tryptophan anchors, buried within its core in the folded state that mediate binding into the hydrophobic pockets of the two CaM domains. These observations lead us to propose that binding to CaM occurs via transiently unfolded states of the FF_3 domain. The mechanism, coupling unfolding to functional interaction, is similar to that observed for phosphorylation of p190A RhoGAP FF_1 at Tyr_{308} by the PDVF receptor α kinase (28). In that case, phosphorylation is dependent on unfolding of the FF_1 domain to expose the buried tyrosine located in the core. Interestingly, the T_m of the p190A RhoGAP FF_1 (52 °C) is similar to the T_m of hPrp40A FF_3 and other FF domains. Together, our results suggest that binding to a transiently populated unfolded state may be a more general mechanism than previously appreciated to regulate FF domain-mediated interactions.

Appendix References

1. Kao, H-Y. and P.G. Siliciano. 1996. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 16:960-967, doi: 10.1128/mcb.16.3.960.
2. Becerra, S., E. Andrés-León, S. Prieto-Sánchez, C. Hernández-Munain, and C. Suñé. 2016. Prp40 and early events in splice site definition. *WIREs RNA*. 7:17-32, doi: 10.1002/wrna.1312.
3. Li, X., S. Liu, L. Zhang, A. Issaian, R.C. Hill, S. Espinosa, S. Shi, Y. Cui, K. Kappel, R. Das, K.C. Hansen, Z.H. Zhou, and R. Zhao. 2019. A unified mechanism for intron and exon definition and back- splicing. *Nature*. 573: 375-380, doi: 10.1038/s41586-019-1523-6.
4. Faber, P.W., G.T. Barnes, J. Srinidhi, J. Chen, J.F. Gusella, and M.E. MacDonald. 1998. Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet*. 7: 1463-1474, doi: 10.1093/hmg/7.9.1463.
5. Passani, L.A., M.T. Bedford, P.W. Faber, K.M. McGinnis, A.H. Sharp, J.F. Gusella, J.P. Vonsattel, and M.E. MacDonald. 2000. Huntingtin's WW domain partners in Huntington's disease post-mortem brain fulfill genetic criteria for direct involvement in Huntington's disease. *Hum. Mol. Genet*. 9: 2175-2182, doi: 10.1093/hmg/9.14.2175.
6. Jiang, Y.J., M.X. Che, J.Q. Yuan, Y.Y. Xie, X.Z. Yan, and H.Y. Hu. 2011. Interaction with polyglutamine-expanded huntingtin alters cellular distribution and RNA processing of huntingtin yeast two-hybrid protein A (HYPA). *J. Biol. Chem*. 286: 25236-25245, doi: 10.1074/jbc.M110.216333.

7. Buschdorf, J.P. and W.H. Strätling. 2004. A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. *J. Mol. Med.* 82:135-143, doi: 10.1007/s00109-003-0497-9.

8. Chung, S., M.R. McLean, and B.C. Rymond. 1999. Yeast ortholog of the Drosophila crooked neck-protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA*. 5: 1042-1054, doi: 10.1017/s1355838299990635.

9. Vincent, K., Q. Wang, S. Jay, K. Hobbs, and B.C. Rymond. 2003. Genetic interactions with CLF1 identify additional pre-mRNA splicing. *Genetics*. 164: 895-907, doi: 10.1093/genetics/164.3.895.

10. Gasch, A., S. Wiesner, P. Martin-Malpartida, X. Ramirez-Espain, L. Ruiz, and M.J. Macias. 2006. The structure of Prp40 FF1 domain and its interaction with the crn-TRP1 motif of Clf1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem*. 281: 356-364, doi: 10.1074/jbc.M508047200.

11. Ester, C. and P. Uetz. 2008. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochemistry.* 9: 1-11, doi: 10.1186/1471-2091-9-29.

12. Li, X., S. Liu, J. Jiang, L. Zhang, S. Espinosa, R.C. Hill, K.C. Hansen, Z.H. Zhou, and R. Zhao. 2017. CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat. Commun*. 8: 1035, doi: 10.1038/s41467-017-01241-9.

13. Díaz Casas, A., W.J. Chazin, and B. Pastrana-Rios. 2017. Prp40 homolog A is a novel centrin target. *Biophys. J.* 112: 2529-2539, doi: 10.1016/j.bpj.2017.03.042.

14. Hesselberth, J.R., J.P. Miller, A. Golob, J.E. Stajich, G.A. Michaud, and S. Fields. 2006. Comparative analysis of Saccharomyces cerevisiae WW domains and their interacting proteins. *Genome Biol*. 7: R30, doi: 10.1186/gb-2006-7-4-r30.

15. Bao, J., A.H. Sharp, M.V. Wagster, M. Becher, G. Schilling, C.A. Ross, V.L. Dawson, and T.M. Dawson. 1996. Expansion of polyglutamine repeat in huntingtin leads to abnormal protein interaction with calmodulin. *Proc. Natl. Acad. Sci. USA*. 93: 5037-5042, doi: 10.1073/pnas.93.10.5037.

16. Zainelli, G.M., C.A. Ross, J.C. Troncoso, J.K. Fitzgerald, and N.A. Muma. 2004. Calmodulin regulates transglutaminase 2 cross-linking of huntingtin. *J. Neurosci*. 24: 1954-1961, doi: 10.1523/JNEUROSCI.4424-03.2004.

17. Durek, N.L., Y. Dai, and N.A. Muma, NA. 2008. Protective effects of interrupting the binding of calmodulin to mutant huntingtin. *J. Neuropathol. Exp. Neurol.* 67: 355-365, doi: 10.1097/NEN.0b013e31816a9e60.

18. Dai, Y., N.L. Durek, S.C. Fowler, and N.A. Muma. 2009. Striatal expression of calmodulin fragment improved motor function, weight loss, and neuropathology in the R6/2 mouse model of Huntington's disease. *J. Neurosci*. 29: 11550-11559, doi: 10.1523/JNEUROSCI.3307-09.2009.

19. Durek, N.L., Y. Dai, and N.A. Muma. 2010. Neuroprotective effects of calmodulin peptide 76-121aa: disruption of calmodulin binding to mutant huntingtin. *Brain Pathol*. 20: 176-189, doi: 10.1111/j.1750-3639.2008.00258.x.

20. Keryer, G., J.R. Pineda, G. Liot, J. Kim, P. Dietrich, C. Benstaali, K. Smith, F.P. Cordelières, N. Spassky, R.J. Ferrante, I. Dragatsis, and F. Saudou. 2011. Ciliogenesis is regulated by huntingtin-HAP1-PCM1 pathway and is altered in Huntington disease. *J. Clin. Invest.* 122: 4372-4382, doi: 10.1172/JCI57552.

21. Karam, A., L. Tebbe, C. Weber, N. Messaddeq, L. Morlé, P. Kessler, U. Wolfrum, and Y. Trottier. 2015. A novel function of Huntingtin in the cilium and retinal ciliopathy in Huntington's disease mice. *Neurobiol. Dis.* 80: 15-28, doi: 10.1016/j.nbd.2015.05.008.

22. Díaz Casas, A., G. Casanova Sepúlveda, O. Sánchez Negrón, A.P. Caro Muñiz, S.R. Malavé Ramos, A.P. Cebollero López, and B. Pastrana-Rios. 2018. Molecular biophysical characterization of the third FF domain of *Homo sapiens* Prp40 homolog A. *J. Mol. Struct.* 1167. 174-179, doi: 10.1016/j.molstruc.2018.04.059.

23. Šali, A. and T. L. 1993. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234: 779-815, doi: 10.1006/jmbi.1993.1626.

24. Topolska-Woś, A.M., N. Sugitani, J.J. Cordoba, K.V. Le Meur, R.A. Le Meur, H.S. Kim, J-E. Yeo, D. Rosenberg, M. Hammel, O.D. Schärer, and W.J. Chazin. 2020. A key interaction with RPA orients XPA in NER complexes. *Nucleic Acids Res.* 48: 2173-2188, doi: 10.1093/nar/gkz1231.

25. Schneidman-Duhovny, D., M. Hammel, J.A. Tainer, and A. Sali. 2013. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* 105: 962-974, doi: 10.1016/j.bpj.2013.07.020.

26. Schneidman-Duhovny, D., M. Hammel, J.A. Tainer, and A. Sali. 2016. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res*. 44: W424-W429, doi: 10.1093/nar/gkw389.

27. Grant, T.D. 2018. *Ab initio* electron density determination directly from solution scattering data. *Nat. Methods.* 15: 191-193, doi: 10.1038/nmeth.4581.

28. Bonet, R., L. Ruiz, E. Aragón, P. Martín-Malpartida, and M.J. Macias. 2009. NMR structural studies on human p190-A RhoGAPFF1 revealed that domain phosphorylation by the PDGF-Receptor α requires its previous unfolding. *J. Mol. Biol.* 389: 230-237, doi: 10.1016/j.jmb.2009.04.035.

29. Sánchez-Hernández, N., L. Ruiz, M. Sánchez-Álvarez, M. Montes, M.J. Macias, C. Hernández-Munain, and C. Suñé. 2012. The FF4 and FF5 domains of transcription elongation regulator 1

(TCERG1) target proteins to the periphery of speckles. *J. Biol. Chem.* 287: 17789-17800, doi: 10.1074/jbc.M111.304782.

30. Allen, M., A. Friedler, O. Schon, and M. Bycroft. 2002. The structure of a FF domain from human HYPA/FBP11. *J. Mol. Biol*. 323: 411-416, doi: 10.1016/S0022-2836(02)00968-3.
31. Bonet, R., X. Ramirez-Espain, and M.J. Macias. 2008. Solution structure of the yeast URN1 splicing factor FF domain: Comparative analysis of charge distributions in FF domain structures-FFs and SURPs, two domains with a similar fold. *Proteins*. 73: 1001-1009, doi: 10.1002/prot.22127.
32. Bonet, R., L. Ruiz, B. Morales, and M.J. Macias. 2009. Solution structure of the fourth FF domain of yeast Prp40 splicing factor. *Proteins*. 77: 1000-1003, doi: 10.1002/prot.22547.
33. Xu, D. and Y. Zhang. 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*. 80:1715-1735, doi: 10.1002/prot.24065.
34. Xu, D. and Y. Zhang. 2013. Toward optimal fragment generations for ab initio protein structure assembly. *Proteins*. 81:229-239, doi: 10.1002/prot.24179.
35. Johnson, C.N., F. Potet, M.K. Thompson, B.M. Kroncke, A.M. Glazer, M.W. Voehler, B.J. Knollmann, A.L. George Jr., and W.J. Chazin. 2018. A mechanism of calmodulin modulation of the human cardiac sodium channel. *Structure*. 26: 683-694, doi: 10.1016/j.str.2018.03.005.
36. Matsushima, N., N. Hayashi, Y. Jinbo, and Y. Izumi. 2000. Ca^{2+}-bound calmodulin forms a compact globular structure on binding four trifluoperazine molecules in solution. *Biochem. J.* 347. 211-215, doi: 10.1042/bj3470211.
37. Chow, J.Y.H., C.M. Jeffries, A.H. Kwan, J.M. Guss, and J. Trewhella. 2010. Calmodulin disrupts the structure of the HIV-1 MA protein. *J. Mol. Biol.* 400: 702-714, doi: 10.1016/j.jmb.2010.05.022.

38. Mazzorana, M. and T. L-M. Sørensen. 2019. Calcium-induced protein folding in calumenin and calmodulin. *Methods Mol. Biol.* 1929: 517-537, doi: 10.1007/978-1-4939-9030-6_32.

39. Chattopadhyaya, R., W.E. Meador, A.R. Means, and F.A. Quiocho. 1992. Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.* 228: 1177-1192, doi: 10.1016/0022-2836(92)90324-d.

40. Rodríguez-Castañeda, F., M. Maestre-Martínez, N. Coudevylle, K. Dimova, H. Junge, N. Lipstein, D. Lee, S. Becker, N. Brose, O. Jahn, T. Carlomagno, and C. Griesinger. 2010. Modular architecture of Munc 13/calmodulin complexes: dual regulation by Ca^{2+} and possible function in short-term synaptic plasticity. *EMBO J.* 29: 680-691, doi: 10.1038/emboj.2009.373.

41. Vlach, J., A.B. Samal, and J.S. Saad. 2014. Solution structure of calmodulin bound to the binding domain of the HIV-1 matrix protein. *J. Biol. Chem.* 289: 8697-8705, doi: 10.1074/jbc.M113.543694.

42. Köster, S., T. Pavkov-Keller, W. Kühlbrandt, and Ö. Yildiz. 2011. Structure of human Na^{+}/H^{+} exchanger 1 NHE1 regulatory region in complex with calmodulin and Ca^{2+}. *J. Biol. Chem.* 286: 40954-40961, doi: 10.1074/jbc.M111.286906.

43. Pastrana-Ríos, B., M. Reyes, J. De Orbeta, V. Meza, D. Narváez, A.M. Gómez, A. Rodríguez Nassif, R. Almodovar, A. Díaz Casas, J. Robles, A.M. Ortíz, L. Irizarry, M. Campbell, and M. Colón. 2013. Relative stability of human centrins and its relationship to calcium binding. *Biochemistry.*52: 1236-1248, doi: 10.1021/bi301417z.

44. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. 2004. UCSF Chimera -- a visualization system for exploratory research and analysis. *J. Comput. Chem*. 25: 1605-1612, doi: 10.1002/jcc.20084.