UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Studies on the substrate interactions of the bacterial primase

Permalink
https://escholarship.org/uc/item/3wr7f65j

Author
Rymer, Richard Untermayer

Publication Date
2012

Peer reviewed|Thesis/dissertation

Studies on the Substrate Interactions of the Bacterial Primase

by

Richard Untermayer Rymer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor James M. Berger, Chair
Professor Kathleen Collins
Professor Jennifer Doudna
Professor Amy Herr

Spring 2012

Abstract

Studies on the Substrate Interactions of the Bacterial Primase

by

Richard Untermayer Rymer

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor James M. Berger, Chair

Primases are DNA-dependent RNA polymerases found in all cellular organisms. In bacteria, primer synthesis is carried out by DnaG, an essential enzyme that serves as a key component of DNA replication initiation, progression, and restart. A long-standing question in the field of polymerase biochemistry is how the striking structural differences between DnaG and other single-subunit polymerases translate to differences in function. A specific barrier to addressing how DnaG relates to other polymerases is a relatively poor understanding of how DnaG associates with its own substrates. We investigated this issue through a multipronged approach. First, we have examined one of the earliest stages in primer synthesis by solving crystal structures of the *S. aureus* DnaG catalytic core bound to metal ion cofactors and either individual nucleoside triphosphates. Second, we investigated how the known classes of primase inhibitors, including two naturally-prevalent inhibitors, the nucleotidyl alarmones, pppGpp and ppGpp, and 2´-deoxyribonucleoside triphosphates, control primer synthesis. When taken together with both biochemical analyses and comparative studies of enzymes that use the same catalytic fold as DnaG, these studies pinpoint the predominant nucleotide-binding site of DnaG, leading to a reconciled model for primer synthesis by DnaG. Our studies of the inhibitors of DnaG illustrate how the induction of the stringent response in bacteria interferes with primer synthesis, and explain the ability of DnaG to differentially discriminate between dNTPs and rNTPs depending on the stage of primer synthesis. Finally, many new tools were developed, which combined with our improved understanding of how DnaG functions in substrate recognition and catalysis could motivate renewed interest in the research and development of DnaG-targeted antimicrobial agents.

1

Table of Contents

Chapter 3 - An introduction to inhibitors of DnaG and investigation of the mechanism of inhibition of DnaG by drug-like inhibitors

Chapter 4 – Elucidation of the mechanism of action of primer synthesis inhibition by the stringent response alarmones, (p)ppGpp.

Chapter 5 - Investigation of the mechanism initiation specific inhibition of primer synthesis by 2´-dNTPs

Appendex III – Synthesis and purification of catalytically competent hairpin-crosslinked DnaG RPD complexes —99

Acknowledgements

Chapter 1 – Introduction to the DNA primase of bacteria

DNA REPLICATION AND THE ROLE OF DNAG

DNA replication is a highly choreographed and tightly-regulated event in the lifecycle of all cells (Kornberg 1992). Carried out by a dynamic, multi-protein complex known as the replisome, the process of replication relies on the coordinated and coupled action of DNA unwinding with strand synthesis. Although many differences exist among bacterial, archaeal, and eukaryotic replisomes, they all utilize a specialized type of DNA-dependent RNA polymerase – termed a primase – to synthesize short RNA oligonucleotides (Frick and Richardson 2001). These RNA primers in turn serve as indispensible starting points for extension by DNA polymerases, which are incapable of initiating strand synthesis *de novo* (**Fig. 1.1a**) (Frick and Richardson 2001; Kuchta and Stengel 2010).

In bacteria, a protein known as DnaG is responsible for catalyzing primer synthesis during DNA replication (Rowen and Kornberg 1978; Kitani, Yoda et al. 1985). DnaG was discovered along with 6 other genes in a seminal study with *E. coli* strains having temperature sensitive DNA synthesis mutations (Carl 1970). It was suggested early on that DnaG played a role in the initiation of DNA synthesis (Schekman, Weiner et al. 1975), a prediction borne out by the isolation (Bouche, Zechel et al. 1975) and initial biochemical characterization (Rowen and Kornberg 1978) of the protein a few years after its discovery. During this same time, two other systems for studying DnaG-type primases were also beginning to emerge. The phage T7 helicase-primase gp4 was discovered simultaneously in 1977 by Kolodner and Richardson (Kolodner and Richardson 1977) and by Scherzinger and coworkers (Scherzinger, Lanka et al. 1977). This enzyme was shown to be a primase in 1978 (Kolodner, Masamune et al. 1978). The phage T4 primase, gp61, was shown to be important for priming in 1979 (Silver and Nossal 1979), and had been isolated and characterized by 1982 (Silver and Nossal 1982).

Primer synthesis by DnaG is now known to play a critical role in the regulation of numerous replicative processes, including replisome assembly (Makowska-Grzyska and Kaguni 2010), the control of fork progression (Salzberg, Salzberg et al. 1998; Lee, Hite et al. 2006; Tanner, Hamdan et al. 2008; Chintakayala, Machon et al. 2009), the regulation of Okazaki fragment length (Wu, Zechner et al. 1992; Wu, Zechner et al. 1992; Zechner, Wu et al. 1992), and replication fork restart (Heller and Marians 2006). Recent work has further shown that DnaG is a target for inhibiting DNA replication under conditions of nutrient-deprivation (Wang, Sanders et al. 2007; Maciag, Kochanowska et al. 2010), a process known as the stringent response. Thus, primer synthesis, and DnaG itself, is a central node for regulating the replication process in bacteria as a whole (**Fig. 1.1b**).

DNAG IS A STRUCTURALLY UNIQUE POLYMERASE

Despite the central of role of DnaG in DNA replication, many aspects of the structure-function relationships underlying DnaG substrate-interactions and catalysis remain poorly understood. DnaG is a member of a large class of enzymes known as nucleic-acid polymer-

ases. Polymerases can be classified according to sequence or structural similarity into 13 families: 12 families of single-subunit polymerases (Rothwell and Waksman 2005; Cameron, Moustafa et al. 2009; Steitz 2009; Heitz, Harter et al. 2010), and a single family of multisubunit RNA polymerases (Cramer 2002; Cramer, Armache et al. 2008). Although important for a diverse array of functions in cells, viruses, and self-replicating plasmids, the single-subunit nucleic acid polymerases are surprisingly closely related to each other, a feature that reflects the common substrates of the biological processes they are involved in. Eleven of the single-subunit families share a mechanism of action termed the two-metal mechanism, which is also utilized by the multisubunit RNA polymerases (Yang, Lee et al. 2006). The majority of the single-subunit polymerases (11 families) also share a global structural arrangement of polymerase domains with an anatomical resemblance to a right-hand (**Fig. 1.2a**), with the fingers, palm, and thumb representing distinct domains. The structures and sequence similarities of each domain vary significantly among the families, but to date all of these "canonical polymerases" display this domain structure (Rothwell and Waksman 2005). The twelfth family of single-subunit polymerases, the DnaG-type primases, utilizes the same substrates, and they perform the same reaction as the other single-subunit polymerases, but they are otherwise unrelated these canonical polymerases.

OVERVIEW OF THE STRUCTURES AND FUNCTIONS OF THE DOMAINS OF DNAG

The differences between DnaG-type primases and canonical polymerases begin with the global domain structure of DnaG, which is distinct from canonical polymerases. Instead of a classic fingers-palm-thumb domain organization, DnaG is comprised of three other types of domains (**Fig. 1.2b**, **Fig. 1.3a**): 1) an N-terminal **Z**inc **B**inding **D**omain (**ZBD**), 2) a central **R**NA **P**olymerase **D**omain (**RPD**), and 3) a C-terminal **H**elicase **I**nteraction **D**omain (**HID**) (Wu, Zechner et al. 1992; Sun, Tormo et al. 1994; Mustaev and Godson 1995). Most members of the family share this domain structure, although a subset typified by T7 gp4 have lost the HID and become covalently attached to the helicase (Matson, Tabor et al. 1983).

While the three domains of DnaG would seem to relate it functionally to other polymerases, these elements are structurally unrelated, and are probably only loosely related functionally to the fingers, palm and thumb paradigm. In particular, the fingers, palm and thumb of other polymerases collaborate in substrate recognition and catalysis (Brautigam and Steitz 1998), whereas the domains of DnaG are more loosely-tethered to each other, and appear to have distinct functions (Larson, Griep et al. 2010).

Substrate specificity and autoregulation are conferred upon DnaG through its N-terminal ZBD. This domain is a single, globular fold that resembles the zinc-ribbon motifs found more commonly in transcription factors (**Fig. 1.3b**) (Pan and Wigley 2000). Although two crystal structures of the ZBD have been solved (**Fig. 1.3c**), exactly how the ZBD carries out its diverse functions is not understood at the structural level. These aspects of the ZBD are discussed later in more detail.

The central RPD contains the DnaG active site. This region shows no similarities with the catalytic palm domain of any other family of polymerase; rather, it is more closely related

to topoisomerases, OLD family nucleases, and RecR via the TOPRIM (**TO**poisomerase **PRI-M**ase) fold (Aravind, Leipe et al. 1998; Keck, Roche et al. 2000; Podobnik, McInerney et al. 2000; Kato, Ito et al. 2003). Several RPD structures have been solved that together illustrate the high degree of structural conservation within the bacterial DnaGs (see **Chapter 2** for more detail). The RPDs of DnaG and T7 gp4 are related mostly through their respective TOPRIM folds (**Fig. 1.3d/e**). The N-terminal subdomain of T7 gp4 is smaller and structurally distinct from the corresponding domain of DnaG, and T7 gp4 lacks the C-terminal helical bundle found in its bacterial orthologs. The functional implications of these structural differences, if any, are not well understood, although the biochemical characteristics and properties of T7 are generally applicable to DnaG.

The C-terminal helicase interaction domain directs the association between primase and the replicative helicase (discussed below) by binding to the N-terminus of DnaB (Chintakayala, Larson et al. 2007). Two crystal structures from *E. coli* and *G. stearothermophilus* show that the HID is a small helical domain structurally related to the DnaB N-terminal domain, and that the two elements likely arose from a gene duplication event (Oakley, Loscha et al. 2005; Syson, Thirlway et al. 2005). The HID is relatively poorly conserved among bacterial species (Syson, Thirlway et al. 2005), which may reflect the high degree of species specificity of the primase-helicase interaction (Koepsell, Larson et al. 2008; Larson, Griep et al. 2010).

DETERMINANTS OF SUBSTRATE BINDING AND CATALYSIS BY DNAG

Both biochemical and structural efforts have pursued an understanding of the structure-function relationships underlying substrate recognition and catalysis by DnaG. Overall, the catalytic and substrate recognition roles of the conserved residues of the primase catalytic core are only loosely understood. However, some structural features of DnaG appear to be analogous to those of canonical polymerases.

Canonical polymerases utilize two or three acidic, metal-binding residues to catalyze strand synthesis (Rothwell and Waksman 2005). DnaG-type primases share this feature, but only in part. Several studies have investigated catalytically-important acidic amino acids in the DnaG active site, and found that these enzymes require a complement of either five (T7) (Lee and Richardson 2005), or six (*E. coli*) (Godson, Schoenich et al. 2000; Rodina and Godson 2006), invariant glutamate and aspartate residues to carry out synthesis (**Fig. 1.4a/b**). Thus far, however, only a subset of residues in both proteins has been shown to bind divalent metal ions (Godson, Schoenich et al. 2000; Kato, Ito et al. 2003). Moreover, the functional topology of the DnaG active site differs drastically from the classic catalytic triad observed in the canonical polymerases (**Fig. 1.4c**).

A common feature of RNA and DNA binding proteins is the use of basic amino acids to interact with their nucleic-acid substrates. DnaG and T7 gp4 both possess conserved, basic amino acids in and around the active site, some of which are indeed involved in binding DNA (**Fig. 1.4d**). For *E. coli* DnaG, several of these residues form a shallow basin that was recently shown to be involved in binding the 5´-tail of a ssDNA template (See **Chapter 2** for

more detail) (Corn, Pelton et al. 2008). Although other conserved residues in both T7 gp4 and *E. coli* DnaG are clearly important for catalysis, their exact role in DNA binding has been less well established (Lee and Richardson 2001; Rodina and Godson 2006).

Structural efforts to understand the mechanisms of substrate binding and catalysis have met with similarly limited success. Two substrate bound structures are available: the Mg^{2+} bound T7 ZBD-RPD structure (Kato, Ito et al. 2003), and the structure of *E. coli* RPD cross-linked to a ssDNA template (Corn, Pelton et al. 2008). The structure of a full ternary complex of DnaG is not available at this time, although the aforementioned basin was identified in part through capturing the ssDNA bound structure of the *E. coli* DnaG RPD (Corn). In summary, it is unclear how the distinctions between the structures of the DnaG-type primases and the canonical polymerases translate to differences in function.

DETERMINANTS OF SUBSTRATE SPECIFICITY BY DNAG

In addition to their structural differences, DnaG-type primases are also enzymologically different from canonical polymerases, possessing at least two unique characteristics in the polymerase pantheon: start-site sequence specificity and autoregulation of product length. In the absence of accessory factors, E*. coli* DnaG will specifically synthesize 8-12 nucleotide (nt) primers (Kitani, Yoda et al. 1985), using the middle base of CTG or CAG triplet sequences in the template as a start site (Swart and Griep 1993; Khopde, Biswas et al. 2002). T7 gp4 also begins synthesis from opposite the central base in a template-encoded GTC triplet (Tabor and Richardson 1981), but stops once the product reaches 4 nt (Scherzinger, Lanka et al. 1977; Romano and Richardson 1979); T4 gp61 begins synthesis opposite GTT or GCT (Valentine, Ishmael et al. 2001), and synthesizes 4-5 nt primers (Liu and Alberts 1980; Nossal 1980). More recently, several surveys of the biochemical properties of DnaGs from different organisms found this framework of sequence specificity and length regulation to be a highly conserved feature of the enzyme class (Koepsell, Larson et al. 2006; Larson, Bressani et al. 2008; Larson, Griep et al. 2010). In all cases, the 3´ base in the recognition sequence (e. g., "G" of CTG with *E. coli* DnaG) is omitted from the primer product.

The source of sequence specificity and length regulation was first determined using a variant of T7 gp4 lacking the ZBD (Bernstein and Richardson 1988). This truncated form retained some synthesis capabilities (albeit at a reduced level), but no longer discriminated against primer sequence and length. Subsequent studies using a variety of methods have added substantial support to this initial finding (Mendelman and Richardson 1991; Hine and Richardson 1994; Mendelman, Beauchamp et al. 1994; Mendelman, Kuimelis et al. 1995; Kusakabe, Hine et al. 1999; Akabayov, Lee et al. 2009; Lee, Zhu et al. 2010). An identical role for the *E. coli* ZBD was later established in a series of studies (Sun, Tormo et al. 1994; Mustaev and Godson 1995; Griep and Lokey 1996; Griep, Adkins et al. 1997; Powers and Griep 1999). The mechanisms of sequence-specificity and length regulation are an area of active study. It is clear that the ZBD interacts directly with the template and the nascent primer through sequence-specific interactions (Mendelman, Kuimelis et al. 1995; Kusakabe and Richardson 1996; Frick and Richardson 1999; Kusakabe, Hine et al. 1999; Kato, Frick et al. 2001; Valentine, Ishmael et al. 2001; Corn, Pease et al. 2005; Akabayov, Lee

et al. 2009). To date, however, a crystal structure of the ZBD bound to a substrate has not been solved. One point of contention is whether the ZBD interacts with the RPD in *cis* or in *trans* (Lee and Richardson 2002; Corn, Pease et al. 2005; Qimron, Lee et al. 2006; Akabayov, Lee et al. 2009; Lee, Zhu et al. 2010; Zhu, Lee et al. 2010), a distinction that is only relevant in the context of the replicative helicase with which the DnaG-type primases interact.

INTERACTIONS WITH THE REPLICATIVE HELICASE

Most polymerases require a single-stranded template, and therefore rely on the action of a helicase to separate a duplex DNA into individual strands. DnaG-type primases are no different in this respect, a property which likely led nature to evolve a direct interaction with the replicative helicase (Marians 1992; W-, Zechner et al. 1992; Tougu, Peng et al. 1994; Lu, Ratnakar et al. 1996). It has been suggested that the helicase interaction is a mechanism for limiting priming to the replication fork, rather than any available ssDNA with the appropriate sequence (Corn and Berger 2006). This connection may also couple primer synthesis to replication progress in general (Zhu, Lee et al. 2010). The association of a DnaG-type primase and a RecA-family replicative hexameric helicase is found in three forms (**Fig. 1.5a-c**), although all occur as an array of at least three primase protomers to each helicase hexamer (Bird, Pan et al. 2000; Thirlway, Turner et al. 2004; Thirlway and Soultanas 2006; Bailey, Eliason et al. 2007). T7 gp4 exemplifies an in-frame fusion of a DnaG-type primase and a helicase. *G. stearothermophilus* DnaG and T4 gp61 form highly stable, non-covalent complexes with their replicative helicases (DnaB and gp41 respectively). DnaGs from *E. coli*, and other mesophilic bacteria form only transient interactions with DnaB that are observable indirectly through activity (Zechner, Wu et al. 1992; Johnson, Bhattacharyya et al. 2000; Koepsell, Larson et al. 2006), or through sensitive trapping experiments and biophysical methods that probe transient states (Mitkova, Khopde et al. 2003).

One reason for the reduced stability of mesophilic DnaG-DnaB arrays may be that repeated cycles of primase-helicase association and dissociation are an important feature in systems with separate proteins (Lee, Hite et al. 2006; Nelson, Kumar et al. 2008; Spiering, Nelson et al. 2008; Manosas, Spiering et al. 2009). In the case of T7 gp4, primer-handoff occurs through a direct interaction of a template bound ZBD and a waiting polymerase (Kato, Frick et al. 2001; Kato, Ito et al. 2004). In bacteria and phage T4, intermediate complexes form with other components of the replisome, which could necessitate dissociation of the primase from the helicase (Yuzhakov, Kelman et al. 1999; Nelson, Kumar et al. 2008). In bacteria, for example, it was found originally that a "three-point switch" between primase, single-stranded DNA binding protein (SSB), and the χ-subunit of the clamp loader complex shepherded the primer to the replicative polymerase (Yuzhakov, Kelman et al. 1999). Recent evidence, though, does not support the requirement for the formation of a stable χ-SSB-primer subcomplex in Okazaki fragment synthesis (Marceau, Bahng et al. 2011), which is a defining feature of the three-point switch. Instead, the interaction between χ and SSB is more important in coupling leading and lagging strand synthesis (Marceau, Bahng et al. 2011).

Beyond restricting priming to appropriate times and locations, along with coupling primer

synthesis to primer hand-off, the association of a DnaG-type primase with a replicative helicases both greatly stimulates the rate of primer synthesis (Johnson, Bhattacharyya et al. 2000; Koepsell, Larson et al. 2006; Koepsell, Larson et al. 2008) and loosens sequence specificity (Bhattacharyya and Griep 2000). The marked effect on primase upon binding to DnaB occurs primarily through an enhancement of its apparent affinity for template (Johnson, Bhattacharyya et al. 2000), which likely arises at least in part from an increase in the local concentration of template bound to DnaB (Mitkova, Khopde et al. 2003).

INTERACTION OF THE RPD WITH THE ZBD

The primase-helicase array may also promote the association of the ZBD with the RPD *in trans* (**Fig. 1.6a**). This interaction is known to lead to additional stimulation of the primer synthesis rate, and further insures that primer synthesis is initiated at the appropriate sequence (Corn, Pease et al. 2005). A *trans* interaction between two neighboring DnaG protomers has also been demonstrated to be functionally important for a bacterial primase-helicase array (Larson, Griep et al. 2010) using chimeric bacterial DnaG enzymes. In this latter study, the ZBD from one species has been swapped onto the RPD of an alternative species and vice versa. The chimeras only achieve maximum activity under helicase-stimulated conditions using mixtures of chimeric enzymes.

This model could extend to other members of the DnaG family, including T7 gp4. In this case, subunits of T7 gp4 having a mutant RPD deficient in primer synthesis can complement subunits lacking a ZBD (Lee and Richardson 2002). Curiously, subsequent studies have indicated that T7 gp4 can switch between *cis* and *trans* activated synthesis (Qimron, Lee et al. 2006), although what causes the switch and what purpose it serves is not understood. The crystal structure of the T7 gp4 primase fragment (Kato, Ito et al. 2003) suggests that the ZBD can access an extended conformation more readily, which is consistent with the ability of the enzyme to switch between a *cis* and *trans* synthesis (**Fig. 1.6b/c**).

Bacterial primases, on the other hand, are unlikely to engage in a *cis* interaction. In both *E. coli* and *A. aeolicus* DnaG, the ZBD and RPD primarily adopt a compact conformation relative to one another. Although an extended conformation is accessible in high salt (Corn, Pease et al. 2005), it is unclear from the *Aa*DnaG ZBD-RPD crystal structure how the ZBD could associate with the active site of its own RPD (**Fig. 1.6d**). DNA binding data also support the association of two DnaG protomers on a single template strand (Khopde, Biswas et al. 2002). Finally, a distal interaction between the ZBD and the template or nascent product is unlikely, as recent studies demonstrate that an interaction between the ZBD and the initiating nucleoside triphosphate bound in the primase active site of T7 gp4 is critical for activity (Lee, Zhu et al. 2010).

The possibility of a *trans* interaction between the ZBD and the RPD represents one of the most fundamental differences between DnaG family polymerases and canonical polymerases. Although it is unknown exactly how the ZBD binds to template and other substrates, the unique manner in which it collaborates with the RPD to carry out primer synthesis suggests its roles in substrate recognition and catalysis are unlike those of the potentially function-

ally analogous fingers and thumb domains of other single-subunit polymerases.

PRIMER SYNTHESIS KINETICS BY DNAG

One last difference between canonical polymerases and DnaG type primases is their poor enzymatic activities. *E. coli* synthesizes primers with a k_{cat} (0.00089 s^{-1}) that is more than a hundred times lower than *E. coli* RNA polymerases (Swart and Griep 1995). Addition of the helicase improves activity ~13-fold (to 0.012 s^{-1}) (Johnson, Bhattacharyya et al. 2000), by rate of product formation is still almost 10-fold slower than a "slow" enzyme like *E. coli* RNA polymerase. This feature of DnaG-type primases may be another adaptation to limit priming to only specific times and locations.

While the chemical mechanism of primer synthesis by DnaG has not been determined, these enzyme likely employ a catalytic cycle similar to RNA polymerases (**Fig. 1.7**) (Swart and Griep 1995; Frick, Kumar et al. 1999; Frick and Richardson 1999; Steitz 2009). In the first and rate-limiting synthesis stage, two NTPs are condensed into a dinucleotide in a process called initiation. Following this, the primer is elongated during the extension phase of synthesis by serial additions of NTPs.

SUMMARY AND EXPERIMENTAL APPROACH

As outlined above, the DnaG family of polymerases differs remarkably from the other well-understood members of the nucleic acid polymerase cohort. This disparity raises the possibility that DnaG substrate recognition and catalysis mechanisms also differ from the other polymerases in fundamental ways. Defining the mechanisms of action and substrate engagement by DnaG could lead to new paradigms of nucleic acid polymerase biochemistry. Moreover, a molecularly detailed understanding of DnaG mechanisms would help to evaluate its prospects as a target for antibacterial development. Currently, the argument for primases as an antibacterial target is purely hypothetical: they are completely unrelated to the human primase (Frick and Richardson 2001; Kuchta and Stengel 2010), and so their use would likely avoid target-based lethality to the patient. However, DnaG-type primases are sufficiently related to each other that inhibitors might serve as broad-spectrum antibiotics(Robinson, Causer et al. 2012). Determining whether DnaG is a reasonable target for small-molecule intervention requires a deeper understanding primase structure and function. A specific example of a polymerase-targeted antibiotic is rifampycin, which binds to and inhibits bacterial multisubunit RNA polymerases (mRNAPs) (Kohanski, Dwyer et al. 2010). This drug would not be as specific for these polymerases were it not for the vast structural and functional differences between the bacterial mRNAPs and the multitude of other polymerases in a human cell. One wonders if DnaG is another example of a divergent polymerase that is ripe for exploitation as a drug-target, or if it is functionally too highly-related to other single-subunit polymerases to avoid the severe side-effects that would be the inevitable result of inhibiting them. In summary, there is a significant need for a better understanding of DnaG-substrate interactions.

My approach to this problem has involved two strategies. First, we studied the structural details of DnaG-substrate interactions by x-ray crystallography, and leveraged these data into a residue-by-residue analysis of the structural requirements for primer synthesis by DnaG. Second, we elucidated the mechanisms of action of a series of primase inhibitors.

Through this effort, we have been able to produce the first glimpse of substrate-bound primase (**Chapter 2**). Notably, this accomplishment leads to a unified model for primer synthesis. We also developed new tools for assaying DnaG activity (**Chapter 3/6**). These developments in turn have allowed us to determine the mechanisms of primer synthesis regulation by physiologically important small molecule inhibitors of DnaG (**Chapters 4/5**). The net result is the realization that DnaG interacts with its substrates through fundamentally different structural paradigms.
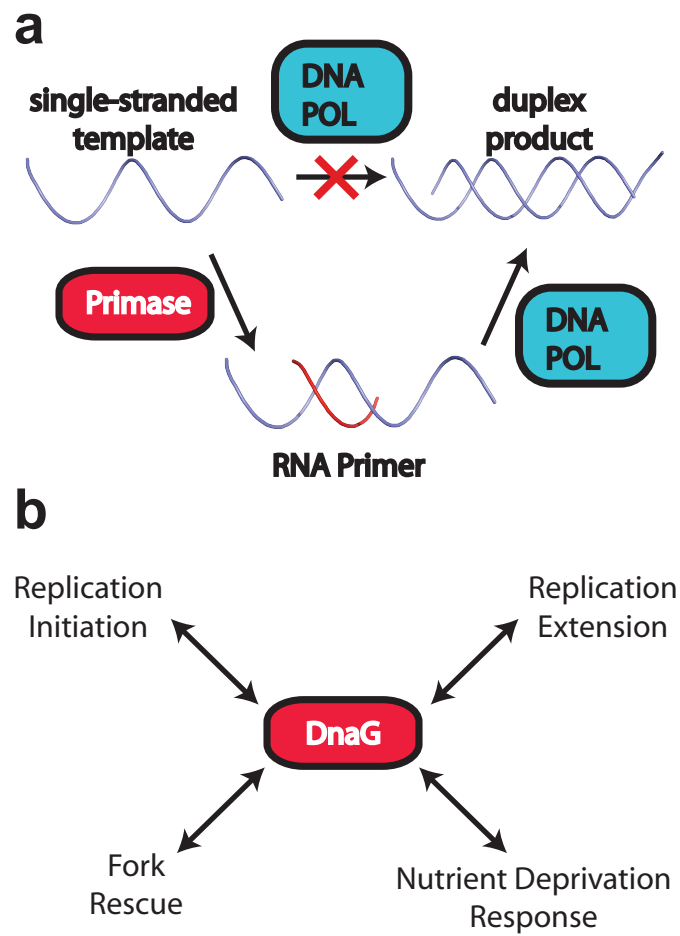
Figure 1.1: The roles of DnaG

a) A diagram illustrating the role of a primase in DNA replication. In bacteria, primer synthesis is carried out by DnaG. b) A diagram illustrating the overall involvement of DnaG in multiple cellular processes.
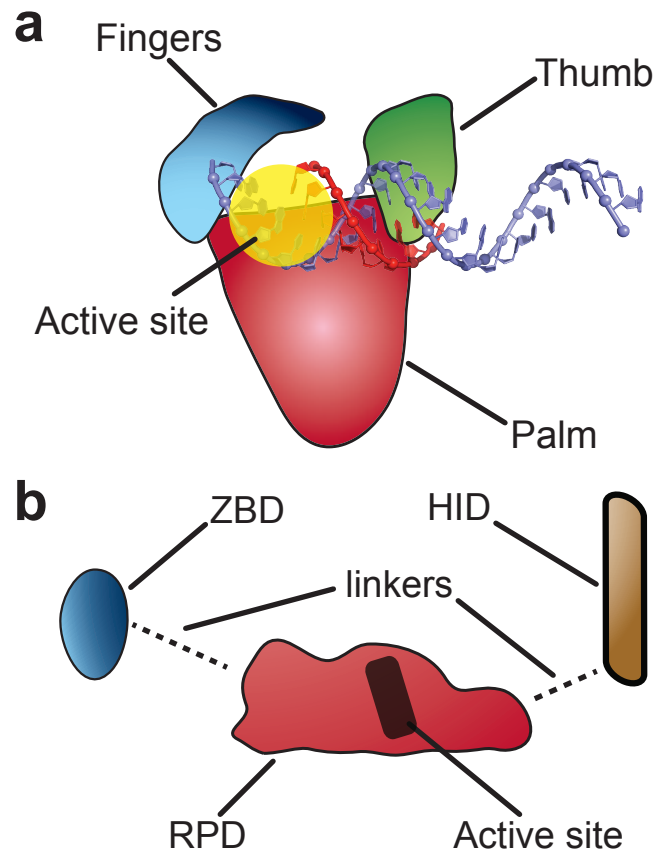
Figure 1.2: A comparison of the domain structures of canonical polymerases and DnaG.
a) A schematic of the domain structure of a canonical polymerase with the active site high-lighted, and the mode of substrate engagement indicated by the duplex DNA diagram. The blue helix strand represents the template strand, and the red helix strand represents the product strand. b) Schematic of DnaG domain structure. ZBD, Zinc Binding Domain. RPD, RNA Polymerase Domain. HID, Helicase Interaction Domain. The active site is denoted by a gray box. White sphere on ZBD represents a zinc ion.
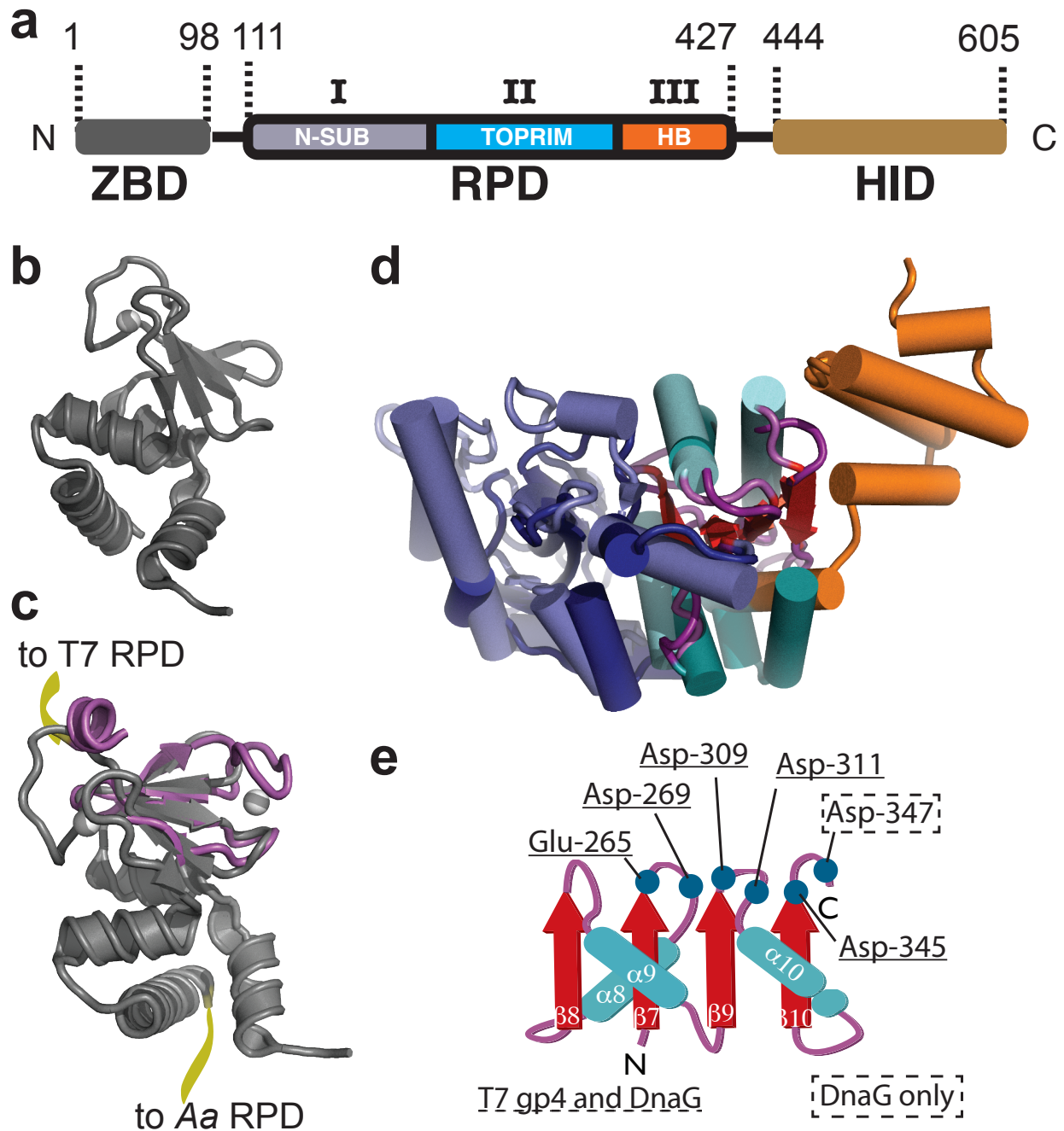
Figure 1.3: The structures of the domains of DnaG

a) A schematic of the primary structure of DnaG. Subdomains of the RPD: I) N-SUB – N-terminal subdomain, II) TOPRIM fold, and III) HB – Helical Bundle. b) A cartoon representation of the crystal structure of the Zinc Binding Domain of DnaG as represented by the *A. aeolicus* 1-401 structure (PDB code 2AU3), only the ZBD is shown, The bound zinc ion is shown as a white sphere. c) A superposition of and the *A. aeolicus* and T7 gp4 ZBDs. Both proteins are shown as cartoon representations, with the *Aa*ZBD shown as in (b), and the T7 gp4 ZBD from the T7 primase fragment crystal structure (PDB code 1NUI) is shown in purple. d) A superposition of and the *A. aeolicus* and T7 gp4 RPDs. Continued on next page.

Both proteins are shown as cartoon representations, with the *Aa*RPD colored by subdomain according to (a). The T7 gp4 RPD from the T7 primase fragment crystal structure (PDB code 1NUI) is shown colored according to subdomain, but using a darker shade of the relevant subdomain color as is depicted in (a). e) A schematic of the TOPRIM fold shared by DnaG and topoisomerase II. β-strands are shown as cyan arrows, α-helices as red ovals, and linkers as purple lines. Helices and sheets are numbered according to *Ec*DnaG. Conserved acidic amino acids are shown as blue circles, and numbered according to *Ec*DnaG. Only a subset of acidic amino acids are preserved between DnaG and T7 gp4.
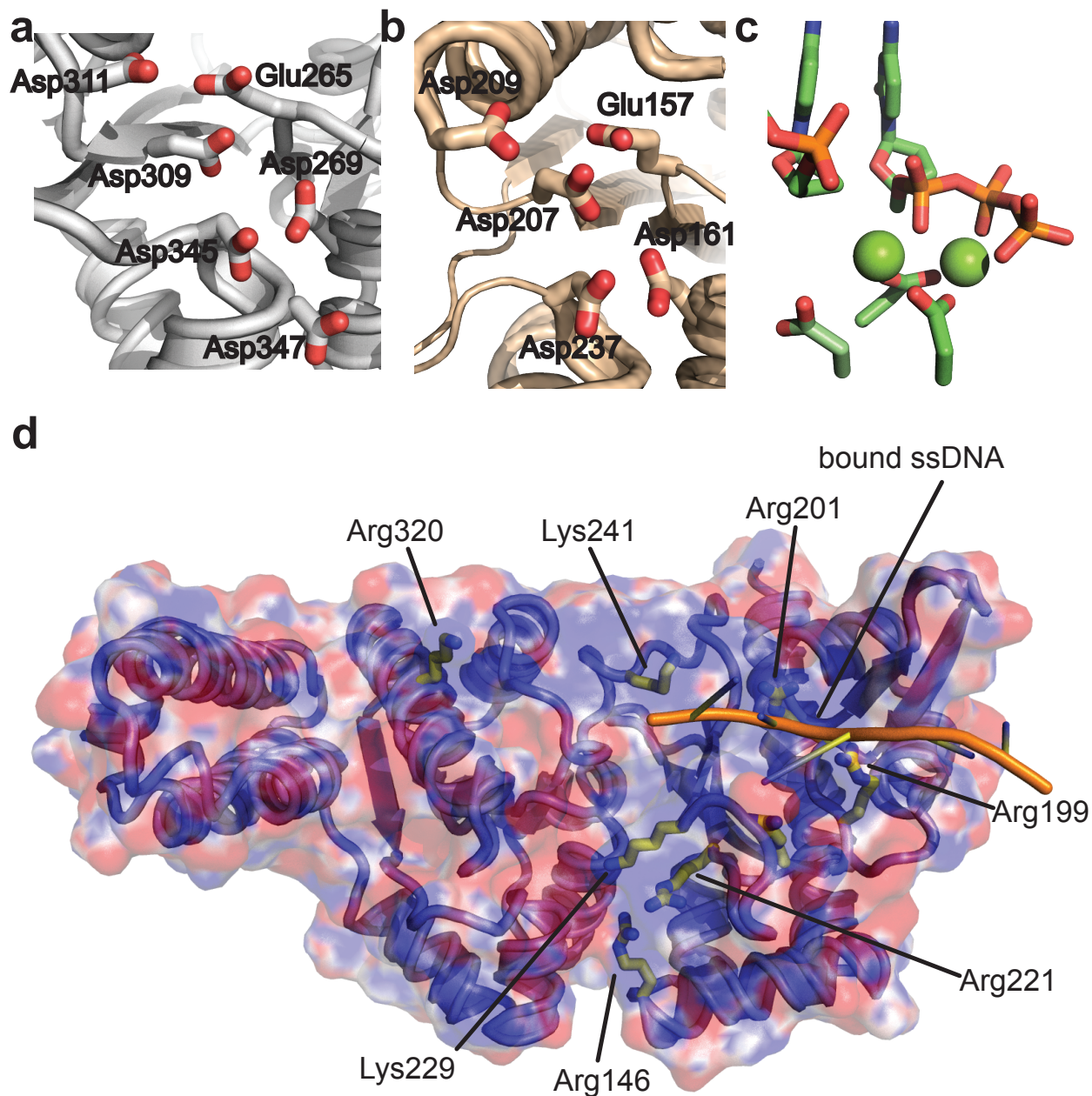
Figure 1.4: The unique active site architecture of DnaG-type primases

a) View of the apo E. coli active site with backbone shown in cartoon form, catalytic acidic residues shown as sticks with oxygens colored red, and labeled according to their identity and E. coli primary structure position. b) View of the magnesium-bound T7 gp4 primase fragment (PDB code 1NUI) active site, with backbone shown in cartoon format, catalytic acidic residues shown as sticks with oxygen colored red, and labeled according to their identity and T7 gp4 primary structure position. c) A model of the active site in a canonical polymerase. The three sidechains shown represent the catalytic triad of canonical polymerases. Green spheres represent Mg^{2+}. d) A surface and cartoon representation of the *E. coli* DnaG RPD, colored according to surface charge potential. Surface charge potential was calculated with DELPHI and CHARMM23b forcefields. The location of ssDNA binding site is indicated. Conserved, catalytically important basic residues are highlighted.

13
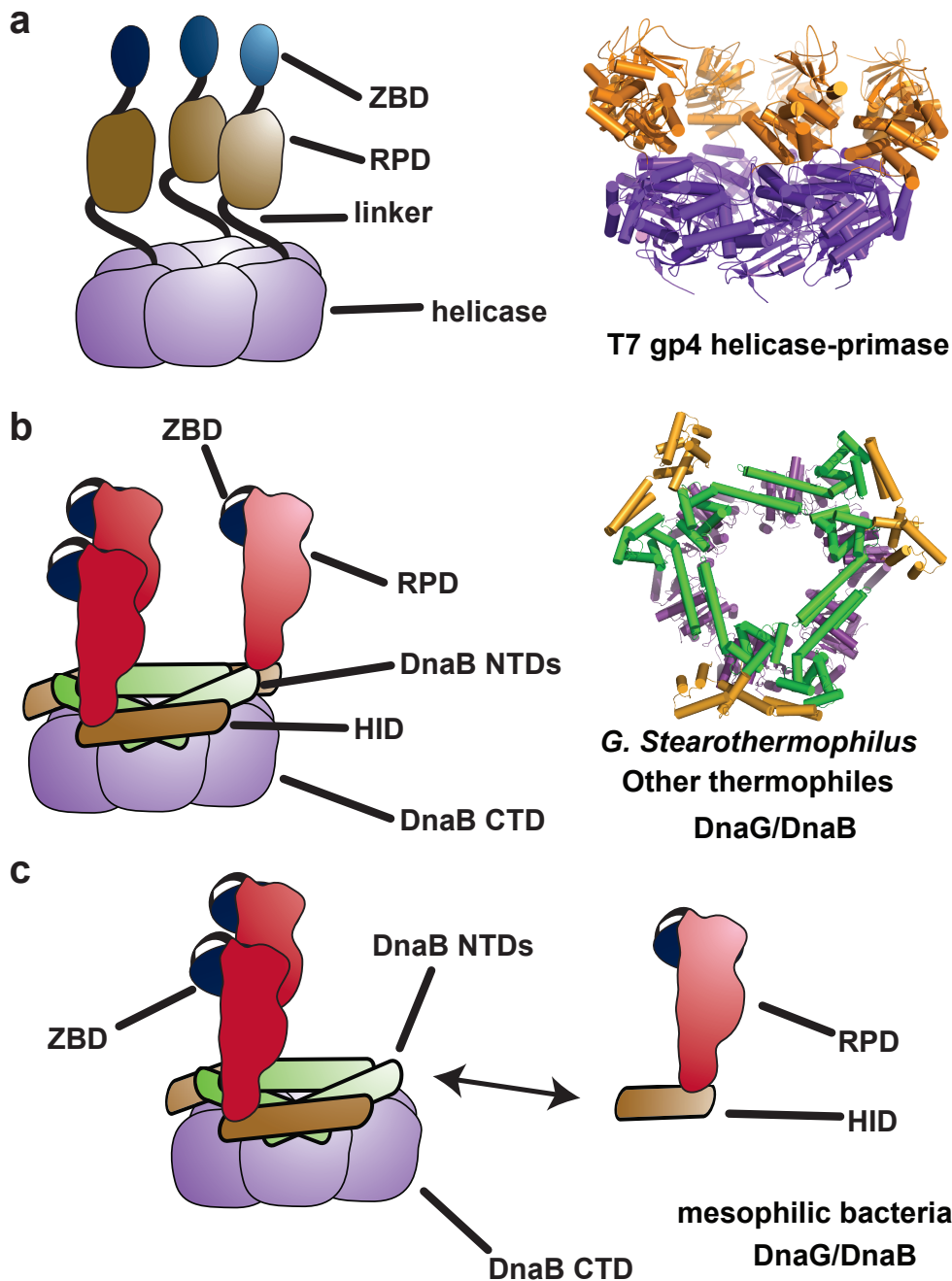
Figure 1.5: Forms of primase-helicase interactions.
a) Covalently tethered. The T7 gp4 primase-helicase is a prime example. Crystal structure: Toth EA and coworkers (2003). ZBD, Zinc Binding Domain, RPD, RNA Polymerase Domain.
b) Stable, non covalent interaction. Both T4 gp61-gp41 and *B. stearothermophilus* are examples. HID, Helicase Interaction Domain, DnaB NTD, N-terminal Domain, DnaB CTD, C-terminal Domain. Crystal structure of the Bst. DnaB hexamer bound to 3 DnaG helicase interaction domains. Bailey S. and coworkers (2007). c)Non-covalent, transient interaction. Most mesophilic bacteria fall into this category. Double headed arrow refers to repeated cycles of association and dissociation.
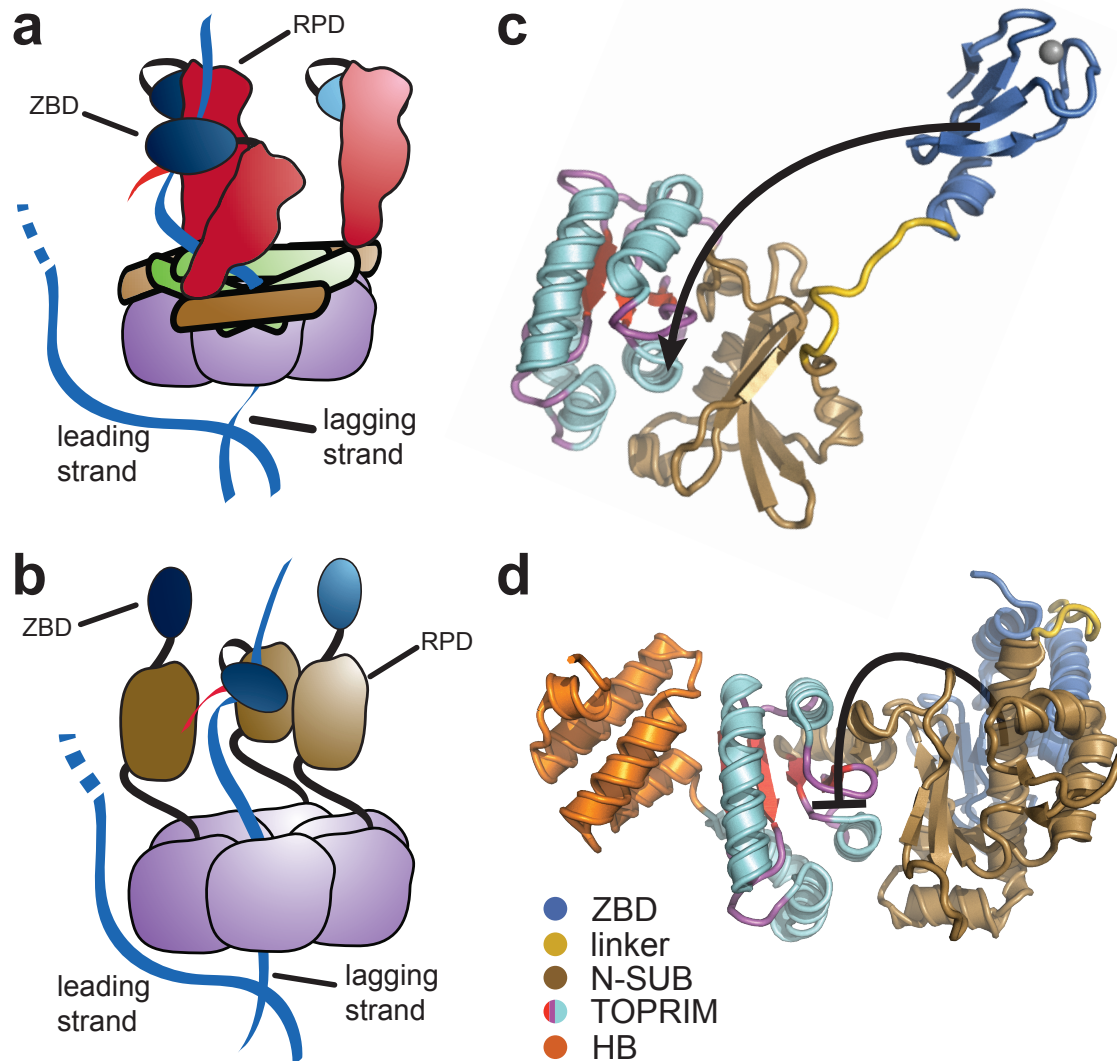
Figure 1.6: A comparison of the locations of the Zinc Binding Domain relative to the active site in bacteria and T7 gp4.

a) A schematic of synthesis in *trans*, as is thought to occur in bacterial systems. Coloring as in **Fig. 1.5c**. b) A schematic of synthesis in *cis*, as is thought to occur with T7 gp4. Coloring as in **Fig. 1.5a**. c) Cartoon representation of the T7 gp4 primase fragment crystal structure (PDB code 1NUI) colored by domain as in (d). T7 gp4 lacks a c-terminal helical bundle. Black arrow denotes the conformational change required in order for the ZBD to dock with the active site from its current location above the RPD as proposed by Kato et al. (2003). d) Cartoon representation of the *A. aeolicus* ZBD-RPD crystal structure (PDB code 2AU3) colored by domain, and showing the location of the ZBD relative to the RPD, and the conformational change required (black arrow) for the ZBD to reach the active site from its current location behind the RPD, which would involve rotating around the RPD to reach the active site.
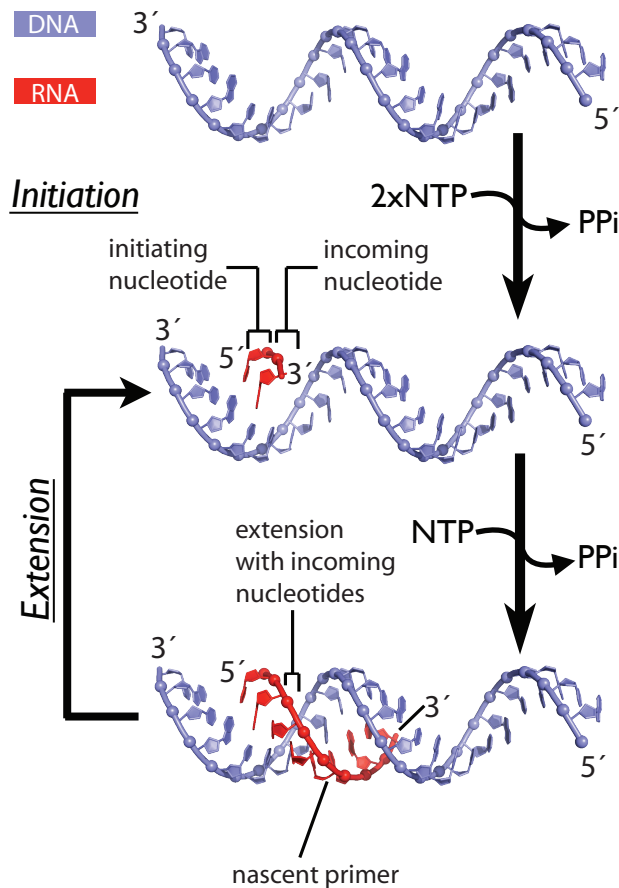
Figure 1.7: A diagram of the catalytic cycle of DnaG.

Two modes of synthesis were identified by Frick DN and coworkers (1999), initiation and extension, which can be arranged into a catalytic cycle with template binding. The cycle begins with template (blue cartoon single-stranded DNA) binding, followed by association with a 5′ initiating nucleoside triphosphate, and a 3′ incoming nucleoside triphosphate. Catalysis of phosphodiester bond formation leads to a dinucleotide product, represented here by two cartoon nucleotides in red. In the second mode of synthesis, the extension phase, the enzyme processively extends the dinucleotide with additional rounds of synthesis, resulting in an RNA oligonucleotide primer, represented here by a red cartoon strand.

Chapter 2: Discovery of the incoming nucleotide-binding site of DnaG leads to a reconciled model of primer synthesis

INTRODUCTION

Structures of DNA and RNA polymerase ternary complexes (consisting of protein, a nascent product•template duplex, and an incoming nucleotide) have provided the primary basis for elucidating how disparate domains of these enzymes – the so-called fingers, palm, and thumb – function in substrate recognition and catalysis (Brautigam and Steitz 1998). Unfortunately, the starkly different domain structure of DnaG diminishes the utility of this framework in understanding bacterial primase function. Moreover, the same methods that led to an understanding of the mechanisms of substrate binding by polymerases have attained only limited success with DnaG-type primases.

In the absence of substrate-bound states, researchers have advanced several theoretical models to explain how DnaG might form a productive ternary complex with template, NTPs, and/or a newly-formed primer. The first crystal structures of apo *E. coli* DnaG allowed two teams to simultaneously propose an initial model (**Fig. 2.1a**) (Keck, Roche et al. 2000;(Podobnik, McInerney et al. 2000), in which template DNA is bound by a highly-conserved region, termed the basic-ridge, and the product primer extends from the active site into a positively-charged basin. More recently, the crystal structure of the *E. coli* RPD bound to single-stranded DNA concretely defined the location of template binding on DnaG, and indicated that the product duplex might instead extend across the basic-ridge, with template turning away from the active site and into the positively charged basin, which is now known to bind ssDNA (**Fig. 2.1b**) (Corn, Pelton et al. 2008). This last model is more consistent with how canonical polymerases are known to bind substrates, and provides access of the product primer to single-stranded DNA binding protein (SSB), a critical interaction in replisomal dynamics.

An alternative model has been proposed for the phage T7 gp4 primase, a distant relative of DnaG (**Fig. 2.1c**) (Kato, Ito et al. 2003). Here, the crystal structure of the protein bound to Mg^{2+}, combined with NMR data in the presence and absence of a preformed primer-template complex, suggest that the regulatory N-terminal Zinc-binding domain of the enzyme stabilizes the primer product in *cis*. This interaction would lead the template strand to traverse diagonally across the protein, from the middle of the T7 gp4 N-terminal subdomain to the TOPRIM fold. A particularly cogent feature of this model is that it provides a mechanism for both primer start site selection and length regulation. However, the relevance of this model to substrate engagement mechanisms by DnaG itself is unknown, as T7 gp4 and DnaG are poorly related outside of the TOPRIM fold. In particular, T7 gp4 lacks a portion of the conserved basin, which plays important roles in models of the bacterial primase bound to substrate.

Despite their explanatory power, all of these schemes suffer from a similar deficiency;

namely, the paucity of structural data for substrate binding to the active site of the enzyme. Without these insights, the orientation of primer synthesis with respect to the TOPRIM fold has remained unknown, a feature that differentiates many of the proposed models. As a consequence, the mechanisms of substrate recognition and catalysis by DnaG likewise remain ill-defined, particularly in comparison to other polymerases.

Barriers to conducting structural investigations of DnaG have stemmed largely from its challenging biochemical properties. In particular, the enzyme has a relatively low affinity for substrates such as RNA·DNA heteroduplexes (~50 μM) (Mitkova, Khopde et al. 2003), which has precluded the straightforward determination of pertinent co-crystal structures. However, DnaG-type primases are known to productively bind individual nucleotides in the mid-micromolar range (Mendelman, Beauchamp et al. 1994; Swart and Griep 1995), a property that should in principle allow for the use of high NTP concentrations to drive complex formation. Here, we have taken advantage of this property to obtain several crystal structures of the *S. aureus* DnaG RPD in complex with individual nucleotides and their metal-ion cofactors. These structures, together with biochemical studies and structural comparisons that validate these interactions, reveal an unexpected location and orientation for nucleotide binding within the TOPRIM region of the enzyme. The observed manner of binding in turn both highlights a mechanistic role for a number of conserved amino acids in the DnaG family and reconciles the disparate models of DnaG-substrate engagement present in the literature.

RESULTS AND DISCUSSION

Structures of S. aureus DnaG RNA Polymerase Domain•NTP complexes
Initial efforts to soak NTP into crystals comprising the RPDs of either *E. coli* or *Aquifex aeolicus* DnaG were unsuccessful, likely due to tight packing environments that occluded their respective catalytic centers (Keck, Roche et al. 2000; Podobnik, McInerney et al. 2000; Corn, Pease et al. 2005). To circumvent this problem, we surveyed the DnaG RPDs from several bacterial species for a more suitable crystal form. This investigation identified crystallization conditions for a previously uncharacterized RPD from *Staphylococcus aureus* (*Sa*DnaG), which diffracted to 2 Å resolution and contained an arrangement of protomers with solvent-accessible active sites (**Fig. 2.2a**). **M**olecular **R**eplacement (MR) was used to generate initial phases for these data, which allowed for the subsequent building and refinement of the structure (**Methods**). The final model has an R_{work}/R_{free} of 18.4%/21.9%, with excellent geometry as judged by MolProbity (**Table 2.1**) (Chen, Arendall et al. 2010). The *S. aureus* RPD superposes well with other RPDs solved to date, having an average RMSD of ~ 1 Å (**Fig. 2.2b**); only the relative orientation of a poorly-conserved C-terminal helical bundle differs appreciably among the structures.

The *Sa*DnaG RPD crystallographic system facilitated the soaking of divalent metal ions and various NTPs to obtain nucleotide-bound complexes (**Methods**). Manganese initially was chosen over magnesium for these studies, since primase works equally well with either cofactor (Godson, Schoenich et al. 2000; Rodina and Godson 2006), and because the spectral properties of manganese enabled the unambiguous determination of its presence from

anomalous diffraction data. Crystals treated in this manner diffracted to a similar resolution as apo crystals, indicating that the soaking procedure did not induce crystal damage. Structures were determined for each of the four ribonucleotide substrates using MR, and refined to a similar level of quality as the nucleotide-free model (**Methods, Table 2.1**). The resultant maps all showed a single region of strong, contiguous difference density inside the RPD active site that could be best fit by a single nucleoside triphosphate and three Mn^{2+} ions (**Fig. 2.3a**). Anomalous difference maps confirmed the binding of manganese. Inspection of the four NTP-bound structures revealed that the first two subdomains of the RPD form the nucleotide-binding pocket (**Fig. 2.3b**). One subdomain consists of the metal-binding center of the TOPRIM fold, which coordinates all three divalent metals using a constellation of six highly-conserved acidic amino acids (**Figs. 2.3c, S2.1, S2.2a**). The other forms an abutting ridge that is rich in invariant, basic amino acids. One of the observable Mn^{2+} ions (denoted here as metal A) is offset from the other two, and makes no direct contacts with the nucleotide; by contrast, metals B and C closely associate with one side of the nucleotide triphosphate moiety, which is further sandwiched by two arginines (Arg146, Arg221) and a lysine (Lys230) from the basic ridge. Overall, the B-factors of the bound ligands (35 $Å^2$) are close to the average B-factor for the entire protein (27 $Å^2$), indicating that the ligands are well ordered and bound at high occupancy (**Table 2.1**). Three of the nucleotides (CTP, GTP and UTP) were seen to bind in a single conformation, whereas ATP adopted two orientations due to a rotation between its ribose and a-phosphate groups. An alignment of the GTP-bound structure with the *A. aeolicus* or *E. coli* DnaG RPD structures reveals that the nucleotide occupies a consistently solvent-exposed region in all of the structures (**Fig. S2.2b**). By varying the substituents of the soaking conditions (**Methods**), we established that: 1) NTP binding is metal dependent (**Fig. S2.2c**), 2) nucleotide is required for metal C to bind (**Fig. S2.2d**), and 3) Mg^{2+} and Mn^{2+} are essentially interchangeable (**Fig. S2.2e**).

The DnaG nucleotide-binding site is preserved with other TOPRIM-dependent enzymes
Having established the non-proteinaceous density associated with the *Sa*DnaG RPD after soaking was a metal•NTP complex, we next asked if the nucleotide-binding configuration was adventitious, or if it reflected a pertinent mode of association consistent with primase function. One clue to resolving this question came from the relationship between DnaG and type IA/II topoisomerases. Both enzyme families employ a TOPRIM fold to promote metal-assisted catalysis of nucleotidyl phospho-transfer reactions – nucleotide addition in the case of DnaG, and reversible DNA strand scission through a catalytic tyrosine in topoisomerases (Aravind, Leipe et al. 1998) (**Figs. 2.4a-b, S2.3**). The availability of multiple topoisomerase structures, in particular a crystal structure of yeast topoisomerase II (topo II) captured as a metal-associated product complex with DNA (Schmidt, Burgin et al. 2010), afforded the opportunity to compare our metal-NTP *Sa*DnaG complexes with a homologous, catalytically competent system.

To carry out this analysis, we superposed the catalytic centers of topo II and the *Sa*DnaG RPD. Only conserved C_α positions within the TOPRIM folds of the two enzymes were used in the alignments, so as not to bias the relative relationship between bound substrates. Although significant variation is seen in the position and length of the outer a-helices that

comprise both TOPRIM folds, the three internal b-strands and turns bearing the preserved catalytic acidic residues align closely (**Fig. 2.4c**). The resulting comparison shows that the single catalytic metal of topo II, which engages both the reactive phosphate and the 3′-OH of the cleaved DNA strand, is coordinated by a pair of acidic amino acids (Glu449/Asp526), and that the homologous counterparts to these residues (Glu266/Asp310) ligand metal A in *Sa*DnaG. The NTP bound to the *Sa*DnaG active site also occupies the same location as the +1 nucleobase in topo II, which marks the 5′ portion of the cleaved strand resulting from DNA cleavage by this enzyme (**Fig. 2.4d**). In the yeast ternary complex, the scissile phosphate associated with this nucleotide is linked covalently to the active-site tyrosine of the enzyme; superposition of the *Sa*DnaG and topo II TOPRIM folds results in a near-exact overlap of this phosphate and the a-phosphate of the bound NTP. Thus, the NTP-binding site of the *Sa*DnaG RPD maps to the same region of the TOPRIM domain that supports phosphoryl transfer in topo II, and further appears to pre-position the metal cofactors and reactive groups of the bound NTP in an analogous manner.

A more global inspection of the superposition between the DnaG and topo II TOPRIM folds highlights additional congruencies (**Fig. 2.4e**). For example, the orientation of the NTP in *Sa*DnaG allows for the clash-free stacking of its base against the -1 nucleobase of the DNA bound to topo II (which bears a free 3'-OH in the cleavage complex). This orientation also positions the hydrogen-bonding groups of the NTP base in-line to engage the complementary strand of the topoisomerase-bound duplex through Watson-Crick pairing (**Fig. 2.5a**). Moreover, the polarity and general position of the non-scissile DNA strand paired with the TOPRIM-bound segment in topo II overlaps with a previously-observed binding site for single-stranded template DNA on the DnaG RPD (Corn, Pelton et al. 2008). Interestingly, inspection of the DNA modeled from topo II onto DnaG placed the intact DNA strand next to both an invariant lysine in subdomain I of the RPD (Lys242) and a highly-conserved, basic amino acid (Lys321) in the RPD TOPRIM fold (**Fig. S2.4a**). In *E. coli* DnaG, the first amino acid has been shown to be important for template binding (Sun, Schoneich et al. 1999), while the latter has been shown to crosslink to the 3′-terminus of thiolated ssDNA oligonucleotides when mutated to cysteine (Corn and Berger 2007). To determine if Lys321 also plays a role in primer synthesis by DnaG through template binding, we substituted the corresponding amino acid in *E. coli* DnaG (Arg320) with either alanine or glutamate; both mutations decreased activity in accord with the severity of the mutation (**Fig. 2.5b**), and further disrupted binding to an RNA·DNA heteroduplex that mimics a preformed primer-template substrate (**Fig. S2.4b, Methods**). Thus, the position of DNA binding to topo II, as defined by its interaction with the TOPRIM fold, provides a prospective model for considering how DnaG engages its own nucleic acid substrates.

The nucleotide-binding site is critical for primase activity
Because modeling efforts such as these are by nature correlative, we next turned to mutagenesis studies to examine the manner of NTP binding. The observed NTP-binding site in *Sa*DnaG is formed by multiple highly-conserved residues, many of which have been demonstrated previously to be critical for catalysis (**Fig S2.1, S2.2a, Table 2.2**). For example, the invariant acidic amino acids that ligand the metal ions involved in nucleotide coordination have been shown to be critical for metal binding and priming activity in both *E. coli* DnaG

and in the related phage protein, T7 gp4 (Urlacher and Griep 1995; Godson, Schoenich et al. 2000,; Lee and Richardson 2005; Rodina and Godson 2006). The positively-charged amino acids that emerge from the basic ridge to contact the nucleotidyl triphosphate moiety likewise have been examined, and found to be important for primase function (Keck, Roche et al. 2000; Rodina and Godson 2006). In particular, substitutions at the amino acid equivalent to Lys229 in T7 gp4 (Lys128) interfere with both template-dependent and template-independent synthesis (Lee and Richardson 2001), directly implicating this residue in nucleotide binding.

The portion of the RPD seen to surround the sugar and base of the bound nucleotide in *Sa*DnaG has received somewhat less attention. There are two conserved loops in this region, each of which bears an invariant glycine that buttress the ribose (Gly267 and Gly287 in *Sa*DnaG) (**Fig. 2.6a**). The close contacts in this interaction suggested that substitution of these amino acids with larger sidechains should occlude binding of the nucleotide and abrogate primer synthesis. Mutation of the more N-terminal glycine to alanine in *E. coli* DnaG (Gly266) has been shown previously to disrupt priming activity (Rodina and Godson 2006); using the same system, we investigated the role of the second glycine (Gly286), and found that it, too, is required for priming activity (**Fig. 2.6b**). By contrast, neither glycine proved important for binding to an RNA·DNA heteroduplex (**Methods**, **Fig. S2.5**), indicating that disruption of primer synthesis arises from a defect in the catalytic center. Further examination of the NTP binding environment identified an invariant asparagine (Asn233) in subdomain I that forms a water-mediated hydrogen bond with the ribose 2′-OH (**Fig. 2.6c**). As with the glycine substitutions, changing the corresponding residue in *E. coli* DnaG (Asn232) to alanine abrogated primer synthesis, but had little effect on heteroduplex binding (**Fig. S2.5**).

In aggregate, the available biochemical data, combined with the comparative studies detailed above, strongly argue that the manner of $Mn^{2+} \bullet NTP$ binding to *Sa*DnaG is representative of the site and configuration used to coordinate the incoming nucleotide during strand synthesis. They also suggest that, despite their distinct active-site architecture, DnaG-type primases may employ a nucleotide-addition strategy that reflects the well-established two-metal mechanism seen in other polymerases (Steitz, Smerdon et al. 1994; Doublie and Ellenberger 1998; Kiefer, Mao et al. 1998). In this vein, we speculate that metal A would serve to coordinate and activate a 3′ acceptor hydroxyl for nucleophilic attack the a-phosphate of the incoming NTP, whereas metal B would promote dissociation of the pyrophosphate leaving group and aid metal A in transition state stabilization. Metal C, which is generally not seen in other polymerase structures, would be a distinctive feature of DnaG that could play a role in binding avidity and/or nucleδotide positioning. We note that metal C is coordinated in part by two acidic amino acids (Asp343 and Asp345 in *Sa*DnaG) that are known to be important for synthesis (Godson, Schoenich et al. 2000), but that also comprise a unique "DPD" signature motif to DnaG-family TOPRIM folds (**Fig. S2.3a**) (Aravind, Leipe et al. 1998; Godson, Schoenich et al. 2000).

Concluding Remarks
The mechanisms by which bacterial, DnaG-type primases synthesize primers in support of

21

DNA replication has long remained enigmatic. By determining how bacterial primases engage nucleotide substrates, this work helps clarifies one of the first steps that occurs during RNA strand polymerization. We find that NTPs bind to the active site of DnaG in an unanticipated manner that reconciles prior structural studies with biochemical work in both bacterial and phage systems, and accounts for the roles of several highly conserved sequence motifs in primase function. The similarities between the active sites of DnaG and type IA/II topoisomerases, together with mutagenesis efforts, support the mode of binding we observe and implicate the NTP-binding locus as the site of nucleotide addition. Interestingly, our comparisons with topo II support earlier proposals suggesting that the newly-formed heteroduplex is guided into a shallow, positively charged basin adjacent to the C-terminal helical subdomain as polymerization occurs (**Fig. 2.7**) (Keck, Roche et al. 2000; Podobnik, McInerney et al. 2000; Kato, Ito et al. 2003), rather than past the basic ridge, which actually serves as a site for triphosphate binding. However, unlike these early models, the current work agrees with later models that predict the product RNA will be highly solvent exposed, a feature consistent with the known properties of primer management by the replisome (Kusakabe, Baradaran et al. 1998; VanLoock, Chen et al. 2001). Future efforts, aimed at imaging higher-order ternary complexes between DnaG and distinct primer•template complexes, will be necessary to better define the chemical basis for primer initiation, elongation and termination, as well as to more thoroughly compare the catalytic approaches employed by DnaG to those used in other polymerase systems.

MATERIALS AND METHODS

Cloning, expression, purification
*Ec*DnaG full-length and RPD constructs were generated previously (Keck, Roche et al. 2000). The *Sa*DnaG RPD (residues 111–437) was cloned into a pET28b (Novagen) derivative with a tobacco etch virus (TEV) protease-cleavable, N-terminal hexahistidine tag. The coding DNA sequence was verified for all constructs (Elim Biopharmaceuticals). All proteins were expressed in BL21 codon+ cells and purified by Co^{2+}-affinity chromatography over a 5-mL General Electric HisTrap HP column. Binding buffer consisted of 500 mM NaCl, 10 mM imidazole, 20 mM HEPES, and 10% glycerol at pH 7.5. Gradient elution was carried out from 100% binding buffer to 100% of an otherwise equivalent buffer containing 500 mM imidazole over 10 column volumes. His6-tags were then removed with 1.5 mg of TEV protease and incubation on ice overnight. Proteins were then exchanged into binding buffer, repassaged over the HisTrap column, and finally run over a Sepharose S-200 gel filtration column (GE) in either 100 mM potassium glutamate, 20 mM HEPES (pH 7.5), 10% glycerol (*E. coli* constructs, Full-length and RPD), or 500 mM NaCl, 20 mM HEPES (pH 7.5), 10% glycerol (*S. aureus* RPD). All proteins were stored in SEC buffer. Full-length *Ec*DnaG was stored at 30 mg/mL, *Ec*DnaG RPD at 100 mg/mL, and *Sa*DnaG RPD at 8 mg/mL. Purity was assessed using polyacrylamide gel electrophoresis and Coomassie staining. Concentration was determined by absorption at 280 nm using the following extinction coefficients: 44080 $M^{-1}cm^{-1}$ for FL *Ec*DnaG, 28260 $M^{-1}cm^{-1}$ for *Ec*DnaG RPD, and 28310 $M^{-1}cm^{-1}$ *Sa*DnaG RPD. Mutant *E. coli* proteins were generated using QuikChange (Stratagene) site-specific mutagenesis, sequenced, and expressed and purified using the above procedure.

Crystallization of the SaDnaG RPD
Sparse matrix screening in the hanging-drop format was used to find initial hits, which were subsequently improved by screening precipitant, salt, buffer and additive conditions. Final crystallization conditions for the *Sa*DnaG RPD employed a well solution of 150 mM sodium thiocyanate, 100 mM Bis-tris pH 6.5, 13% PEG 3350, and 2% benzamidine. Crystals used for data collection were grown by hanging-drop in 24-well plates at 18°C using protein dialyzed into 100 mM NaCl, 10 mM HEPES pH 7.5 at 5 mg/mL. For soaks, crystals were first exchanged into well solution containing 5 mM $MnCl_2$ and 2.5 mM nucleotide, unless otherwise noted, and incubated overnight prior to cryoprotection and harvesting. Crystals were cryoprotected using a multi-step, serial-dilution process in which 2 μL of precipitant solution (supplemented with 20% glycerol and containing the ligand of interest) was diluted into 2 μL drops, followed by removal of 2 μL of the mixture. The exchange was repeated five times, resulting in a replacement of ~97% of the soaking solution with cryoprotectant solution. All crystals were flash-frozen and stored in liquid nitrogen prior to data collection at 100 K.

Data collection and refinement
Diffraction data were collected at the Advanced Light Source Beamline 8.3.1 at Lawrence Berkeley National Laboratory (MacDowell, Celestre et al. 2004). Diffraction data were indexed and integrated using XDS (Kabsch 2010) and scaled with SCALA in the CCP4 crystallography suite (Winn, Ballard et al. 2011). Phases and an initial model were obtained by MR using a polyalanine model of the *E. coli* DnaG RPD (PDB code 1DDE) wherein residues 115-249, residues 260-363, and residues 366-427 were each searched for separately using PHENIX AutoMR (Adams, Afonine et al. 2011). For soaked *Sa*DnaG RPD crystals, phases were obtained using the apo *Sa*DnaG RPD model. R_{free} flags for data collected from soaked crystals were copied from the apo data set to avoid model bias. Refinement consisted of a single round of rigid body and TLS refinement, followed by multiple rounds of manual rebuilding and refinement with the PHENIX software suite (Adams, Afonine et al. 2011). NTPs and metals were not added to models from soaked data sets until the protein portion of the model had reached convergence in refinement (defined as the R_{work}/R_{free} remaining stable for two rounds of refinement). During model building in Coot (Emsley, Lohkamp et al. 2010), real-space refinement was carried out using electron density maps calculated without the R_{free} set.

De novo primer synthesis assays
For all *de novo* synthesis experiments, reactions were carried out in triplicate, unless otherwise noted, and are reported here as an average of the three values, along with the Standard Error of the Mean (SEM) for each data point. Primer synthesis assays are based on the method described in (Koepsell, Hanson et al. 2005). All reactions were carried in a total volume of 20 μL containing a buffer (referred to as Reaction Buffer) consisting of 100 mM potassium glutamate, 20 mM HEPES (pH 7.5), 0.2 mg/mL BSA, 20 mM MgOAc and 1 mM DTT. ATP and UTP were present at 0.1 mM. Full-length *Ec*DnaG and *Ec*DnaB (expressed and purified as in (Corn, Pelton et al. 2008)) were each present at 500 nM. A 3′-phosphorylated, single-stranded oligonucleotide (5′-CACACACACACACACTGAAAGCCAAAAG-3′) was used as a template DNA at 600 nM in each reaction. The preferred primer start site is

underlined, and the 3´-cryptic G nucleotide is not copied into the primer, causing all primers to be initiated by the condensation of a 5´-ATP and 3´-GTP; hence, GTP is always an incoming nucleotide under these conditions. GTP dilutions at 4x concentration (from 4 mM to 3.9 µM) were prepared by serially diluting GTP stock solution (4 mM, in TE) 1:1 into TE. Reaction mixtures (10 µL) lacking GTP were incubated at room temperature for 10 min prior to addition of a diluted GTP solution (10 µL) to form a complete reaction mixture. Complete reactions were incubated in a sealed 384-well plate (Bio-Rad) for 1 h at 37 °C, and then stopped by addition of 10 µL of a 1:67 dilution of PicoGreen stock solution (Invitrogen) in 20 mM Tris, 50 mM EDTA, pH 7.5, giving a final stain dilution of 1:200, as per the manufacturer's instructions. Stopped reactions were incubated in the dark for 5 min, spun at 3000 x g for 2 min, and raw fluorescence measured in a PerkinElmer Victor3V multilabel plate reader. Raw fluorescence was background corrected with a no-NTP control to give Fluorescence Intensity (FI). The wild-type V_{max} was calculated by fitting resultant curves to the Michaelis-Menten model (Eqn. 1.1) using nonlinear regression. All data within a given experiment were then normalized to the wild-type V_{max}, and are reported here as Relative Primer Synthesis (RPS).

Equation 2.1 $$\text{FI} = \frac{V_{max} \times [GTP]}{K_{M,app} + [GTP]} \text{ ,}$$

where V_{max} is the FI for the maximum primer synthesis extent, and $K_{m,app}$ is the concentration of GTP that yields a primer synthesis extent equal to one half the V_{max}.

<u>*Heteroduplex binding assays*</u>
Reactions were carried out in triplicate, and are reported here as an average of the three values along with the Standard Deviation (SD) of each datapoint. For assays performed with the full-length enzyme, a published protocol was used (Corn, Pelton et al. 2008), except the ssDNA was replaced with a 5´-fluorescein labeled RNA·DNA heteroduplex with a 5´-overhang (DNA strand (IDT) sequence: 5´-AAAACTGCCGCCGC-3´, RNA strand (Dharmacon) sequence: 5´ -GCGGCGGCA-3´), and 10 mM Mg^{2+} was included in the reaction buffer. The heteroduplex was annealed by combining the DNA and RNA strands in 100 µL of TE at 100 µM each, boiling the mixture for 5 min in a 500 mL water bath, and allowing it to cool to room temperature for 8 h. The annealed heteroduplex was diluted to 10 µM, and aliquoted prior to storage at -80 °C. Enzymes were titrated from 25 µM to 25 nM. Fluorescent Polarization (FP) was measured with a PerkinElmer Victor3V multi-label plate reader, and converted to Fluorescence Anisotropy (FA) prior to background correction with a no-protein control to obtain ΔFA values.

For binding assays carried out with the *E. coli* DnaG RPD, assays conditions consisted of 1x Reaction Buffer (see *de novo* primer synthesis assay methods) supplemented with 50 nM of the same heteroduplex. FA was measured and processed as with the full-length protein experiments. Data were fit to the explicit solution to single site binding, Eqn. 1.2 (Stein,

Wilkinson et al. 2001), by nonlinear regression,

Equation 2.2

$$\Delta\text{FA} = \frac{(\Delta FA_{\max} + [P_t] + K_{D,app}) - \sqrt{(\Delta FA_{\max} + [P_t] + K_{D,app})^2 - 4 \times [P_t] \times \Delta FA_{\max}}}{2} \, ,$$

in which $\Delta FA_{\max}$ is the signal at 100% ligand binding, and $[P_t]$ is the total protein concentration.

<u>Visualization and data analysis methods</u>
Structures were visualized and figures generated using PyMOL (Schrodinger 2010). Data for activity and binding assays were processed in Microsoft Excel. Plots and corresponding fits were generated with Kaleidagraph. Sequences for sequence alignments were obtained through BLAST (Altschul, Gish et al. 1990). Alignments were calculated with MAFFT (Katoh, Misawa et al. 2002), and visualized and edited with JalView (Waterhouse, Procter et al. 2009).
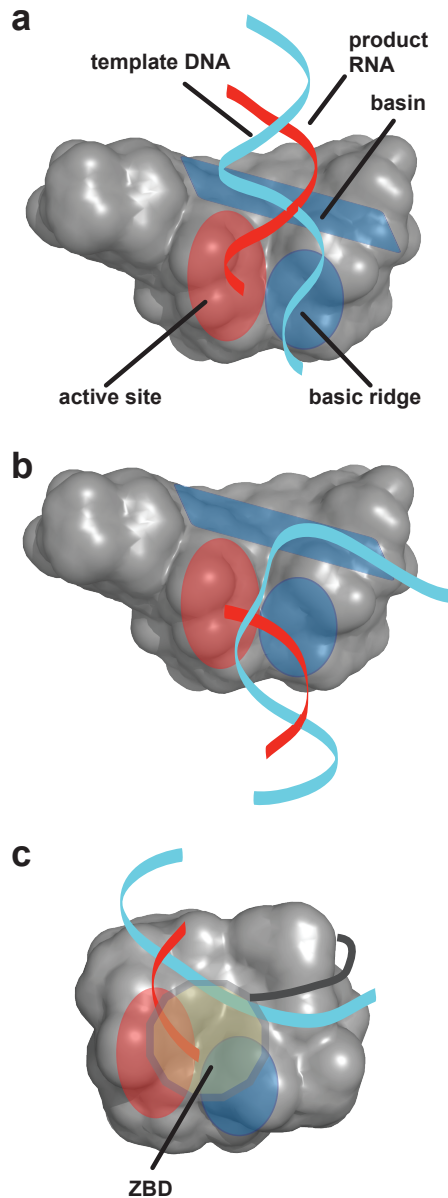
Figure 2.1: Models for primer synthesis by DnaG-type primases
a) Schematic of the model proposed simultaneously by Keck, J. L. et al. (2000) and Podobnik, M. P., et al. (2000). Gray blob represents the DnaG RPD, with regions of interest highlighted and labeled. The product RNA is extruded out the right side of the basin, while the temple interacts with the left side of the basin and the basic ridge. b) Schematic of the model proposed by Corn, J. E., et al. (2008). Coloring as in (a). The product RNA is extruded past the basic ridge out the bottom of the RPD, while the template tracks into the basin, and exits the side of the RPD. c) Schematic of the model for T7 gp4 proposed by Kato, M., et al. (2003). The RPD is represented by the gray blob, the ZBD by the yellow polygon, and template DNA and product RNA are colored as in (a). The template DNA tracks diagonally across the RPD, passing above the active site. The product RNA is extruded parallel to template, but highly solvent exposed. The active site is also solvent exposed, and is available to interact with the ZBD.

a

b

- ● E. coli (apo)
- ● A. aeolicus
- ● S. aureus (apo)
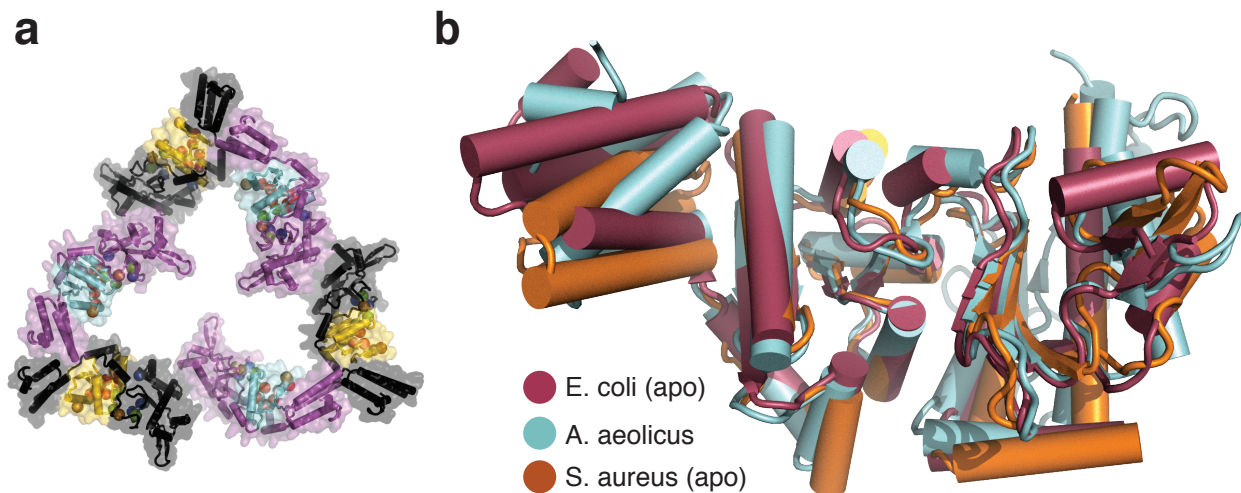
Figure 2.2: Structure of the *S. aureus* DnaG RPD

a) The *S. aureus* DnaG RPD active site is solvent accessible. Cartoon and surface representation is shown for the crystal packing arrangement of *S. aureus* DnaG RPD protomers. Six protomers, related by crystal symmetry, are shown. Three are colored purple with cyan active sites; three are black with yellow active sites. Spheres denote the Cα positions of functionally important residues in the active site. b) Superposition of several DnaG RPDs solved to date. Structures, corresponding to known apo states, are shown as cartoons and colored as follows: *E. coli* RPD (PDB ID: 1DDE, maroon), *A. aeolicus* RPD (PDB ID: 2AU3, cyan), *S. aureus* RPD (PDB ID: 4E2K, orange).
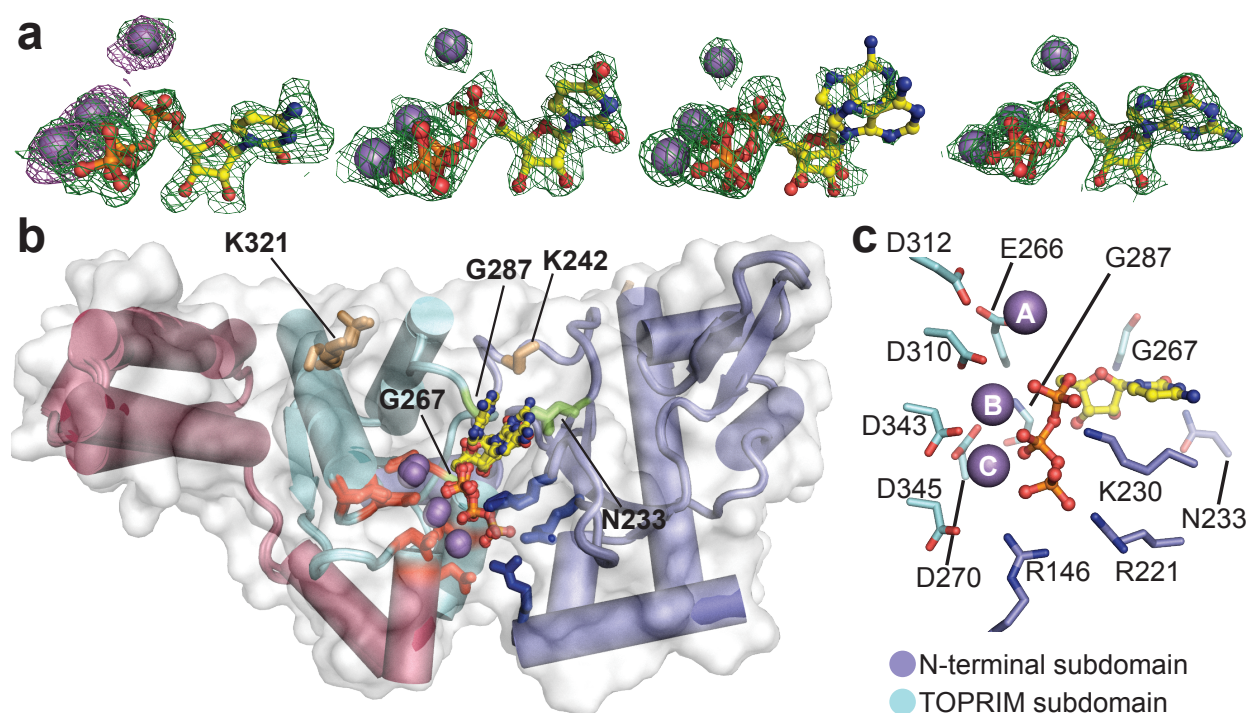
Figure 2.3: Overview of NTP-bound *Sa*DnaG RPDs

a) Difference density present in the SaDnaG RPD active site for Mn^{2+}•NTP-soaked crystals. From left to right, density corresponds to CTP, UTP, ATP, and GTP. The resulting difference density (mF_o-DF_c) is shown contoured at 3σ as a green mesh. Anomalous difference density maps obtained by collecting data on a CTP and Mn^{2+} soaked crystal at the Mn K-edge is shown contoured at 5σ as purple mesh. Individual NTPs are shown from right to left as ball-and-stick representations, while Mn^{2+} ions are shown as gray spheres. Maps were calculated after refinement of the protein portion of the model converged, but prior to modeling of the nucleotide (**Methods**). b) All four metal•NTP complexes bind to a common site on the *Sa*DnaG RPD. A superposition of the apo *Sa*DnaG RPD structure (white surface) and all four NTP bound *Sa*DnaG RPD structures (cartoons colored by subdomain, as per **Fig. 2.2a**) is shown. Sidechains of metal-binding residues are shown as red sticks, and sidechains of basic ridge amino acids as blue sticks. NTPs and Mn^{2+} ions are shown and colored as in (a). The sidechains of the individuals residues investigated in this study are labeled (*Sa*DnaG numbering), and are shown as stick representations. c) Coordination of metal•NTP complexes. The CTP bound structure is displayed as a representative example, and the NTP and Mn^{2+} ions are colored as in panel (a). Sidechains of residues that contact the nucleotide and metal ions are shown in stick representation, and colored according to their respective subdomain. Residue numbering is for *Sa*DnaG.
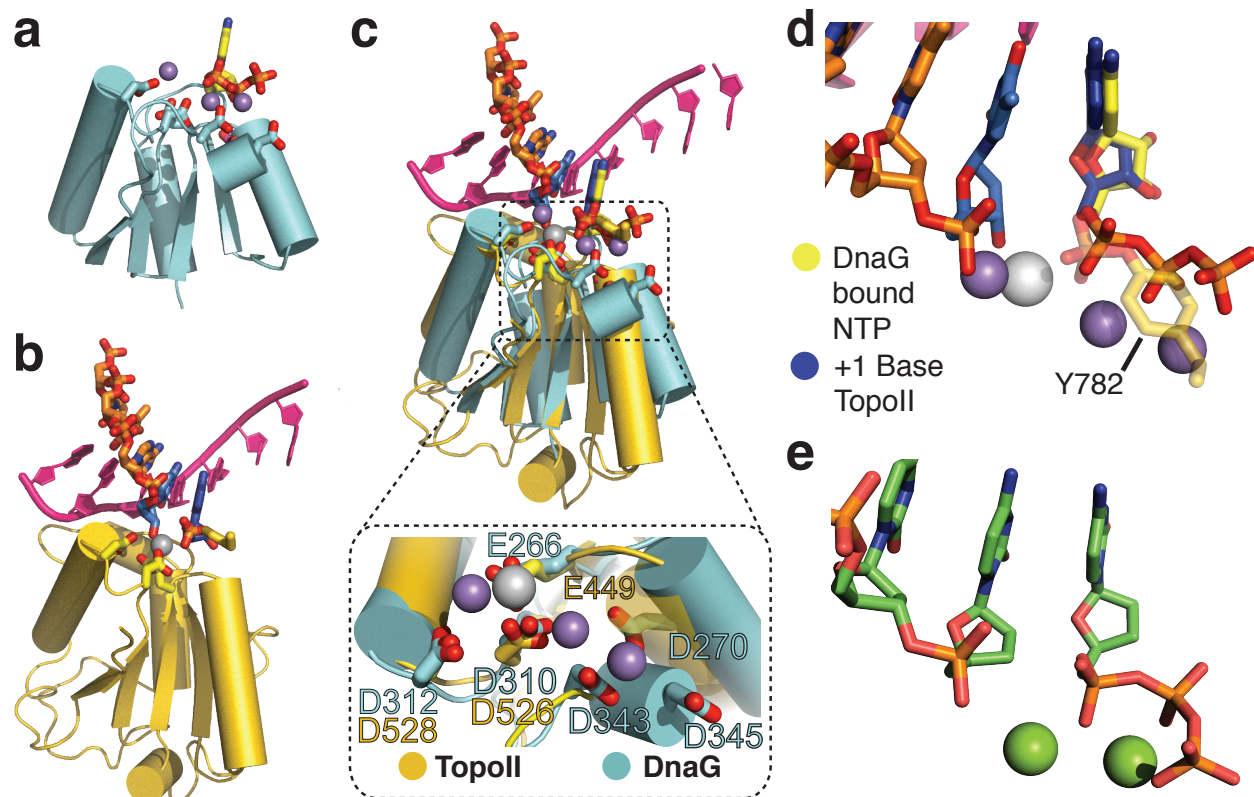
Figure 2.4: Overview of NTP-bound *Sa*DnaG RPDs

a) Cartoon representation of the *Sa*DnaG RPD TOPRIM fold bound to nucleotide (CTP). Sidechains of the conserved acidic residues in the DnaG TOPRIM fold are shown as sticks with carbon colored cyan and oxygen red. CTP and its associated Mn^{2+} ions are colored as in **Fig. 2.2a**. b) Cartoon representation of the *S. cerevisiae* topo II TOPRIM fold as found in a DNA-cleavage complex (PDB ID 3L4K). Sidechains of the conserved acidic residues and the covalently attached catalytic tyrosine (Tyr782) from the neighboring protomer are shown as sticks with carbon colored gold and oxygen in red. The scissile strand is shown as orange sticks. The -1 and +1 nucleobases at the cleavage site are colored blue. The complementary strand is shown in bright pink as a cartoon representation. The metal ion bound to the topo II catalytic center (Zn^{2+}, in this particular complex) is shown as a white sphere. c) Superposition of TOPRIM folds from CTP-bound SaDnaG RPD and the yeast topo II•DNA cleavage complex. The DnaG and topo II TOPRIM folds and sidechains, along with their associated substrates, are depicted as per panels (a) and (b), respectively. Boxed panel: close-up of the aligned metal binding regions (region highlighted by dashed outline in main panel) of both TOPRIM folds, with conserved acidic residues shown as stick representations and colored as per panels (a) and (b). The metal ions bound to topo II and SaDnaG are colored white and gray, respectively. d) Close-up of substrate configuration based on a TOPRIM-fold alignment between DnaG and yeast topo II. Only the substrates from both structures are shown, and are colored as in panel (c). The covalent linkage between Tyr782 and DNA observed in topo II is shown as a semi-transparent stick representation. e) Schematic of the active site in the T7 DNA polymerase ternary complex structure (PDB ID: 1T7P, (Doublie, Tabor et al. 1998)). The incoming nucleotide and primer strand are shown in stick representations with carbon colored dark green. Mg^{2+} ions are shown as green spheres.

a

Bound to DnaG
- ● NTP
- ● ssDNA

Modeled from topo II
- ● Scissile DNA
- ● Nonscissile DNA

b
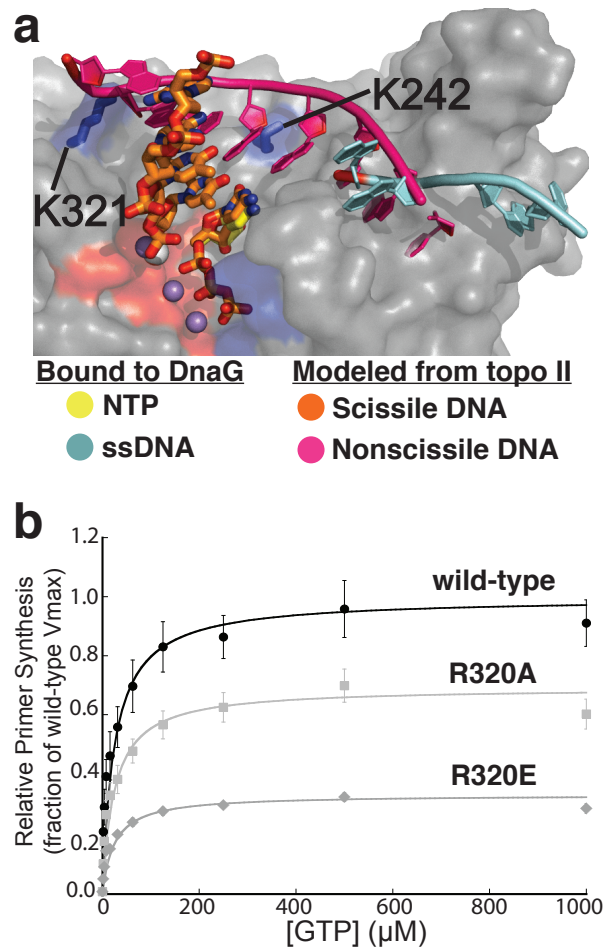
Figure 2.5: Congruencies between the binding sites for the complementary DNA strand in topo II DNA and *Sa*DnaG.
a) The polarity and terminal position of the complementary DNA strand bound to the topo II TOPRIM fold lines up with the site for ssDNA binding in the DnaG RPD. Superposition of the TORPIM folds between an S. cerevisiae topo II•DNA cleavage complex and an ssDNA-bound state of the *E. coli* DnaG RPD (PDB ID 3B39) was used to examine how a nucleic acid duplex might bind the primase active site. The scissile strand in topo II is shown as orange sticks, and its complement as a magenta cartoon. ssDNA bound to *Ec*DnaG, thought to mark the site of template binding (Corn and Berger 2007; Corn, Pelton et al. 2008), is shown as a cyan cartoon. The polarities of the magenta and cyan DNAs match, while the DNA strands partially overlap, as a result of the TORPIM-based alignment. The protein portion of the *Sa*DnaG RPD bound to CTP is displayed as a surface representation in gray, with the metal binding cluster of DnaG highlighted in red, and the basic ridge in blue. The modeling implicates residues K321 and K242 of SaDnaG (blue sticks with corresponding blue surfaces) as possibly playing a role in binding a primer•template product. b) Arg320 of *Ec*DnaG (corresponding to Lys321 in *Sa*DnaG) is required for de novo primer synthesis. Helicase-stimulated, GTP-dependent primer synthesis was assayed for the wild-type enzyme (black), Arg320Ala (light gray) and Arg320Glu (gray) mutant enzymes in a fluorometric de novo primer synthesis assay (**Methods**). Curves represent fits to a standard Michaelis-Menten kinetics model (**Methods**); error bars represent ±SEM (**Methods**).
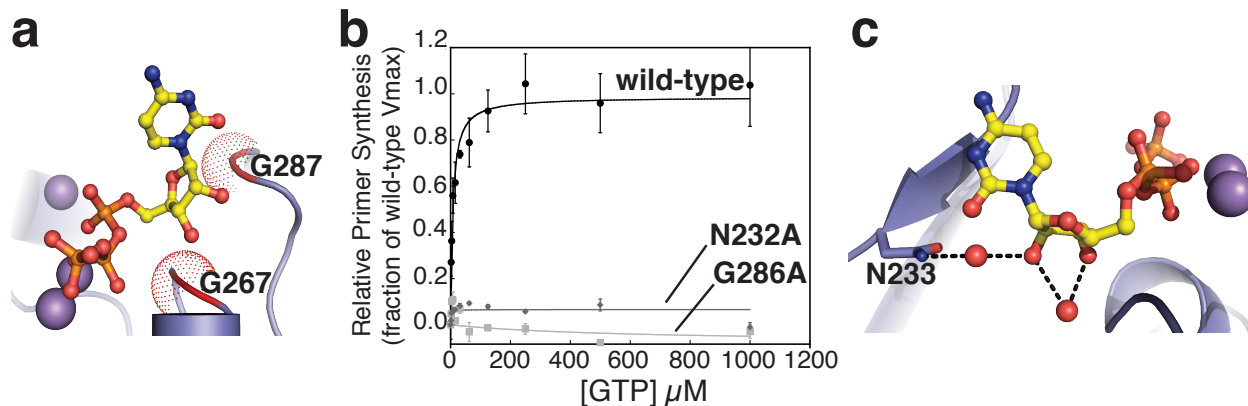
Figure 2.6: Analysis of conserved residues that engage the ribose of bound nucleotide
a) Views of the region near the ribose in the CTP-bound *Sa*DnaG RPD. The protein portion of the RPD is shown as a cartoon. Conserved residues Gly267 and Gly287 (corresponding to Gly266 and Gly286 in *Ec*DnaG) are highlighted in red, with their Cα van der Waals radii represented by red dots. b) Ribose-binding amino acids are critical for DnaG function. The results of a GTP-dependent, helicase-stimulated de novo primer synthesis assay are shown for wild-type (black circles), Gly286Ala (gray diamonds), and Asn232Ala (light-gray squares) *E. coli* DnaG. Primer synthesis extent is reported as a fraction of the wild-type Vmax. Curves represent fits to a standard Michaelis-Menten kinetics model (**Methods**); error bars represent ±SEM (**Methods**). c) Asn233 (Asn232 in *Ec*DnaG) associates with the 2′-OH of the bound nucleotide through a water-mediated hydrogen bond. Dashed lines represent hydrogen bonds.
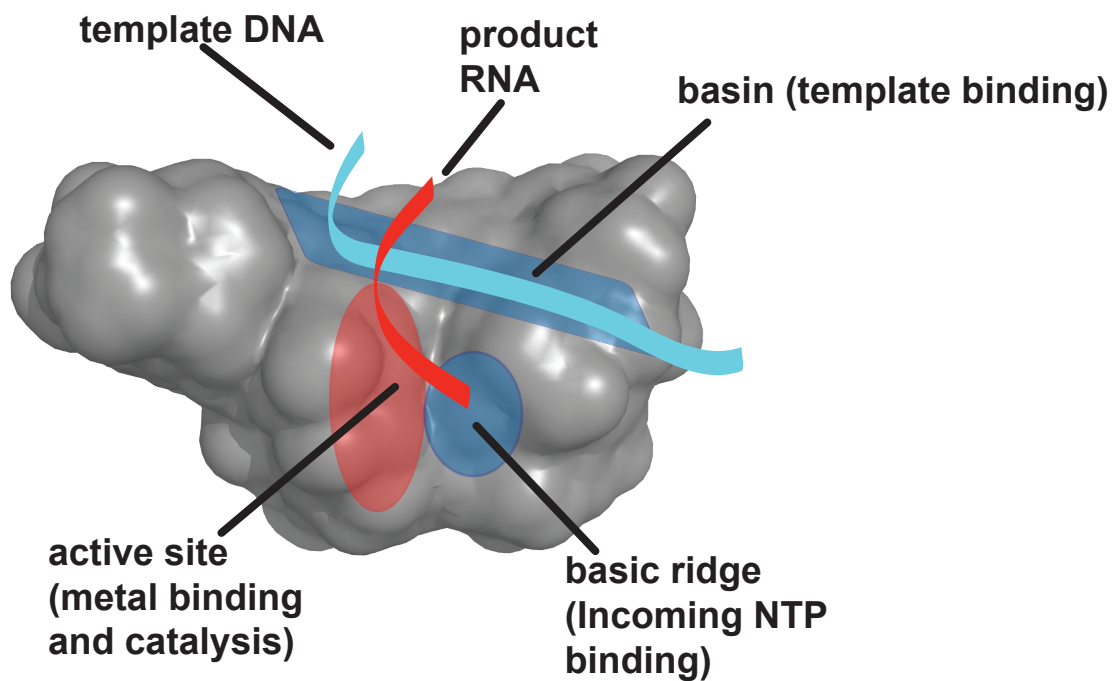
Figure 2.7: A reconciled prospective model for primer synthesis
A schematic showing the reconciled model for primer synthesis. Coloring as in **Fig. 2.1a**. As with the other models, the basin binds the template strand, but more completely than was proposed previously. The basic ridge binds the incoming NTP. The product primer is highly solvent exposed.

Figure S2.1: Sequence conservation between DnaG-family RPDs.
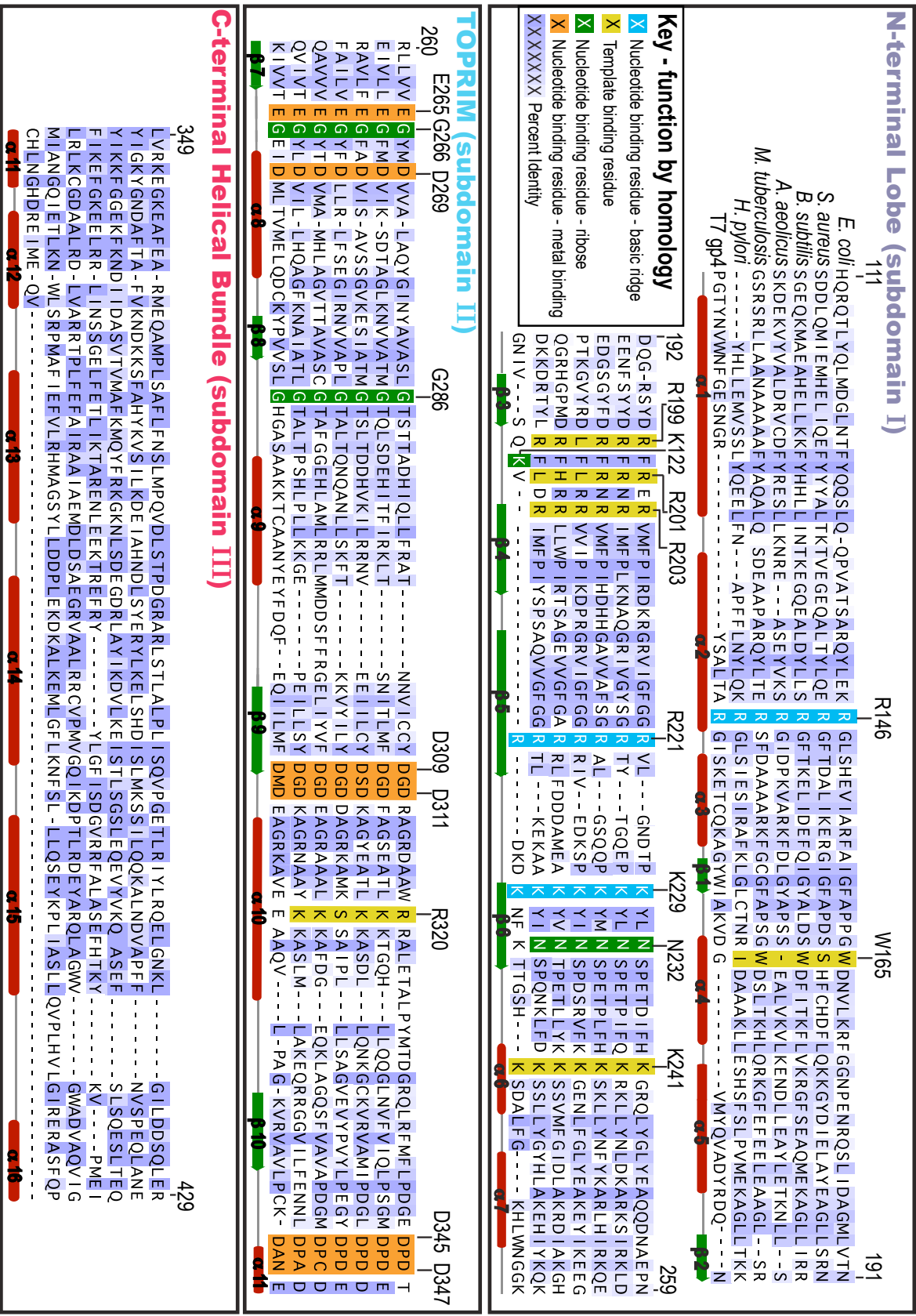
Figure Legend S2.1: Sequence conservation between DnaG-family RPDs
Sequence conservation between DnaG-family RPDs. An alignment of DnaG sequences from six bacterial species and from bacteriophage T7 gp4 is shown. Sequences were obtained and aligned as described in the **Methods**. Position numbering refers to *E. coli* DnaG, and the secondary structure of the fold as determined from the *E. coli* protein is shown below each alignment section. The percent identity at each position, shown as shaded blue boxes, was calculated by JalView (Waterhouse, Procter et al. 2009). Residues discussed in Chapter 2 are highlighted.

a

G287　G267　N233

E266

D312　D310

D343　K230

D345　R146　R222

● metal binding

● basic ridge

● ribose pocket

● peripheral

b

S. aureus DnaG RPD bound to GTP

A. aeolicus DnaG RPD modeled with GTP

E. coli DnaG RPD modeled with GTP

c

d

A　D312

x　E266

D310

B　y　D270

C

D343

D345

● NTP, Mn^{2+} soaked

○ Mn^{2+} only soaked

e
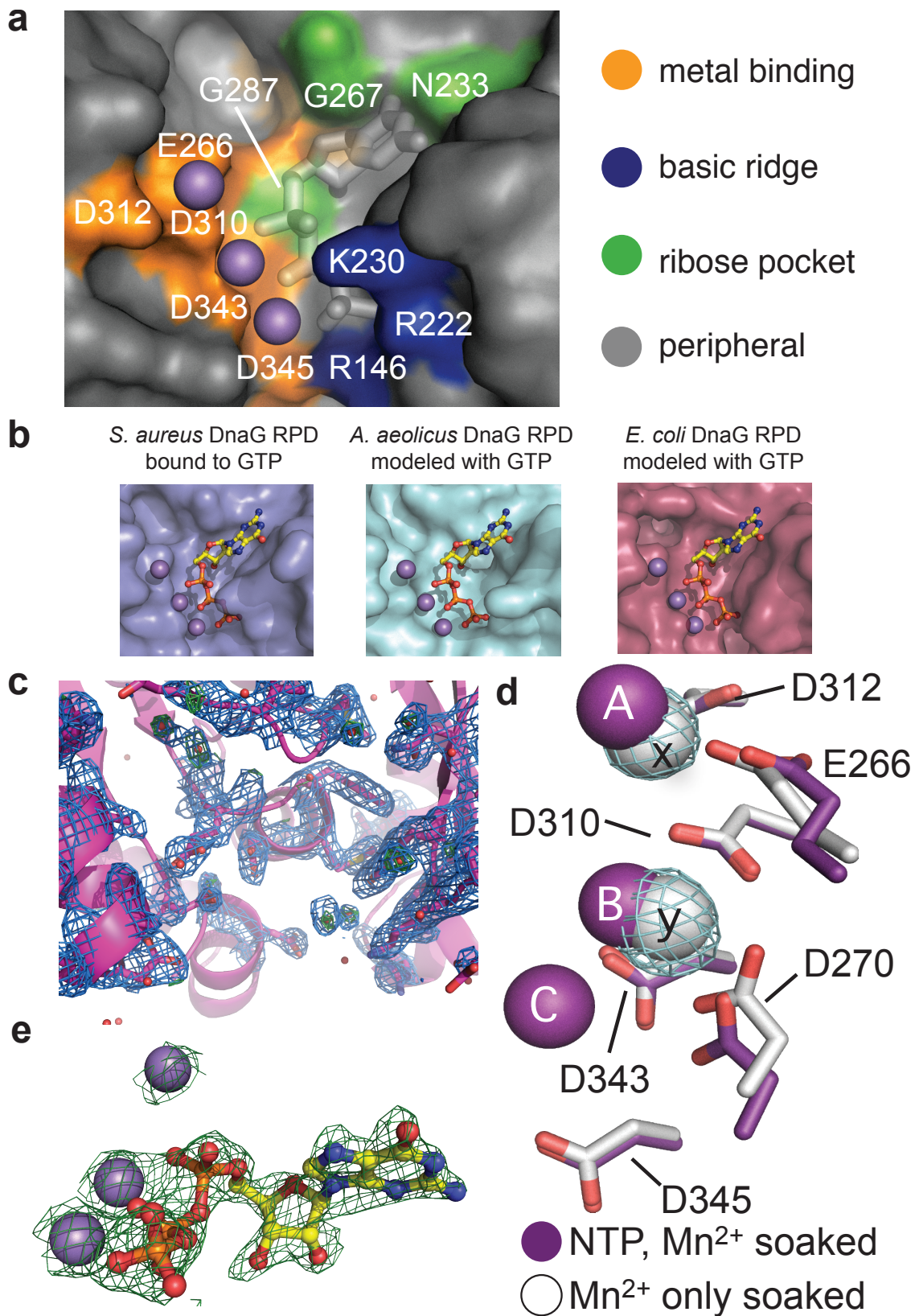
Figure S2.2: Conservation and solvent accessibility of the DnaG NTP-binding site across species.

Figure Legend S2.2: Conservation and solvent accessibility of the DnaG NTP-binding site across species.

a) Surface representation of CTP-bound *Sa*DnaG RPD with highly conserved residues highlighted. Residue numbering is for *Sa*DnaG. Bound nucleotide is shown as transparent white sticks, and associated Mn^{2+} ions as spheres. b) The DnaG nucleotide-binding site is preserved across species. Mn^{2+}•GTP was docked into the *A. aeolicus* and *E. coli* DnaG RPD structures by superposition with *Sa*DnaG RPD. GTP is shown as a ball-and-stick representation with carbons colored yellow. Mn^{2+} ions are shown as gray-spheres. Docking does not result in any clashes with either protein surface. c) Divalent metals are required for NTP binding to *Sa*DnaG. RPD crystals were soaked with 10 mM ATP in the absence of divalent salts; after data collection, processing and initial model and phase generation by MR, the active site (shown) was inspected for the presence of difference density (mFo-DFc, green mesh, 3σ). Blue mesh: 2Fo-Fc electron density at 1σ. The absence of strong or contiguous regions of difference density indicates that no ligand is bound. d) NTPs are required for the binding of metal C. Crystals were soaked with a high concentration of $MnCl_2$ (50 mM) in the absence of nucleotide to determine the specificity of the observed metal binding locations. Diffraction data were collected at the Mn-K edge, processed, and phases obtained by MR. Anomalous difference maps were calculated with PHENIX. Metals were identified as high (>10σ) peaks in anomalous (F_o-F_c) difference maps (cyan mesh). The GTP-bound structure was then aligned with the metal-soaked model to compare the position of metal binding in the presence (purple spheres) or the absence (white spheres) of nucleotide. Sidechains are colored according to their respective dataset. e) Analysis of the influence of high concentrations of Mg^{2+} on nucleotide binding. Difference density (mF_o-DF_c) contoured at 3σ from a data set collected from a crystal soaked with 100 mM $MgCl_2$, 2 mM GTP and 5 mM $MnCl_2$ is shown as green mesh. Maps were calculated after a single round of rigid body and TLS refinement. The difference density signal (green mesh) and the position of the bound nucleotide (yellow sticks) do not change under high Mg^{2+} ion concentrations.
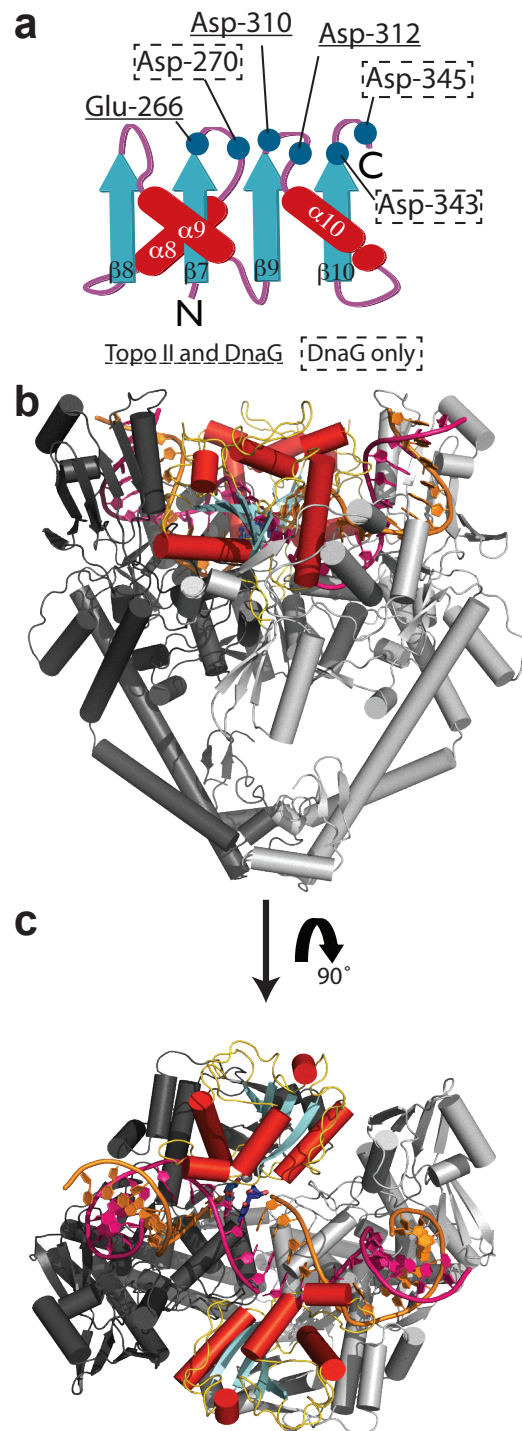
Figure S2.3: The *S. cerevisiae* topoisomerase II TOPRIM fold.

Figure Legend S2.3: The *S. cerevisiae* topoisomerase II TOPRIM fold

a) Schematic of the TOPRIM fold shared by DnaG and topoisomerase II. β-strands are shown as cyan arrows, α-helices as red ovals, and linkers as purple lines. Helices and sheets are numbered according to *Ec*DnaG. Conserved acidic amino acids are shown as blue circles, and numbered according to SaDnaG. Only a subset of acidic amino acids are preserved between DnaG and topo II, which likely explains the ability of the topo II catalytic center to bind only a single chemically-compentent metal ion (Bax, Chan et al. 2010; Laponogov, Pan et al. 2010; Schmidt, Burgin et al. 2010). b) Cartoon representation of the functional dimer that comprises the topo II cleavage complex (PDB ID 3L4K). One subunit is colored dark gray, the other light gray. The bound DNA substrate is shown as a cartoon representation, colored orange for the scissile strand and bright pink for the nonscissile strand (as defined by the dark-gray subunit on the left). The TOPRIM fold of each subunit is shown with β-strands colored cyan, and α-helices red. c) 90° rotated view of the structure displayed in panel (b). The +1 and -1 nucleobases of the cleavage site are colored blue and shown as sticks.

a Conservation of position 320 or equivalent across species

b Heteroduplex binding of *Ec*DnaG wild-type and R320 mutants
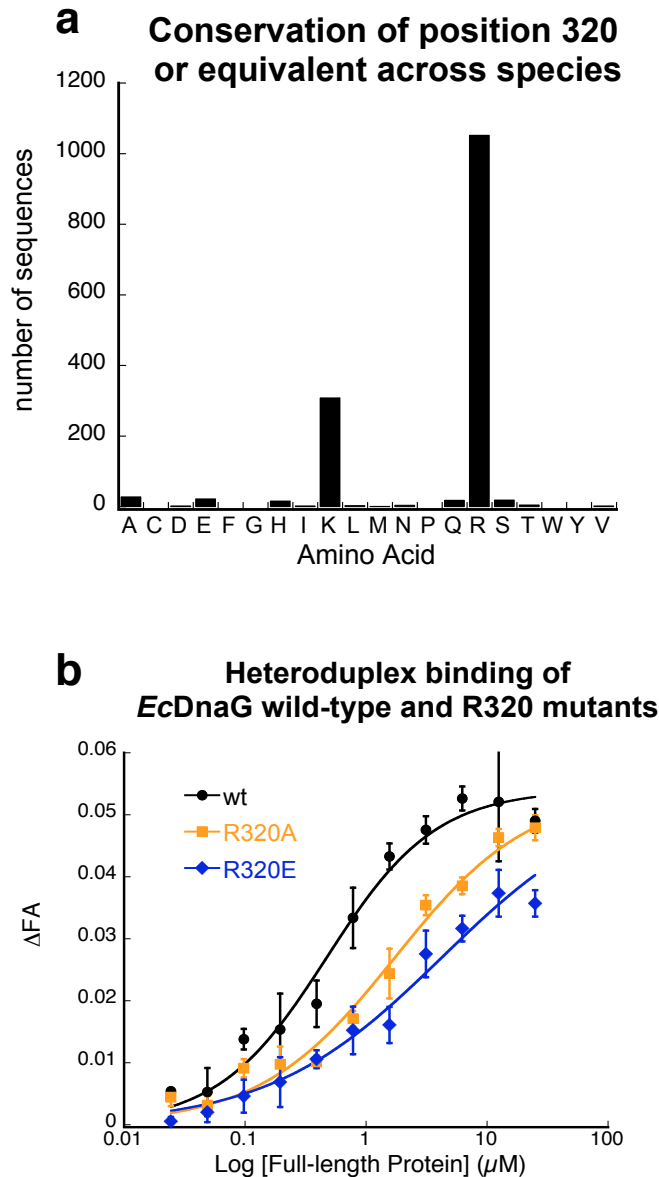
Figure S2.4: Arg320 is both highly conserved and important for heteroduplex binding in *E. coli* DnaG.
a) Plot showing the distribution of amino acids found at position Arg320 of *E. coli* DnaG. Bin height represents the number of times a given amino acid is present in ~1400 non-redundant DnaG sequences (**Methods**). b) Plot comparing the ability of wild-type *E. coli* DnaG and two Arg320 mutants to bind an RNA•DNA heteroduplex based on the change in fluorescence anisotropy of a fluorescently-labeled substrates as a function of protein concentration (**Methods**). Full-length wild-type *E. coli* – black circles; Arg320Ala – orange squares; Arg320Glu – blue diamonds. The inability to reach binding saturation for the mutant proteins precludes an accurate determination of the change in apparent Kd, but corresponds to at least a 10-fold drop in affinity for the most severe mutation; because of the uncertainty in the final saturation values, curve fits are for visual purposes only. Error bars represent ±SD (**Methods**).
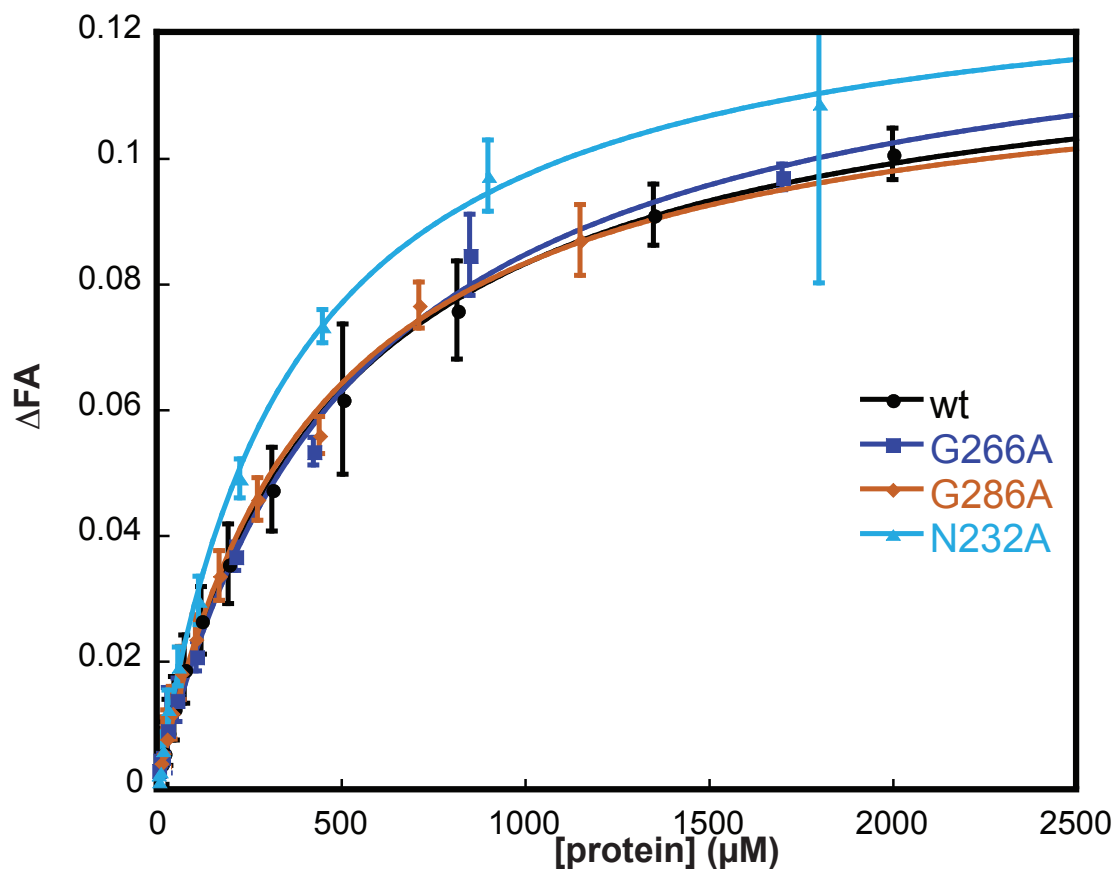
Figure S2.5: Amino acids associated with the ribose of bound NTPs are not required for heteroduplex binding.
The plot shows the change in fluorescence anisotropy of a fluorescently-labeled hetero-duplex as a function of DnaG concentration (**Methods**). Data are shown for binding to the wild-type *E. coli* RPD (black circles), and to the Gly286Ala (orange diamonds), Gly266Ala (blue squares) and Asn232Ala (cyan triangles) mutant proteins. Data were fit to the explicit solution to single site binding (**Methods**). Error bars represent ±SD (**Methods**).

Table 2.1: Data Collection and Refinement Statistics

	ApoSaDnaG RPD 4E2K*	ATP Bound SaDnaG RPD 4EDG*	CTP Bound SaDnaG RPD 4EE1*	GTP Bound SaDnaG RPD 4EDK*	UTP Bound SaDnaG RPD 4EDR*
Crystal Data					
Space Group	P 6$_1$	P 6$_1$	P 6$_1$	P 6$_1$	P 6$_1$
Cell Dimensions (a, b, c (Å), α, β, γ(°))	151.3,151.3,38.5 90, 90, 120	152.3, 152.3,38.7 90, 90, 120	152.2,152.2,38.8 90, 90, 120	152.2,152.2,38.8 90, 90, 120	150.6,150.6,38.6 90, 90, 120
Data Collection	Native	Native	Native	Native	Native
Wavelength (Å)	1.115	1.115	1.115	1.115	1.115
Resolution (Å)	50.00-2.15	34.59-2.00	30.62-1.99	49.6-2.00	37.04-2.01
Reflections (measured/ unique)	248,259/28,064	187,094/34,499	368,265/32,369	506,322/35,005	183,488/33,847
†R$_{sym}$ (%) (highest shell)	7.2 (64.7)	4.6 (28.7)	11.3 (68.2)	9.4 (49.5)	4.7 (17.3)
<I/σI> (highest shell)	30.5 (3.9)	22.2 (5.2)	14.8 (4.1)	21.8 (5.3)	10.9 (3.6)
Completeness (%) (highest shell)	100 (100)	99.8 (98.8)	94.7 (100)	100 (100)	99.9 (100)
Redundancy	8.8	5.3	11.4	10.8	5.4
Refinement					
‡R$_{work}$/R$_{free}$ (%)	19.1/21.7	16.2/19.7	18.2/21.6	16.1/20.0	17.2/20.5
Average B factor (Å²)					
protein	39.4	34.8	26.9	32.3	28.8
solvent	44.8	46.8	39.6	43.9	40.9
Rmsd					
Bond lengths (Å)	0.006	0.02	0.002	0.002	0.002
Bond angle (°)	0.77	0.715	0.637	0.681	0.735
Protein residues	319	321	321	321	321
Water molecules	240	355	348	360	377
Ions	0	3	3	3	3
§Ramachandran (favored/ generous/disallowed)	316/3/0	317/4/0	316/5/0	317/4/0	319/2/0
Ligand Refinement					
Avg. B Factor (Å²)	N/A	49.9	35.4	31.0	24.5
Real-space correlation coefficient	N/A	0.887	0.933	0.957	0.954

*PDB ID. **†Rsym** = $\Sigma| I-\langle I\rangle|/\Sigma I$, where I is the integrated intensity for a given reflection. **‡Rwork** = $\Sigma|Fobs - Fc|/ \Sigma Fobs$, where Fobs and Fc are the observed and calculated structure factor amplitudes, respectively. Rfree is calculated as for Rwork, but from a subset of the data (5 %) that was withheld from crystallographic refinement. §As calculated by MolProbity (Chen, Arendall et al. 2010).

Table 2.2: Putative functions of conserved residues in DnaG-family RPDS

E. coli Residue	*S. aureus* Residue	T7 gp4 Residue	Prospective Role	Reference
W165	S165			Corn and Berger 2007
R199	R200			Corn and Berger 2007
R201	R202		Template Binding	Corn and Berger 2007
K241	K242	K137		Sun, Schoneich et al. 1999
R320	K321			Corn and Berger 2007; *This study*
E265	E266	E157	Metal A binding	Godson, Schoenich et al. 2000; Keck, Roche et al. 2000; Kato, Ito et al. 2003; Lee and Richardson 2005; Rodina and Godson 2006
D269	D270	D161	Metals B/C binding	Godson, Schoenich et al. 2000; Keck, Roche et al. 2000; Kato, Ito et al. 2003; Lee and Richardson 2005; Rodina and Godson 2006
D309	D310	D207	Metal A binding	Keck, Roche et al. 2000; Kato, Ito et al. 2003; Lee and Richardson 2005; Rodina and Godson 2006
D311	D312	D209	Metal B binding	Keck, Roche et al. 2000; Kato, Ito et al. 2003; Lee and Richardson 2005; Rodina and Godson 2006
D345	D343	D237	Metals B/C binding	Godson, Schoenich et al. 2000; Keck, Roche et al. 2000; Kato, Ito et al. 2003; Lee and Richardson 2005
D347	D345	N239	Metal C binding	Godson, Schoenich et al. 2000; Keck, Roche et al. 2000; Kato, Ito et al. 2003; Lee and Richardson 2005
R146	R146	R84		Keck, Roche et al. 2000; Rodina and Godson 2006
R221	R222	R124	Basic Ridge Triphosphate binding	Keck, Roche et al. 2000; Rodina and Godson 2006
K229	K230	K128		Keck, Roche et al. 2000; Lee and Richardson 2001; Rodina and Godson 2006
N232	N233	K122		Lee and Richardson 2001; Lee and Richardson 2005; *This study*
G266	G267	G158	Nucleotide binding Incoming ribose	Rodina and Godson 2006; *This study*
G286	G287	G181		*This study*

Chapter 3 - An introduction to inhibitors of DnaG and investigation of the mechanism of inhibition of DnaG by drug-like inhibitors

INTRODUCTION

Mechanisms of primer synthesis regulation

Control of primer synthesis by DnaG is a central regulatory point in bacterial DNA replication (**Chapter 1**). Overall, three general strategies to modulating DnaG function have been identified. One relies on the selective recruitment of DnaG to a specific site of action. For example, *Escherichia coli* DnaG interacts with both the replicative helicase, DnaB (Marians 1992; Wu, Zechner et al. 1992; Tougu, Peng et al. 1994; Lu, Ratnakar et al. 1996), and the single-stranded DNA binding protein, SSB (Sun and Godson 1993; Sun, Tormo et al. 1994), interactions that allow DnaG to interface with distinct replisomal processes (Tougu and Marians 1996; Heller and Marians 2006). A second strategy involves the localized control of primer synthesis. Once bound to DnaB, DnaG can self-associate as a means to both determine the starting position for primer synthesis and control primer length (Tougu and Marians 1996; Bhattacharyya and Griep 2000; Corn, Pease et al. 2005; Chintakayala, Machon et al. 2009; Hamdan and van Oijen 2010; van Oijen and Loparo 2010). The third approach is the use nucleotidyl inhibitors, such as the stringent response effectors (p)ppGpp and 2´-dNTPs, which directly impede primer formation (Swart and Griep 1995; Wang, Sanders et al. 2007; Maciag, Kochanowska et al. 2010). At present, how these various protein/protein and protein/ligand interactions influence the catalytic properties of DnaG is not understood.

Studies on the mechanisms of action of DnaG inhibitors

To address the mechanisms of action underlying the latter strategy, we investigated the mechanisms of primer synthesis inhibition by four known classes of primase inhibitors (**Fig. 3.1**) (Robinson, Causer et al. 2012). The mechanism of primer synthesis inhibition by the stringent responses alarmones, (p)ppGpp is elucidated in **Chapter 4**, and the mechanism of initiation-specific inhibition of DnaG by 2´-dNTPs *in vitro* is investigated in **Chapter 5**.

In this chapter, a study of the mechanisms of primer synthesis inhibition by drug-like primase inhibitors is presented. In the course of this work, we developed a highly sensitive coupled reporter experiment, the Helicase-Stimulated, Polymerase-Coupled (HSPC) primer synthesis assay, that is well suited to the study of primase inhibitors (**Appendix I**). An important feature of this assay is the ability to account for off-target effects on alternative polymerases. We also determined that molecules in one drug-like class, typified by Sch-642305, are unlikely to actually be primase inhibitors, while that a second group of small-molecule agents, the benzopyrimidines, inhibit DnaG by nonspecific, intercalative interactions with DNA.

RESULTS

Failure to observe primer synthesis inhibition by Sch-642305
Drug-like primase inhibitors have received relatively little attention (Robinson, Causer et al. 2012). In 2003, a group at Schering-Plough (now part of Merck, Inc.) discovered a natural product that inhibited DnaG in a screen for bactericidal compounds based on extracts from *Penicillium verrucosum* fermentation cultures (Chu, Mierzwa et al. 2003). The compound, originally named Sch-642305 (compound **2**, **Fig. 3.1b**), is a cyclic, 10-membered macrolide with an IC_{50} of 70 μM. Several methods for the total synthesis of this compound have been reported (Mehta and Shinde 2005; Snider and Zhou 2006; Nicoletti, Lopez-Gresa et al. 2007; Wilson and Trauner 2007). We probed of the mechanism of action of Sch-642305 initially by pursuing a cocrystal structure with this compound and the *E. coli* RPD. As these efforts were unsuccessful, we turned to a biochemical characterization of the primer synthesis inhibition attributed to the agent. Surprisingly, our first control, a titration of Sch-642305 into reactions of the HSPC primer synthesis assay, revealed no apparent inhibition (**Fig. 3.2a**). The inclusion or omission of the helicase did not influence this result (not shown). As the coupled assay was run as an endpoint method, we theorized that the relatively weak reported inhibition by Sch-642305 might be masked by the long incubation times we employed (1 h). However, a dose-response analysis at two earlier time points (10 and 20 min), revealed no evidence of inhibition (**Fig. 3.2b**). No inhibition was apparent with the polymerase alone (not shown). Although we searched for alternative explanations for inhibition by Sch-642305, such as interference with biotin-streptavidin binding (an interaction required to observe signal for the assay originally used to discover the compound), none of these experiments revealed a repeatable source of inhibition. As a consequence of this finding, we ceased further investigations into the mode of action of Sch-642305.

Benzopyrimidines inhibit DnaG and Klenow by nonspecifically interacting with template DNA
Since our efforts to understand Sch-642305 had reached an impasse, we turned our attention to a second class of drug-like primase inhibitors, collectively called benzopyrimidines. These inhibitors were identified at Achillion Biopharmaceuticals by computational docking experiments with commercially available drug-like molecular scaffolds (Agarwal, Louise-May et al. 2007). Compounds **3** and **4** exemplify the hits obtained from this screen. The same researchers also performed a limited SAR study that indicated IC_{50} values in the 5 – 100 μM range, providing biochemical validation of the computational results.

We obtained samples (**Methods**) of the most potent compounds, **3** and **4** (**Fig. 3.1c,d**), as well as several other analogues (not shown). A dose-response analysis of **3** and **4** using the HSPC primer synthesis assay indeed demonstrated robust and concentration dependent inhibition of DnaG (**Fig. 3.3a**, black curves), with IC_{50}s of 46 μM for **3** and 43 μM for **4**. Using a more traditional primer synthesis assay, we validated these findings, showing that at least one of the benzopyrimidines, **3**, inhibited DnaG (**Fig. 3.3c**) (Swart and Griep 1995), **Methods**). However, upon testing for inhibition of the polymerase by our HSPC approach (**Methods**), we unexpectedly discovered that both compounds also inhibited the DNA polymerase alone, with only slightly reduced potency (**Fig. 3.3a**, gray curves).

Because of this cross-reactivity, we set out to determine if benzopyrimidines might be specific inhibitors of both Klenow and DnaG, or if some alternative, but nonspecific mechanism of action could account for the inhibition seen with both enzymes. We began by investigating the response of the combined inhibition of DnaG and Klenow by **3** to substrate DNA. Titration of m13 ssDNA with increasing concentrations of **3** using the coupled assay results in curves that appear to be characteristic of a mixed mode inhibitor (**Fig. 3.4a**). As a means of reducing the impact of the apparent substrate inhibition on the data analysis, we replotted the low concentration data points (where substrate inhibition would be the least influential) using Eadie-Hofstee analyses (Segel 1993). By extrapolating these data to the y-axis, we were able to estimate the V_{max} in the absence of substrate inhibition, revealing a pattern more consistent with a predominantly competitive inhibitor (**Fig. 3.4b**).

We also compared the IC_{50} of **3** obtained on a standard m13 ssDNA template, with that seen when using the same short, single-stranded substrate used in traditional primer synthesis assays (**Fig. 3.3c** and **Methods**). To obtain these data, we used a previously published, uncoupled fluorometric primer synthesis assay (**Methods**, (Koepsell, Hanson et al. 2005)). To our surprise, a simple dose-response curve of **3** using the short template returns an IC_{50} five times higher than an identical titration performed using m13 as a template (**Fig. 3.4c**). These titrations were performed at a template concentration that yielded approximately the same relative primer synthesis rate (**Methods**).

This trend in the template length-dependence of IC_{50} for **3** is not compatible with either a competitive or mixed mode of inhibition, as the m13 ssDNA template is approximately 260 times longer than the ssDNA oligonucleotide template. Consequently, the concentration of template in the titration of **3** using m13 ssDNA was higher on a per-base scale (36 μM) than with the titration using the short oligonucleotide ssDNA (17 μM, **Methods**). Because the IC_{50} of a competitive inhibitor depends inversely on the concentration of the substrate with which it competes (Copeland 2005), the IC_{50} with m13 ssDNA should be higher, not lower, as we observe. To reconcile these two results, we theorized that the template length dependence of inhibition derived from a nonspecific mechanism of inhibition involving a direct association with template, not primase.

Given the aromatic nature of many of the benzopyrimidines, we surmised that the compounds might intercalate into DNA, and that the large, complicated structure of m13 ssDNA would provide more opportunities for the compounds to interact with the template DNA in comparison to the short oligonucleotide template. To test this theory, we performed a canonical induced supercoiling relaxation assay with **3** and **4** (**Fig. 3.4d**). In this assay, **3** and **4** were mixed with relaxed plasmid DNA at concentrations slightly below their respective IC_{50}s with m13 ssDNA as a template. In this assay, a DNA intercalator induces positive supercoiling of the substrate plasmid. Treatment with topo I returns the DNA to a relaxed supercoiling state. After purification of the plasmid away from the compound, the plasmid compensates by adopting a negatively supercoiled state. As can be seen in **Fig. 3.4e**, addition of both **3** and **4** resulted in significantly faster migrating species, the characteristic signature of additional supercoiling in a plasmid, demonstrating that both compounds are DNA intercalating agents. Of the two, compound **3** was particularly robust at altering DNA topology.

45

DISCUSSION

Because of its central role in DNA replication DnaG is an attractive prospective target for therapeutic intervention; however, relatively little has been done to establish its feasibility as a drug target. We investigated the mechanism of primer synthesis inhibition by two previously reported DnaG inhibitors, and show that neither is a *bona fide* primase inhibitor. In the case of **2**, it is apparent that the compound is in fact a false positive, but poor disclosure of the exact methods and conditions employed in the report detailing its discovery prevented us from determining the true source of its apparent inhibition in the Shering-Plough assay. For the second class of inhibitors, benzopyrimidines, we found that they inhibit DNA synthesis by Klenow nearly as well as they do DnaG, but that their mode of inhibition is due to DNA intercalation rather than through specific protein interactions. Together, these studies highlight some the challenges of following up on small-molecule identification efforts in the literature, as a range of poorly-controlled factors can incorrectly identify non-specific inhibitory agents.

MATERIALS AND METHODS

Sources for inhibitors, proteins, and template DNA
2 was synthesized by researchers in the Trauner lab according to their published procedure (Wilson and Trauner 2007). Material was dissolved in DMSO to 100 mM, and stored at -20 ℃ until just prior to use. Solid powder of benzopyrimidines **3** and **4** were purchased from ChemDiv Inc., dissolved in absolute ethanol (Sigma-Aldrich) to 100 mM, and stored at -20 ℃ until just prior to use. DnaG and DnaB were expressed and purified as described (Corn, Pelton et al. 2008). M13 ssDNA was purchased from Affymetrix as a 1 μg/μL stock solution in TE, aliquoted and stored at -20 ℃. Klenow was purchased from Epicentre Biotechnologies as a 50 U/μL stock solution, and was stored at -20 ℃ in the buffer provided by Epicentre. Oligonucleotide templates were purchased from Integrated DNA Technologies, resuspended in TE to 1 mM, aliquoted, and stored at -20 ℃.

Polymerase Coupled Primer Synthesis Assay
All reactions were carried out in a volume of 50 μL. Each complete reaction contained buffer (referred to as Reaction Buffer) consisting of 100 mM potassium glutamate, 20 mM HEPES (pH 7.5), 6% sucrose, 0.2 mg/mL BSA, 20 mM MgOAc and 1 mM DTT. Unless otherwise noted, all nucleotides were present at 0.1 mM, and m13 ssDNA template at 1.4 nM, Klenow was present at 0.1 U/rxn, DnaG at 100 nM and DnaB at 100 nM. Inhibitors were included in the incomplete reaction mixture for template titrations being conducted in the presence of inhibitors. Complete reactions were incubated in a sealed half-area 96-well plate (Corning) for 1 h at 37 ℃, and then stopped by addition of 50 μL of a 1:100 dilution of PicoGreen stock solution (Invitrogen) in 20 mM Tris, 50 mM EDTA, pH 7.5, giving a final stain dilution of 1:200, as per the manufacturer's instructions. Stopped reactions were incubated in the dark for 5 min, spun at 3000 x g for 2 min, and fluorescence intensity measured in a PerkinElmer Victor3V multilabel plate reader. Raw fluorescence was background corrected with a no-NTP control to give Fluorescence Intensity (FI). The wild-type V_{max} was calculated by fitting resultant curves to the Eqn. 2.1 by nonlinear regression. All data within

a given experiment were then normalized to the wild-type V_{max}, and are reported here as Relative Primer Synthesis (RPS). For template titrations, raw fluorescence was background corrected with a no protein control at each template concentration to give FI. The data were then fit to Eqn. 3.1 (Segel 1993) by nonlinear regression, and normalized to the 0 μM titration V_{max}.

Equation 3.1 $$\text{FI} = \frac{V_{max,app} \times S}{K_{M,app} + S \times (1 + \frac{S}{K_{i,app}})} \quad,$$

For inhibitor titrations, dilutions (5 μL) were prepared in TE (**1**) or DMSO (all others), and mixed with 20 μL of an incomplete reaction mixture containing Klenow, DnaG, DnaB, ATP, UTP and template, and allowed to incubate at room temperature for 10 minutes prior to addition of d/NTP solution (25 μL) to start the reaction. Reactions were incubated at 37 °C for 1 h, and stopped, stained and measured using the same protocol as the standard reactions. The data were normalized to the average max FI, and fit to the Eqn. 3.2 (Copeland 2005) by nonlinear regression.

Equation 3.2 $$\text{RPS} = \frac{1}{1 + \left(\frac{[I]}{IC_{50}}\right)^h} \quad,$$

in which [I] is the inhibitor concentration, IC_{50} is the concentration of inhibitor that results in 50% inhibition, and h is the hill coefficient. For all *de novo* synthesis experiments, reactions were carried out in triplicate, unless otherwise noted, and are reported here as an average of the three values, along with the Standard Deviation (SD).

General primer synthesis assay by radiolabel incorporation
The general primer synthesis assay has been described (Swart and Griep 1995), and was performed according to the protocol described elsewhere (Corn, Pease et al. 2005). The template was phosphorylated on the 3'-end.

Uncoupled fluorometric de novo primer synthesis assays
All reactions were carried out in a volume of 25 μL. Each complete reaction contained Reaction Buffer lacking sucrose. Under standard reaction conditions, all nucleotides were present at 0.1 mM, template at 5 nM for the 7,250 nt m13 ssDNA template, or 600 nM for the 28 nt single-stranded oligonucleotide template, which correspond to a concentration of each template that yields activity of ~25% of V_{max} for DnaG at 500 nM and DnaB at 500 nM. **3** was titrated from 3.5 μM to 0.5 mM. Inhibitor dilutions (2.5 μL) were prepared in DMSO, and mixed with 20 μL of an incomplete reaction mixture containing Klenow, DnaG, DnaB, ATP, UTP and template, and allowed to incubate at room temperature for 10 minutes prior to addition of NTP solution (2.5 μL) to start the reaction. Complete reactions were incubated in a sealed half-area 96-well plate (Corning) for 1 h at 37 °C, and then stopped by addition of 25 μL of a 1:100 dilution of PicoGreen stock solution (Invitrogen) in 20 mM Tris, 50 mM EDTA, pH 7.5, as per the manufacturer's instructions. After stopping and staining, reactions were treated as above. The data were normalized to the average max FI to give RPS, and fit to Eqn. 3.1 by nonlinear regression.

Writhe footprinting assay for intercalation

Reactions were carried out in a total volume of 25 µL. Plasmid (300 ng), buffer, and compound were incubated for 10 min at RT. Topo I (1 µL) was added to each reaction, and the reactions incubated at 37 ˚C for 25 min, then stopped by the addition of SDS to 1%, EDTA to 20 mM, proteinase K to 1 mg/mL and incubation at 55 ˚C for 10 min. 5 µL of loading buffer was added to each reactions, 5 µL of sample loaded onto a 1% 10 x 30 cm agarose gel, and the reaction products separated by electrophoresis at 50 V for 24 h in 1xTAE. The gel was then stained by addition of ethidium bromide staining solution, and incubation with shaking for 10 min in the dark. Finally, the gel was destained for 20 min in running buffer, and visualized by UV. Relaxed plasmid and topo I solution were provided by Elsa M. Tretter.

a nucleotidyl

Cordycepin (**1**)

b 10-membered macrolides

(+)-Sch 642305 (**2**)

c Benzopyrimidines

3

4
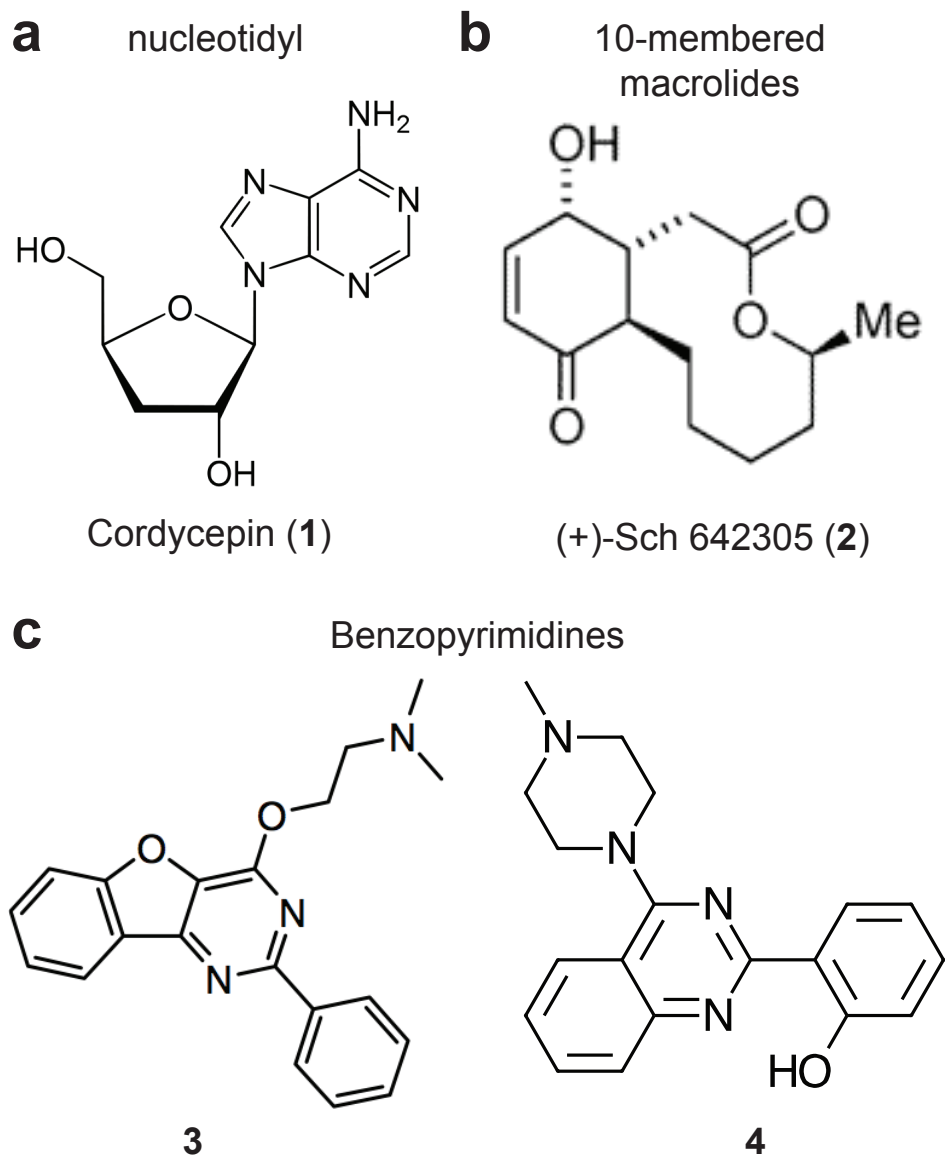
Figure 3.1: The Classes of DnaG-type Primase Inhibitors.
Chemical structures of representative members of the DnaG-type primase inhibitors.
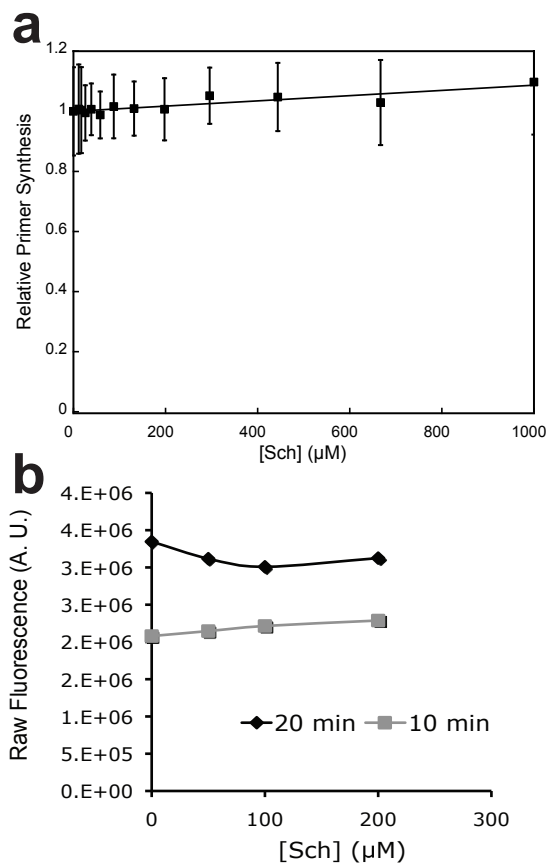
Figure 3.2: Failure to observe Sch-dependent inhibition of primer synthesis by E. coli DnaG.

a) Dose-response of DnaG-DnaB to Sch in the HSPC primer synthesis assay. Sch was titrated from 11 μM to 1 mM into reactions containing 100 nM DnaG and 100 nM DnaB. Relative Primer Synthesis, the primer synthesis extent as a fraction of the synthesis extent for the 0 μM Sch data point (Relative Primer Synthesis). b) Dose-response of DnaG-DnaB to Sch at earlier time points in the HSPC primer synthesis assay. Sch was titration from 50 μM to 200 μM into reactions as in (a), but primer synthesis was measured at 10 (gray squares) and 20 min (black diamonds) as opposed to the standard 1 h. A. U., arbitrary units.

Figure 3.3: Equivalent inhibition of DnaG and Klenow by benzopyrimidines.
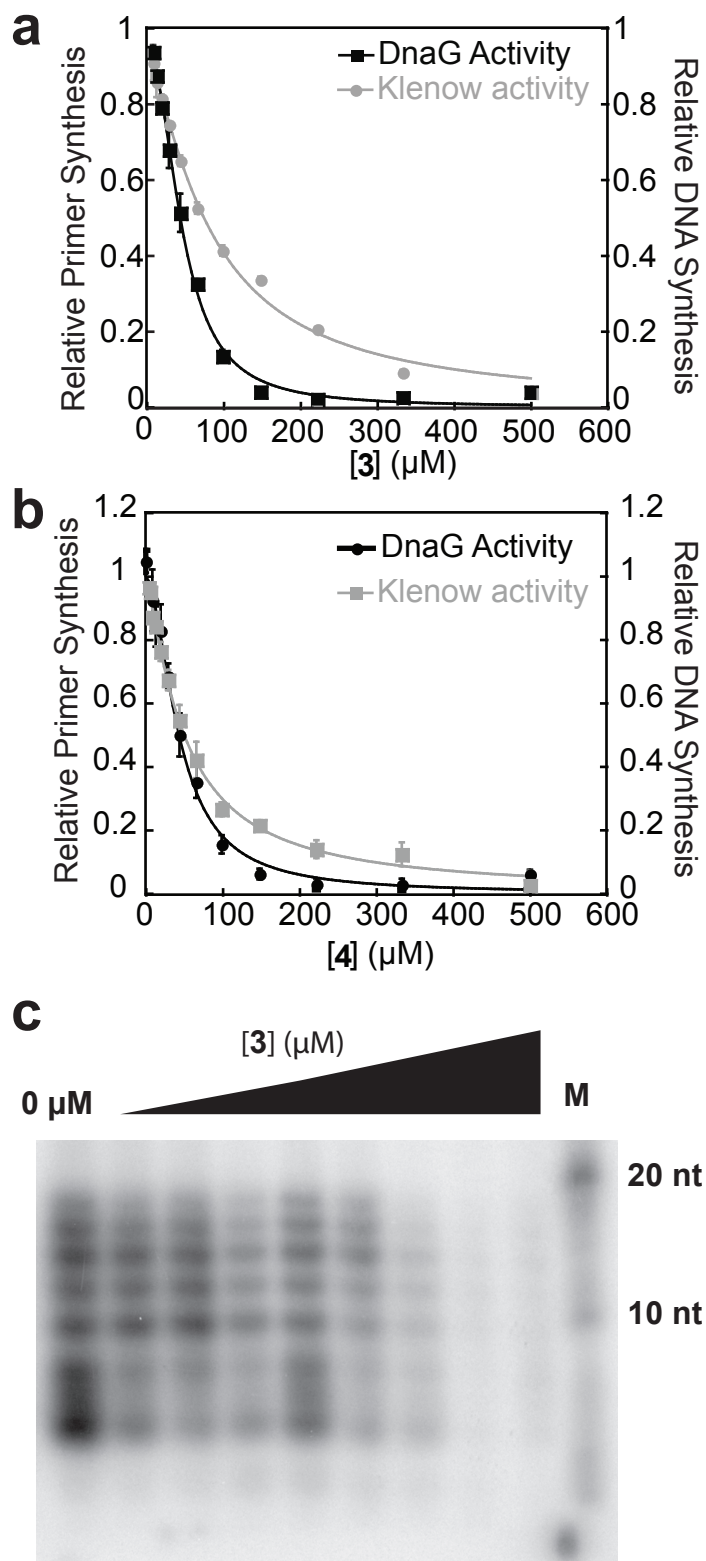
a) Dose response of DnaG-DnaB and Klenow to **3**. **3** was titrated from 5 µM to 0.5 mM into reactions containing 100 nM DnaG and 100 nM DnaB (black squares) or 10 nM of an m13 complementary DNA primer and no DnaG (gray circles). Curves are fit to an IC$_{50}$ model (**Methods**). Synthesis is reported as a fraction of the synthesis extent for the 0 µM **3** data point. Error bars represent ±SD (**Methods**). b) Dose response of DnaG-DnaB and Klenow to **4**. **4** was titrated from 5 µM to 0.5 mM into reactions containing 100 nM DnaG and 100 nM DnaB (black circles) or 10 nM of an m13 complementary DNA primer and no DnaG (gray squares). Curves are fit to an IC$_{50}$ model (**Methods**). Synthesis is reported as a fraction of the synthesis extent for the 0 µM **4** data point. Error bars represent ±SD (**Methods**). c) Dose-response of DnaG to **3** in a radiolabel incorporation primer synthesis assay. **3** was titrated from 30 µM to 300 µM into reactions containing 10 µM DnaG, no DnaB and no polymerase. A 3´-phosphorylated template was used, but other reaction conditions matched the published procedure (**Methods**).
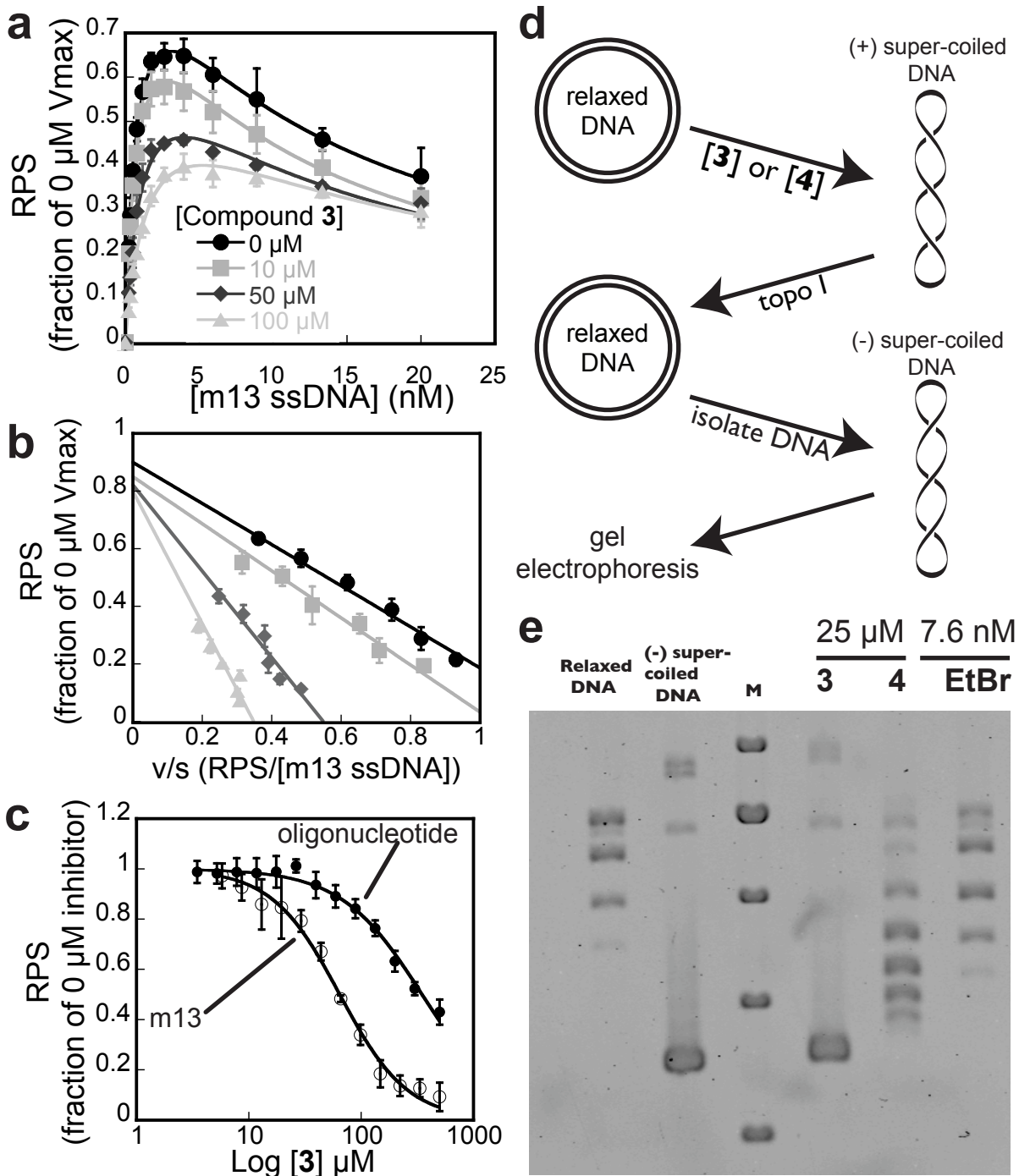
Figure 3.4: Benzopyrimidines inhibit DnaG and Klenow by interacting directly with template DNA.

a) m13 ssDNA dependent synthesis in the presence of increasing concentrations of **3**. m13 ssDNA was titrated from 0.2 to 20 nM into reactions containing 0 (black circles), 10 (gray squares), 50 (dark gray diamonds), or 100 μM (light gray triangles) **3**. Curves were fit to a substrate inhibition model (**Eqn. 3.2**, **Methods**). Error bars represent ±SD (**Methods**). RPS, relative primer synthesis. Continued on next page.

b) Eadie-Hofstee plot (v vs. v/s) of the 0-5 nM m13 ssDNA data points shown in (a). Data were fit by linear regression, and the lines extended to the y-axis to provide an estimate of the Vmax of each titration without substrate inhibition. c) The impact of template length on the IC_{50} of **3**. Titrations of **3** were performed in an uncoupled assay using either 5 nM m13 ssDNA (open circles), or 600 nM of a short ssDNA oligonucleotide template (closed circles). All reactions contained 500 nM DnaG and 500 nM DnaB, and 100 μM rNTPs. All other assay conditions from the coupled assay were maintained. Curves were fit to an IC_{50} model (**Methods**). d) Schematic of a writhe assay used to interrogate supercoiling caused by intercalation of small molecule into DNA. The potential intercalator is removed upon isolation of the DNA, resulting in compensatory supercoiling. e) The presence of **3** and **4** induces supercoiling of relaxed DNA. Gel image showing the results from including **3** and **4** in the assay depicted in (d). Several controls were included, relaxed DNA and negatively supercoiled DNA, as well as a reaction with a small amount of ethidium bromide (EtBr), a known DNA intercalator. The appearance of additional bands with faster mobility than can be seen in the relaxed DNA lane is indicative of additional supercoiling, and therefore intercalation.

Chapter 4 – Elucidation of the mechanism of action of primer synthesis inhibition by the stringent response alarmones, (p)ppGpp.

INTRODUCTION

Nutrient deprivation in bacteria evokes a cellular reaction, termed the Stringent Response, in which DNA replication, transcription and translation become inhibited through the action of two small molecule alarmones, ppGpp and pppGpp (**Fig. 4.1a**). Both of these nucleotides, collectively known as (p)ppGpp, bind to and alter the activity of critical enzymes involved in all three processes (Srivatsan and Wang 2008), leading to slower growth, reduced consumption of metabolic starting materials, and increased expression of various metabolic enzymes. Although the molecular and atomic details of how (p)ppGpp affect targets in transcription and translation have been studied for some time (Potrykus and Cashel 2008), only recently have researchers begun to unlock these details for enzymes involved in DNA replication: interestingly, recent findings indicate that (p)ppGpp-dependent inhibition of primer synthesis by DnaG appears to play a major role in replication inhibition. (Wang, Sanders et al. 2007).

To investigate the mechanism of primer synthesis inhibition by (p)ppGpp, we solved the crystals structures of the *S. aureus* DnaG catalytic core bound individually to each alarmone. Both structures reveal a unique mode of interaction with the DnaG catalytic core that overlaps with both the site of nucleotide binding and with the nascent primer chain. These insights, together with biochemical studies, indicate that stringent response alarmones block DnaG function by complex, mixed mode of inhibition for multiple substrates.

RESULTS

The crystal structures of the S. aureus DnaG RPD bound to (p)ppGpp
To first define how (p)ppGpp might bind to DnaG, we initially attempted to model the two modified nucleotides into our structure of the isolated *S. aureus* RPD (**Chapter 2**). These efforts, however, indicated that neither G5P nor G4P could bind to the *Sa*DnaG RPD in a manner similar to that of $Mn^{2+} \cdot NTP$ without introducing steric clashes between the 3´-phosphates of the alarmones and the protein.

To instead establish the location and orientation of G5P and G4P binding to the DnaG RPD directly, we collected diffraction data and determined structures for crystals soaked in solutions of 2.5 mM of either G5P or G4P, along with 10 mM $MnCl_2$, and 1 mM $MgCl_2$. Clear difference density was evident for both molecules (**Fig. 4.1b/c**), which revealed that the two guanosine analogs associate with the *Sa*DnaG RPD in a similar manner. Interestingly, comparison with the nucleotide-bound *Sa*DnaG complexes shows that the site of binding for G5P and G4P overlaps with the $Mn^{2+} \cdot NTP$ binding locus, but only partly (**Fig. 4.1d/e**). In particular, the 5´-phosphates of both inhibitors occupy the same triphosphate-binding region, using a metal-coordinated configuration identical to that observed for the other

NTPs, with the exception that metal B fails to bind with G4P, causing a shift in the position of its 5´-phosphates compared to that of either G5P or the four standard nucleotides. By contrast, the 3´-phosphates of both alarmones ligand metal A, a connection that requires reorientation of the base and ribose into a configuration distinct from that seen with NTPs. This location and rearrangement suggests that G5P and G4P interfere with primer synthesis by co-occupying both the 3´-OH and 5´-phosphate binding loci, thereby occluding the entire DnaG active site.

(p)ppGpp compete with two substrates for binding to DnaG
Having established a binding modality for (p)ppGpp, we set out to test the mechanism of action the structures implied: that the two molecules impede primase activity not only through blocking binding of an incoming NTP, but by also interfering with the binding of either an initiating 5´-NTP, or the extensible 3´ end of an RNA•DNA heteroduplex. To this end, we examined the concentration-dependent effects of both G4P and G5P on various activities of *E. coli* DnaG. As a control, we first assessed the ability of our alarmone preparations to inhibit *de novo* primer synthesis by full-length DnaG, using an established helicase-stimulated assay (Koepsell, Hanson et al. 2005). Consistent with previously published findings (Maciag, Kochanowska et al. 2010), we found that both nucleotides were weak inhibitors, having IC_{50}s in the low mM range (**Fig 4.2a**).

We next asked if (p)ppGpp could interfere with heteroduplex binding, as might be expected if the 3´-phosphates of the inhibitors were to clash with an acceptor 3´-OH associated with metal A. Using a change in fluorescence anisotropy of a fluorescein labeled-DNA•RNA heteroduplex as a readout for binding to a primer-template substrate (**Methods**) by the *Ec*DnaG RPD, we observed that increasing concentrations of either G5P or G4P progressively impeded the protein from associating with the oligo (**Fig 4.2b**).

To verify that (p)ppGpp compete for binding to the observed NTP binding site, we first examined the ability of G5P and G4P to interfere with binding of the *Ec*DnaG RPD to a fluorescently labeled nucleotide, adenosine 5´-O-(3-thiotriphosphate)-BODIPY fluorescein thioester, **Appendix II**). The results from these experiments were inconclusive, so we instead assessed the ability of (p)ppGpp to compete with GTP in a total primer synthesis assay using full-length DnaG. Since GTP cannot provide an initiating 3´-OH due to the start site preferences of *E. coli* DnaG (5´-C<u>T</u>G and 5´-C<u>A</u>G, where the T or A position marks the 5´ terminal site of the primer), it acts as a specific probe for the accessibility of the incoming nucleotide binding-site, which both alarmones block partially in our structures (Khopde, Biswas et al. 2002). Both alarmones qualitatively exhibited mixed-type inhibitor behavior in blocking the ability of DnaG to synthesize primers in a GTP-concentration dependent manner (**Fig 4.2c,d**), with G5P causing a larger increase in the apparent-K_m compared to G4P (**Fig 4.2e,f**).

DISCUSSION

The ability of bacterial organisms to respond to stressful conditions has a direct bearing on the potential pathogical behavior (Levin and Rozen 2006). To date, the stringent response has been implicated many times in the pathogenicity of several clinically relevant organisms, including *E. coli* and *P. aeruginosa* (Traxler, Summers et al. 2008; Vogt, Green et al. 2011). The stringent response inhibits DNA replication in bacteria by two alarmones, pppGpp (G5P) and ppGpp (G4P) (Jain, Kumar et al. 2006). Long known to block translation and transcription (Potrykus and Cashel 2008), both nucleotides have been shown recently to additionally impede primer synthesis by inhibiting DnaG directly (Wang, Sanders et al. 2007; Maciag, Kochanowska et al. 2010). This study reinforces this finding, while further showing that the two small-molecule alarmones associate with the DnaG catalytic core in a manner that would occlude access to the active site of DnaG (**Chapter 2**), thereby providing a mechanism of action for these agents.

MATERIALS AND METHODS

Cloning, expression, purification
See **Methods**, **Chapter 2**.

Crystallization of the SaDnaG RPD
See **Methods**, **Chapter 2** for the crystallization protocol. For soaks, crystals were first exchanged into well solution containing 10 mM $MnCl_2$, 1 mM $MgCl_2$ and 2.5 mM of either G4P or G5P and incubated overnight prior to cryoprotection and harvesting. Crystals were cryoprotected as described in **Methods, Chapter 2**. All crystals were flash-frozen and stored in liquid nitrogen prior to data collection at 100 K.

Data collection and refinement
See **Methods**, **Chapter 2**.

De novo primer synthesis assays
GTP titrations were performed using the protocol described in **Methods**, **Chapter 2**, except inhibitors were included in the incomplete reaction mixture, and incubated for 10 min at room temperature prior to addition of GTP solution to start the reaction. For inhibitor titrations, G5P and G4P were prepared as described (Mechold, Murphy et al. 2002), and stored in desiccated form until just prior to use. Dry pellets were resuspended in 1x Tris-EDTA (TE), and the concentration measured by absorbance at 258 nm using the extinction coefficient of 13,700 $M^{-1}cm^{-1}$. GDP was prepared by dissolving GDP powder (SigmaAldrich) in TE to 100 mM, and the pH adjusted to 7.5. Inhibitor dilutions (7.5 μL) were prepared in TE, and mixed with an equal volume of an incomplete reaction mixture containing 2.6x Reaction Buffer, DnaG, DnaB, ATP, UTP and template, and allowed to incubate at room temperature for 10 minutes prior to addition of GTP solution (5 μL at 400 μM) to start the reaction. Reactions were incubated at 37 °C for 1 h, and stopped, stained and measured using the same protocol as the standard GTP titrations. G5P was titrated from 8.4 mM to 8.2 μM, G4P was titrated from 6.9 mM to 6.8 μM, and GDP from 10 mM to 78 μM. The data were normalized to the average maximum fluorescence intensity to give RPS, and fit to Eqn. 3.1 by nonlinear regression.

Heteroduplex competition assays
E. coli DnaG RPD (500 µM) was incubated with 100 nM of a 5′-fluorescein labeled RNA·DNA heteroduplex with a 5′-overhang (DNA strand (IDT) sequence: 5′-AAAACTGCCGCCGC-3′, RNA strand (Dharmacon) sequence: 5′ -GCGGCGGCA-3′) and 2x Reaction Buffer conditions at RT for 10 min. Inhibitor at 2x final concentration (10 mM) was serially diluted 1:1 into TE; dilutions were then mixed 1:1 into the 2x RPD and labeled heteroduplex mixture. All inhibitors were titrated from 5 mM to 10 µM. Reactions were incubated at room-temperature for 30 min, and Fluorescent Polarization (FP) was measured with a PerkinElmer Victor3V multi-label plate reader. FP was converted to Fluorescence Anisotropy (FA) prior to background correction with a no-protein control to obtain ΔFA values. The data were fit by nonlinear regression to Eqn. 4.1 (Copeland 2005),

$$\textit{Equation 4.1} \quad = \quad \Delta FA_{min} + \frac{(\Delta FA_{max} - \Delta FA_{min})}{1 + (\frac{[I]}{IC_{50}})^h} \quad ,$$

in which ΔFA_{min} and ΔFA_{max} refer to the signal minimum and signal maximum, respectively, [I] is the concentration of inhibitor, IC_{50} is the concentration of inhibitor that results in 50% displacement, and h is the Hill coefficient.

Visualization and data analysis methods
Structures were visualized and figures generated using PyMOL (Schrodinger 2010). Data for activity and binding assays were processed in Microsoft Excel. Plotting and fitting were carried out in Kaleidagraph.
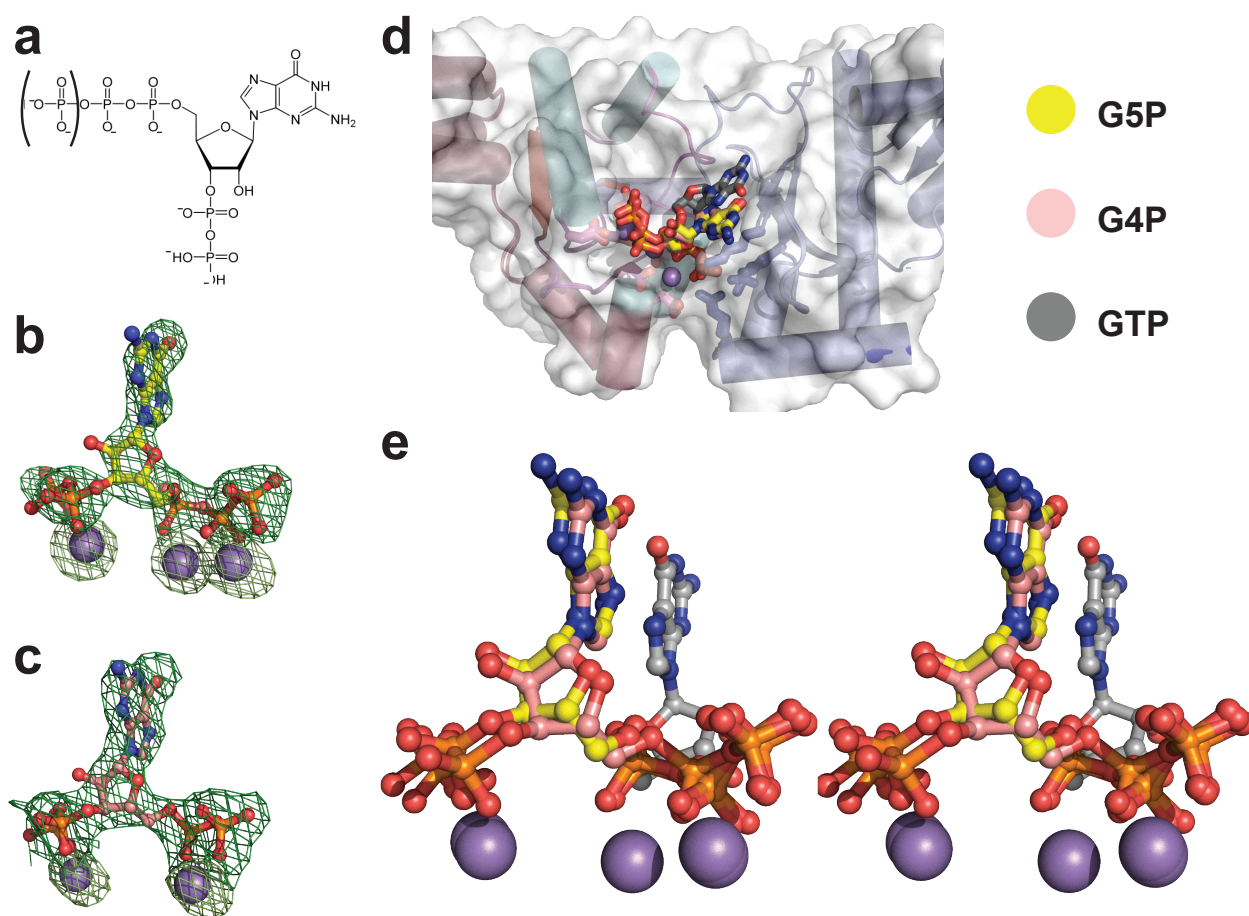
Figure 4.1: Structure of the *S. aureus* DnaG RPD bound to pppGpp and ppGpp

a) The chemical structure of (p)ppGpp. The phosphate enclosed in parentheses is present with pppGpp (G5P), but not ppGpp (G4P). b,c) Difference density present in the *Sa*DnaG RPD active site generated from crystals soaked with b) G5P or c) G4P. Difference density (green mesh, mFo-DFc) is contoured at 3σ after refinement of the protein reached convergence. Anomalous difference density generated from data collected from the same crystal at the Mn-K edge is shown as light green mesh, and is contoured at 8σ. G5P is shown as a ball-and-stick representation with carbons colored yellow. G4P is shown as a ball-and-stick representation with carbons colored salmon. Mn^{2+} ions are shown as gray spheres. d) Alignment of the G5P and G4P bound structures with each other and with the GTP bound structure. GTP is shown as a ball-and-stick model with carbons colored gray. G5P and G4P are colored as in (b) and (c), respectively. Mn^{2+} ions are shown as gray spheres. e) Stereo image of the alignment shown in (d) with protein models omitted.
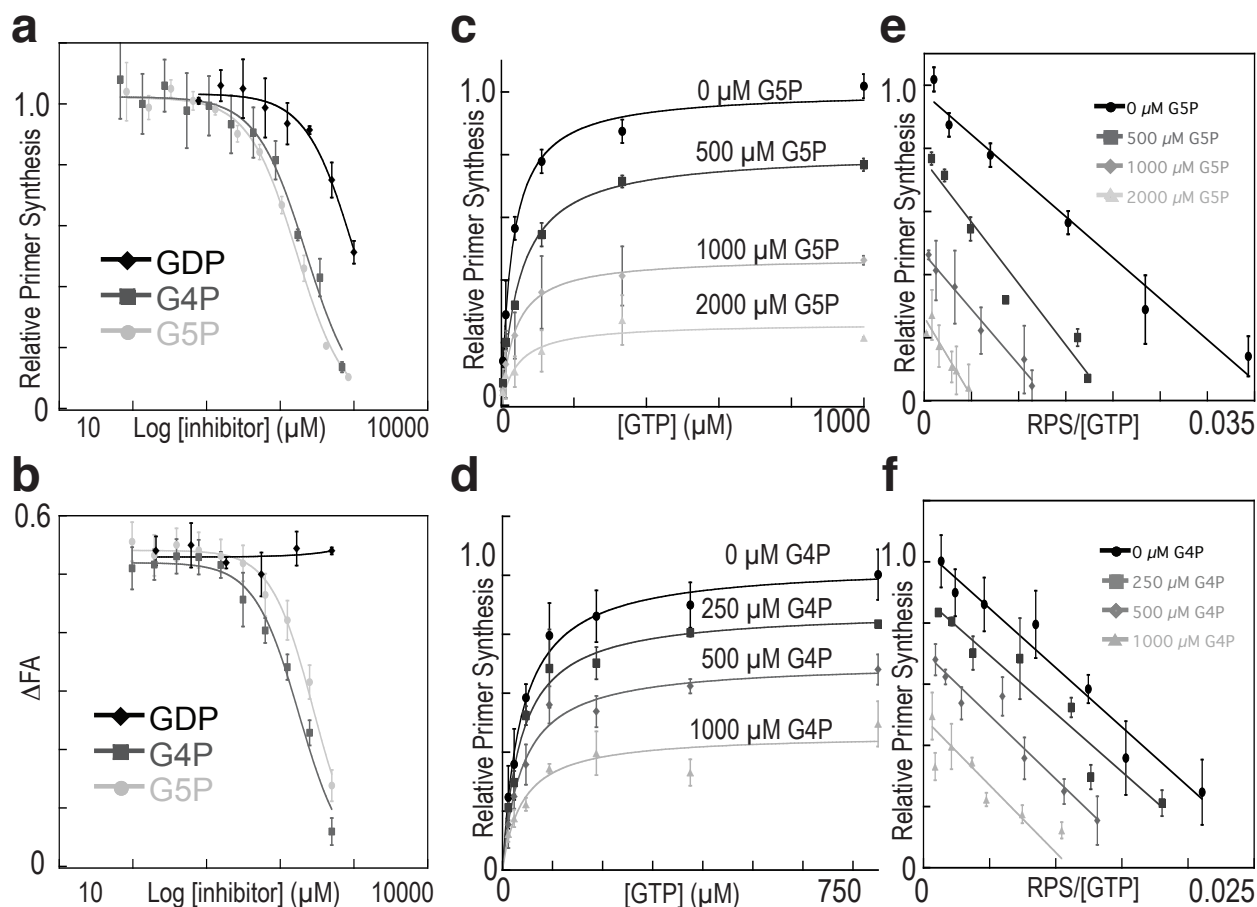
Figure 4.2: The biochemical response of the E. coli DnaG to G5P and G4P

a) Concentration dependent inhibition of *Ec*DnaG by (p)ppGpp in a fluorometric, helicase-stimulated *de novo* primer synthesis assay. Titration of G5P (dark-gray circles), G4P (black squares), and GDP (light-gray diamonds) into reactions with full-length *E. coli* DnaG. Curves are fit to a standard IC50 model (**Methods**). Error bars represent ±SE. b) Competition for binding of the *Ec*DnaG RPD to a fluorescently labeled heteroduplex by G5P (dark-gray circles), G4P (black squares), and GDP (light-gray diamonds). Inhibitors were titrated into reactions containing the fluorescently labeled heteroduplex, and 250 μM RPD. Binding extent was determined by measuring the change in the Fluorescence Anisotropy (DFA) of the labeled heteroduplex. Curves are fit to a standard IC50 model (**Methods**). Error bars represent ±SE. c,d) GTP-dependent inhibition of primer synthesis by G5P and G4P. GTP-dependent, helicase-stimulated primer synthesis activity was measured (**Methods**) with increasing concentrations of c) G5P or d) G4P. Points represent averages of duplicate reactions. Primer synthesis extent is reported as a fraction of the 0 μM inhibitor reaction Vmax. Curves are fit to a standard Michaelis-Menten kinetics model. e,f) GTP-dependent, helicase-stimulated primer synthesis activity from (c)and (d) replotted on an Eadie-Hofstee (v vs. v/s) plot. Linear fits use the same parameters as were determined by nonlinear regression fitting in (c) and (d).

Table 4.1: Data Collection and Refinement Statistics

	G5P Bound SaDg RPD 4EDV*	G4P Bound SaDg RPD 4EDT*
Crystal Data		
Space Group	P 6$_1$	P 6$_1$
Cell Dimensions (a, b, c (Å), α, β, γ(˚))	152.6,152.6,38.6 90, 90, 120	150.2,150.2, 38.6 90, 90, 120
Data Collection	Native	Native
Wavelength (Å)	1.115	1.115
Resolution (Å)	38.64-2.01	43.4-2.00
Reflections (measured/ unique)	277,309/34,211	284,153/35,073
†R$_{sym}$ (%) (highest shell)	8.3 (57.2)	7.5 (63.0)
<I/σI> (highest shell)	17.8 (4.1)	18.7 (3.2)
Completeness (%) (highest shell)	99.8(99.8)	99.7(98.1)
Redundancy	8.2	8.2
Refinement		
‡R$_{work}$/R$_{free}$ (%)	16.7/20.3	18.0/22.0
Average B factor (Å^2)		
protein	36.4	39.4
solvent	47.6	50.3
Rmsd		
Bond lengths (Å)	0.003	0.011
Bond angle (˚)	0.705	1.269
Protein residues	321	321
Water molecules	338	299
Ions	3	2
§Ramachandran (favored/ generous/disallowed)	317/4/0	317/4/0
Ligand Refinement		
Avg. B Factor (Å^2)	57.6	70.1
Real-space correlation coefficient	0.948	0.917

*PDB ID. **†Rsym** $= \Sigma|\ I-\langle I\rangle|/\Sigma I$, where I is the integrated intensity for a given reflection. **‡Rwork** $= \Sigma|Fobs – Fc|/\Sigma Fobs$, where Fobs and Fc are the observed and calculated structure factor amplitudes, respectively. Rfree is calculated as for Rwork, but from a subset of the data (5 %) that was withheld from crystallographic refinement. §As calculated by MolProbity (Chen, Arendall et al. 2010).

Chapter 5 - Investigation of the mechanism initiation specific inhibition of primer synthesis by 2´-dNTPs

INTRODUCTION

The mechanism by which DnaG ensures the synthesis of RNA primers (over DNA oligo-nucleotides) is normally attributed to its specificity for rNTPs during primer synthesis. A curious feature of DnaG, however, is its relatively robust ability to incorporate dNTPs into a primer *in vitro* (Rowen and Kornberg 1978; Swart and Griep 1995). This property suggests that DnaG-type primases may not possess a strong, inherent selectivity for rNTPs. Although primers produced *in vivo* are composed almost exclusively of RNA (Kitani, Yoda et al. 1985), this distribution may simply arise from the overabundance of ribonucleotides compared to deoxyribonucleotides (Neidhardt 1987). Despite this promiscuity, early studies of DnaG have shown that dNTPs can inhibit DnaG *in vitro* (Rowen and Kornberg 1978; Frick, Kumar et al. 1999), but only at an early stage of synthesis (Frick, Kumar et al. 1999). Consequently, DnaG appears to harbor a specific mechanism for rNTP selectivity during the initiation phase of synthesis.

To investigate the mechanism of dNTP utilization by DnaG, we solved the cocrystal structures of the *S. aureus* DnaG RPD bound to Mn^{2+} and several deoxynucleotide analogues. These structures illustrate a possible mechanism, involving aberrant nucleotide conformations, by which DnaG can select against dNTP incorporation during initiation, but occasionally allow it during extension.

RESULTS

We first investigated the sequence-specific incorporation of 2´-dNTPs by DnaG in *de novo* primer synthesis reactions using an established fluorometric primer synthesis assay (**Methods**). As shown in **Fig. 5.1a**, dGTP (GTP is utilized during primer synthesis initiation, **Methods**) impedes synthesis in a concentration-dependent fashion, but only after an initial increase in synthesis. Conversely, when we performed the same titration with a template in which dGTP can only be utilized during initiation (**Fig. 5.1b**), we observed only minimal primer synthesis (**Fig. 5.1a**). dTTP, on the other hand, leads only to a concentration-dependent increase in total primer synthesis, a response in accordance with its use in extension but not initiation (**Methods**). We next examined the effect of dATP using a template in which it can be utilized during extension, but not initiation (**Fig. 5.1b, Methods**). The resulting curve shows an intermediate effect between that seen for dGTP and dTTP, which is characterized by an initial increase, followed by subdued substrate inhibition. These results indicate that all dNTPs are incorporated into nascent primers to some extent, but with nonequivalent efficiencies even for equivalent incorporation positions.

To address the physical basis of these effects, we solved the structures of the *Sa*DnaG RPD bound to 2´- and 3´-dATP, and to both 2´-dGTP and 2´-dTTP (**Fig. 5.2a-d**), using methods described earlier (**Chapter 2**). The resulting structures all globally recapitulate the

previously-identified NTP binding site (**Fig. 5.3a**, **Chapter 2**), consistent with incorporation of all of the nucleotides into primer products. However, alignment of the 2´-dATP with 3´-dATP, rCTP, rGTP, or rUTP-bound models, reveals that that the 2´-deoxyribonucleotide binds in a tilted conformation involving a rotation of the ribose and base relative to the a-phosphate. A similar analysis with 2´-dTTP revealed an identical conformational change relative to 3´-dATP and the panel of rNTPs, while the sugar and base of dGTP were not sufficiently well-resolved to identify their conformation relative to the triphosphate. To investigate the significance of the conformational change observed with 2´-dATP and 2´-dTTP, we superposed a dinucleotide onto 3´-dATP after aligning the protein components of the 3´-dATP- and 2´-dATP-bound structures (**Fig. 5.3b**). This alignment indicates that the base of a nucleotide bound in the 2´-dATP-conformation would clash with a second base stacked against the incoming nucleotide.

As it has not been established whether the region occupied by the 5´-nucleotide of the modeled dinucleotide is involved in catalysis, we tested the catalytic importance of a highly-conserved residue (Ala314, Ala313 in *E. coli*) that is located near the hypothetical second-nucleotide binding site (**Fig. 5.4a**). Substitution of Ala313 in *E. coli* DnaG with glutamine or glutamate resulted in a reduction in total primer synthesis concomitant with the severity of the mutation, but without strongly affecting heteroduplex binding (**Fig. 5.4b/c**). Thus, mutagenesis is consistent with substrate binding to this region of the active site.

DISCUSSION

Synthesis of a ribonucleotide primer is a conserved feature of DNA replication in all cellular organisms. This adaptation may promote primer processing, thereby preventing incorporation of these highly error-prone oligomers into the chromosome, and contributing to the overall fidelity of DNA replication (Kuchta and Stengel 2010). Curiously, DnaG is not highly specific for rNTPs during primer extension, but is for initiation. Here, we have uncovered a possible molecular mechanism for this the rNTP specificity in DnaG during initiation. We discovered that at least some 2´-dNTPs bind in a conformation that would prevent binding of the 5´-most NTP during di-nucleotide condensation. We speculate that extension would not be affected by the tilted binding orientation of the dNTPs, as the steric block afforded by base stacking interactions between a nascent primer and the incoming nucleotide would favor a productive conformation more similar to the binding modality of the other rNTPs. We have shown that even very early extension events are tolerant of the presence of a dNTP, which is consistent with the ability of a nascent primer to induce productive binding of a dNTP.

It is unclear exactly how the dNTP binding modality is achieved, as we were unable to identify the specific residues that guide the dNTPs into the tilted conformation, in part due to the fact many of the interactions we could identify are required for catalysis. Although we do observe that a water-mediated hydrogen bond between Asn233 and the 2´-OH of rNTPs switches to the 3´-OH with dNTPs, we cannot say if this switch is the cause or an effect of the tilted binding conformation. Additional research will be necessary to determine if this conformation is effected by specific interactions within the DnaG active site, and how these interactions are overcome during the extension phase of synthesis.

MATERIALS AND METHODS

Cloning, expression, purification
See **Methods**, **Chapter 2**.

Crystallization of the SaDnaG RPD
Crystals were grown as described in **Chapter 2**. For soaks, crystals were first exchanged into well solution containing 20 mM $MnCl_2$ and 10 mM dNTP, and incubated overnight prior to cryoprotection and harvesting. Crystals were cryoprotected, harvested, and frozen as described in **Chapter 2**.

Data collection and refinement
See **Methods**, **Chapter 2**.

De novo primer synthesis assays
For all *de novo* synthesis experiments, reactions were carried out according to the protocol described in **Chapter 2**. For dNTP titrations, reactions were performed in triplicate, and are reported here as an average of the three values along with the Standard Deviation (SD). rNTPs were present at 0.1 mM, except the rNTP corresponding to the titrated dNTP was omitted. dNTP dilutions at 2x concentration (from 4 mM to 250 nM) were prepared by serially diluting dNTP stock solution (4 mM, in TE) 1:1 into TE. Reaction mixtures (10 µL) lacking the dNTP were incubated at room temperature for 10 min prior to addition of a diluted dNTP solution (10 µL) to form a complete reaction mixture. Complete reactions were incubated in a sealed 384-well plate (Bio-Rad) for 1 h at 37 ˚C, and then stopped by addition of 10 µL of a 1:67 dilution of PicoGreen stock solution (Invitrogen) in 20 mM Tris, 50 mM EDTA, pH 7.5, giving a final stain dilution of 1:200, as per the manufacturer's instructions. Stopped reactions were incubated in the dark for 5 min, spun at 3000 x g for 2 min, and fluorescence intensity measured in a PerkinElmer Victor3V multilabel plate reader. Raw fluorescence was background corrected with a no-NTP control. The V_{max} was calculated by fitting resultant curves to a substrate inhibition model (Eqn. 3.1) by nonlinear regression, where S is the concentration of dNTP, $K_{i,app}$ is the apparent inhibition constant for that dNTP, and h is the Hill coefficient. The data were then normalized to the fitted Vmax to give Relative Primer Synthesis (RPS).

Heteroduplex binding assays
See **Methods**, **Chapter 2**.

Visualization and data analysis methods
See **Methods**, **Chapter 2**.

a

Dose-response of total primer synthesis by DnaG to various dNTPs
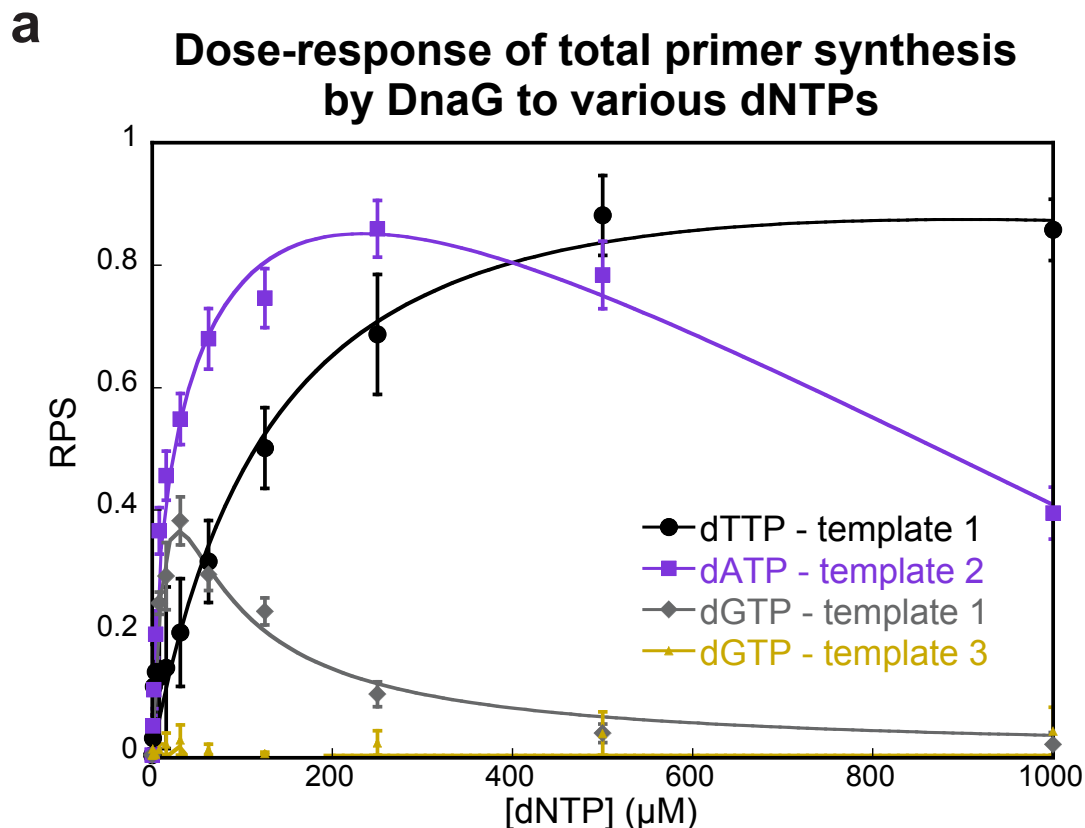
b

1) 5'-CACACACACACACA<u>CTG</u>AAAGCCAAAAG-3'

2) 5'-GTGTGTGTGTGTGT<u>CAG</u>AAAGCCAAAAG-3'

3) 5'-GAGAGAGAGAGAGA<u>CTG</u>AAAGCCAAAAG-3'

Figure 5.1: The effects of dGTP, dTTP and dATP on primer synthesis by E. coli DnaG.

a) Biochemical response of DnaG to the presence of dNTPs during different synthesis phases. The indicated nucleotides were titrated into reactions containing the indicated template, *Ec*DnaG and *Ec*DnaB along with the 3 remaining nucleotides as ribonucleotides. Curves were fit to a substrate inhibition model (Methods). Error bars represent ±SD (Methods). b) Schematic of templates used in (a). All templates were phosphorylated on the 3´-end.
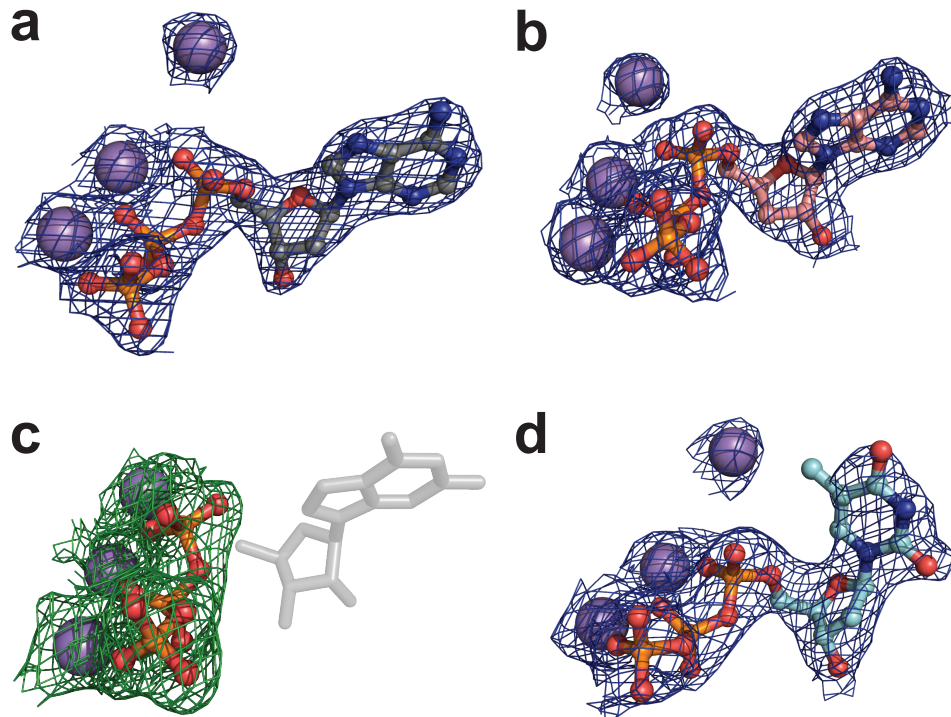
Figure 5.2: The crystal structures of the *Sa*DnaG RPD bound to 3´-dATP, 2´-dATP, 2´-dGTP, and 2´-dTTP.

a) Final refined $2F_o$-F_c density of 2´-dATP and Mn^{2+} bound to the *Sa*DnaG RPD contoured at 1 σ (blue mesh). 2´-dATP is shown as a ball and stick model with carbon colored gray. Mn^{2+} ions are shown as gray spheres. b) Final refined $2F_o$-F_c density of 3´-dATP and Mn^{2+} bound to the *Sa*DnaG RPD contoured at 1 σ (blue mesh). 3´-dATP is shown as a ball and stick model with carbon colored salmon. Mn^{2+} ions are shown as gray spheres. c) Difference density (mF_o-DF_c green mesh) contoured at 3 σ present in the active site from data obtained from a crystal soaked with 2´-dGTP and Mn^{2+}. Only the triphosphate and metal ions are visible in the density. These were modeled in after several rounds of refinement. GTP is shown as white sticks to indicate the location the ribose and base are normally observed. Notably in this structure, all 3 metal ions are collinear, and the α-phosphate interacts with metal A. d) Final refined $2F_o$-F_c density of 2´-dTTP and Mn^{2+} bound to the *Sa*DnaG RPD contoured at 1 σ (blue mesh). 2´-dTTP is shown as a ball and stick model with carbon colored cyan. Mn^{2+} ions are shown as gray spheres.
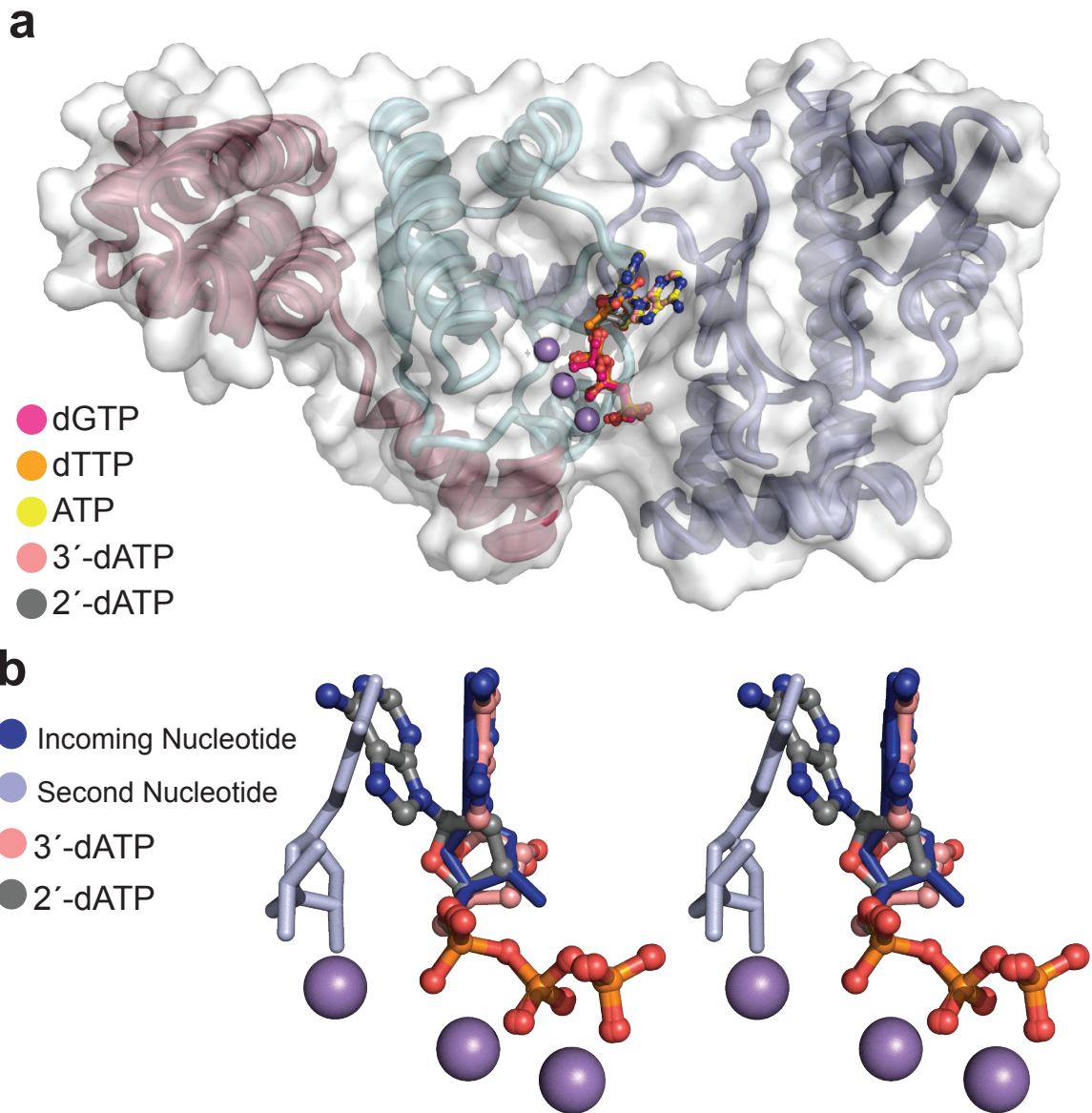
Figure 5.3: The 2´-dATP bound structure illustrates a possible mechanism of initiation specific primer synthesis inhibition by 2´-dNTPs.

a) 2´-dATP binds in a tilted conformation. A superposition of the ATP bound *Sa*DnaG RPD structure (white surface, and yellow nucleotide) and the four dNTP bound SaDnaG RPD structures (cartoons colored by subdomain, as per Fig. 2.1a) are shown. NTPs are shown as ball and stick representations with carbon colored gray (2´-dATP), salmon (3´-dATP), orange (2´-dTTP), or yellow (ATP). Mn^{2+} ions are shown as gray spheres. The tripolyphosphate model represents the position of the 2´-dGTP triphosphate moiety, and is colored magenta. b) The conformation of 2´-dATP would clash with a second nucleotide bound to the active site. Stereo diagram of an alignment of the 3´-dATP bound nucleotide (salmon), and a modeled dinucleotide. The dark blue nucleotide of the dinucleotide represents the +1, or incoming nucleotide, and the light blue nucleotide represents the -1, or initiating nucleotide. The 2´-dATP bound structure was then superposed onto the 3´-dATP bound structure.

Figure 5.4: Ala313 is required for maximal primer synthesis, but not heteroduplex binding with *E. coli* DnaG.

Effect of substitutions to Ala313 on *de novo* primer synthesis by *E. coli* DnaG
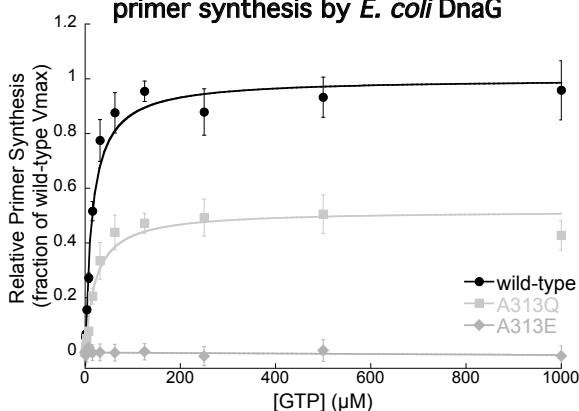
Influence of Ala313Gln substitution on heteroduplex binding by *Ec*DnaG RPD
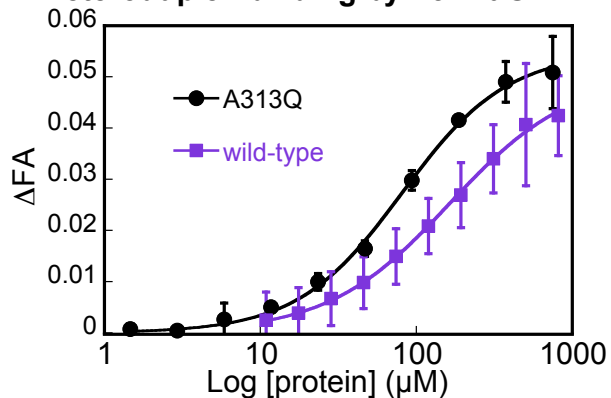
a) View of the DnaG active site in the 3´-dATP bound structure highlighting the tight distance restraints with Ala313 and an initiating nucleotide. The protein portion of the RPD of the 3´-dATP bound structure (**Chapter 2**) is shown as a cartoon and white surface, and the bound 3´-dATP as a ball and stick model with carbon colored salmon. Mn^{2+} ions are shown as gray spheres. A docked dinucleotide (see **Fig. 5.3**) is shown, and is colored as in **Fig. 5.3b**. Conserved residue Ala314 (corresponding to Ala313 in *Ec*DnaG) is highlighted in red. The surface of the initiating nucleotide is shown as a light blue surface to illustrate the distance between the initiating nucleotide and Ala314. b) Ala313 is critical for DnaG function. The results of a GTP-dependent, helicase-stimulated de novo primer synthesis assay are shown for wild-type (black circles), Ala313Gln (gray squares), and Ala313Glu (light-gray diamonds) *E. coli* DnaG. Primer synthesis extent is reported as a fraction of the wild-type Vmax. Curves represent fits to a standard Michaelis-Menten kinetics model (**Methods**); error bars represent ±SEM (**Methods**). c) The plot shows the change in fluorescence anisotropy of a fluorescently-labeled heteroduplex as a function of DnaG concentration (**Methods**). Data are shown for binding to the wild-type *E. coli* RPD (purple squares), and to the Ala313Gln (black circles) mutant protein. Data were fit to the explicit solution to single site binding (**Methods**). Error bars represent ±SD (**Methods**). The fitted apparent dissociation constants differ by less than 2-fold.

Chapter 6 – An improved labeled nucleotide incorporation primer synthesis assay for bacterial primase

INTRODUCTION

Primases are DNA dependent RNA polymerases required by all cellular organisms to synthesize short RNA oligonucleotide "primers" needed by DNA polymerases to replicate DNA (Kornberg 1992). In bacteria, DnaG functions as the replicative primase (Frick and Richardson 2001), and has been cited recently as an underexploited target for the developments of novel antibiotics (Robinson, Causer et al. 2012). The case for DnaG as a drug target stems primarily from the fact that it bears no sequence or structural homology to the eukaryotic primase (Kuchta and Stengel 2010), or any other polymerase. In accordance with this structural divergence, DnaG displays many biochemical behaviors unique to polymerases, such as sequence specificity in the absence of accessory factors, and autoregulation of product length (Corn and Berger 2006).

The structure-function relationships underlying these unique activities are not well understood. In the course of our studies to better understand how DnaG functions in substrate recognition and catalysis, we have most often employed a primer synthesis assay that relies on incorporation of a radiolabeled nucleotide into the product primer. The products are separated by denaturing gel electrophoresis, and visualized by phoshporimaging. An improved assay has been available for some time, in which thermally denaturing HPLC is used in place of gel electrophoresis, and native products are detected directly by UV absorption (Koepsell, Bastola et al. 2004). Importantly, both methods allow the simultaneous measurement of both the distribution of primer length, as well as total primer synthesis. This method has been employed to great effect in the study of the fundamental enzymology of DnaG (Koepsell, Larson et al. 2006; Koepsell, Larson et al. 2008; Larson, Griep et al. 2010). However, implementation of the HPLC method requires a specialized and dedicated instrument that precludes wider and more commonplace usage.

We have recently developed a companion "benchtop" assay to the HPLC assay in the form of a fluorescent nucleotide incorporation primer synthesis assay. This assay combines the ease of execution and information density of the HPLC method, with the uniform availability of the equipment required for the radiolabel incorporation method.

MATERIALS AND METHODS

Full length DnaG was expressed and purified as described (Corn, Pelton et al. 2008). Unless otherwise noted, reactions were carried out with 20 μM primase, 40 μM of a single Fluorescen-12-NTP (PerkinElmer), 200 μM of the other three NTPs, and 30 μM template DNA. The template DNA oligos were phosphorylated on the 3´-end, and were purchased from IDT. Sequences of the templates are reported in **Fig. 6.1**. Reaction buffer consisted of 100 mM potassium glutamate, 10 mM Mg acetate, 20 mM HEPES, 5% glycerol, and 0.2 mg/mL BSA. Reactions (25 μL total volume) were mixed and then incubated at 37 ˚C for 1 h. Stop buffer

(2.5 µL) containing 1 mg/mL proteinase K (Promega), 20 mM EDTA, 2% SDS, was added to each reaction, and the temperature raised to 55 ˚C for 10 min. Free label was removed and the buffer exchanged to 90% Formamide, 10 mM EDTA with a Bio-Rad MicroBioSpin 6 column according to the manufacture instructions. Reaction products were separated on a 15% polyacrylamide (40:1 bis:acrylamide, Fisher Scientific) containing 8 M urea and 1x tris-taureen-EDTA (TTE) buffer poured in a 1.5 mm Novex® Gel Cassettes (Invitrogen). Electrophoresis was conducted in an XCell SureLock™ Mini-Cell (Invitrogen) electrophoresis chamber with 1x TTE running buffer at 20 V/cm for ~ 45 min. Gels were prerun and the wells rinsed extensively prior to sample loading. Fluorescent ladder consisted of fluorescein labeled CA repeat DNA oligos of the specified length (Elim Biopharmaceuticals). Fluorescent products were visualized on a standard UV light box. Quantification was carried out by measurement of the total sample fluorescence prior to electrophoresis with a PerkinElmer Victor3V multilabel plate reader, and normalized to a no-enzyme control.

RESULTS AND DISCUSSION

Fluorescent nucleotide incorporation is a common method utilized with nucleic acid polymerases (Kricka 2002). Although our first attempts to observe fluorescent nucleotide incorporation were unsuccessful, we discovered that high-concentrations of template, 100 µM of template and 20 µM full-length protein, afforded fluorescent-nucleotide incorporation with both fluorescein-12-ATP and fluorescein-12-GTP (not shown). A standard enzyme titration (**Fig. 6.1a**) revealed that the rate of incorporation is dependent on enzyme concentration, both by visual inspection of the gel, as well as analysis of total sample fluorescence just prior to electrophoresis (**Fig. 6.1b**).

As such a high concentration of template might interfere with aspects of DnaG primer synthesis, like sequence specificity, we investigated the effect of a sequence specific inhibitor, 3´-dATP (3DA), on primer synthesis using the fluorescent nucleotide incorporation assay. A titration of 3DA with the standard radiolabel incorporation primer synthesis assay generates a characteristic, and concentration dependent reduction in primer size due to chain termination (**Fig. 6.1c**). We observed a highly similar primer length trend with a 3DA titration conducted using the fluorescent nucleotide incorporation assay (**Fig. 6.1d**). An analogous experiment with GDP, another sequence specific primase inhibitor, produced comparable results (data not shown).

The fluorescent label incorporation primer synthesis assay presented here represents a convenient method for interrogating primer synthesis by DnaG without using radioactivity. One of the strengths of the method is the ability to easily link total activity to primer length distribution by easily quantifying total sample fluorescence just prior to gel electrophoresis. Although possible to do so with the analogous radiolabel incorporation method, the use of a plate reader or fluorometer is far more convenient than scintillation counting, and more accurate than band intensity quantification. A major limitation of the assay is its relatively limited throughput; we are currently addressing this problem by studying alternative label removal methods.

Figure 6.1: Development of a fluorescently labeled nucleotide incorporation assay for DnaG.
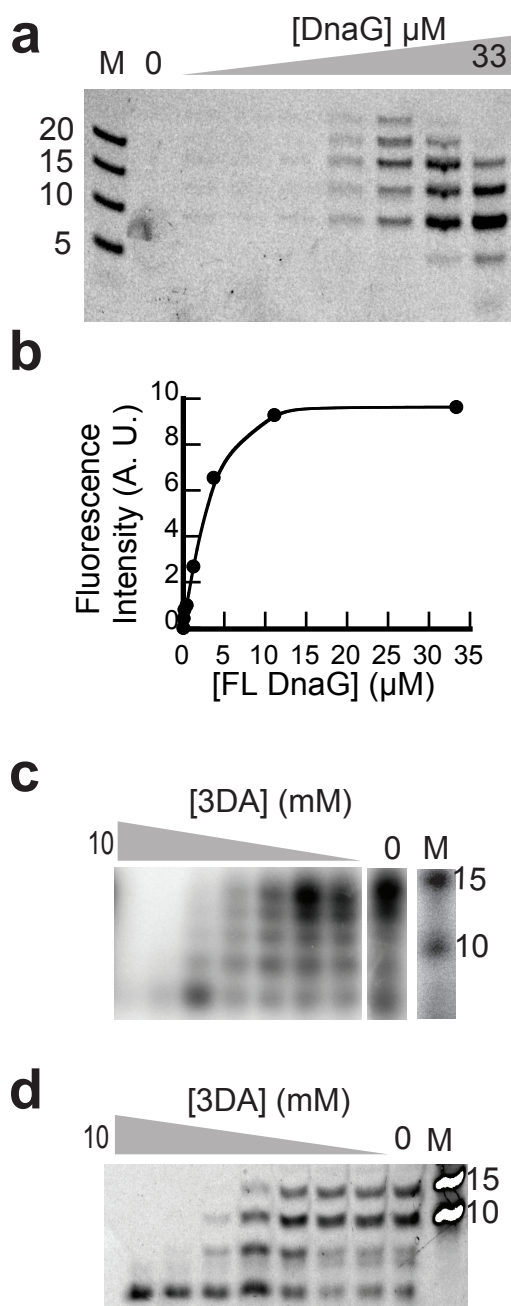
a) Full-length DnaG titration with a 3´-phosphory-lated single stranded template (5´-CACACACA-CACACACTGAAAGCCAAAAG-3´), and fluorescein-12-UTP as the labeled nucleotide. Starting at 33 µM, subsequent samples are 1:3 dilutions of the preceding sample. Image shows the gel electro-phoresis results from the enzyme titration. An equivalent volume (10 µL) was loaded for each reaction. Lane M, size marker. b) Quantification of fluorescent signal in 20 µL of each titration sample from (a). The total fluorescence from the no enzyme sample was subtracted as background. c) A 3´-dATP titration from 10 mM to 14 µM in 1:3 steps using a radiolabel incorporation assay. Reac-tions were carried out with 1 µCi of 80 mCi/mmol alpha-32P-UTP, 400 nM of 3´-phosphorylated tem-plate DNA of sequence 5´-ATATATATATATATCT-GAAAGCCAAAAG-3´, and 200 µM ATP, CTP, and GTP. Lane M, size marker. d) A 3´-dATP titration from 10 mM to 14 µM in 1:3 steps using the fluo-rescent label incorporation assay. Reactions were carried out with 40 µM Fluorescein-12-UTP, 30 µM of the template used in (c), and 200 µM ATP, CTP, and GTP. Lane M, size marker. For all panels, marker sizes are reported in nucleotides.

Chapter 7 – Concluding Remarks

Bacterial DnaG-type primases represent a distinctive class of nucleotide polymerases. First, these enzymes are structurally very different from all other polymerases identified to date. Second, they form a critical component of the bacterial DNA replication machinery, yet synthesize RNA, rather than DNA, oligonucleotides. In the current work, we have examined the structures and regulation of some of the earliest steps of primer synthesis by DnaG, with particular focus on nucleotidyl and small-molecule inhibitor interactions.

Chapter 2 presents evidence for a reconciled model for primer synthesis by DnaG. The structures that lead to this model illustrate how the metal-binding TOPRIM fold and the N-terminal subdomain of DnaG act in concert to bind the incoming nucleoside triphosphate. Moreover, comparisons with the structure of a covalent protein•DNA complex of yeast topoisomerase II reveal the function of additional metal-binding residues unique to the DnaG active site. DnaG apparently utilizes a unique variant of two-metal mechanism, the "2+1-metal mechanism." In this mechanism, two of the metals are analogous to metals A and B of the two-metal mechanism by flanking the a-phosphate. However, instead of metal B chelating the entire triphosphate moiety, a third metal binds the b- and g-phosphates (metal C). At present, it is unclear whether DnaG requires metal C for primer synthesis in general, or whether, for example, all three metals are required only during initiation of primer synthesis.

This work also uncovered structural details of primer synthesis regulation by two types of natural small-molecule DnaG inhibitors, (p)ppGpp (**Chapter 4**) and 2´-dNTPs (**Chapter 5**). Both of these studies support the crystallographic approach employed in Chapter 2, while the former provides a mechanistic basis for the response of DnaG to changing cellular conditions. Specifically, during nutrient starvation, (p)ppGpp occludes the metal-binding region of the DnaG active site, thereby disrupting formation of new primers and bringing Okazaki fragment synthesis to a halt. The latter study, an investigation of the mechanism of initiation specific inhibition of primer synthesis by 2´-dNTPs, provides a physical explanation for a perplexing feature of *in vitro* primer synthesis by DnaG.

Taken together, these studies illustrate how the structural divergence of DnaG from other polymerases translates to differences in function. For example, in most other polymerases, the triphosphate is stabilized by basic amino acids protruding from a mobile element called the fingers domain (Rothwell and Waksman 2005). DnaG employs an analogous set of basic amino acids, but rigidly associates these groups with the metal-binding active site (**Chapter 2**). Thus, the incoming nucleotide-binding site of DnaG is preformed, imparting an inherent rigidity to DnaG that explains why a fingers domain is not necessary. This property may also explain why the nucleotide soaking experiments performed in the present work yielded a productively oriented nucleoside triphosphate bound in the active site of DnaG. Similar experiments with most DNA and RNA polymerases have been somewhat less informative, and typically result only in nucleotide binding to a "pre-insertion" site present in the fingers domain (Li, Kong et al. 1998; Thompson, Albertini et al. 2007). For most other single-subunit polymerases, a primer•template complex must be bound to the enzyme for

an incoming nucleotide to be captured and positioned in a catalytically-competent conformation (Li, Korolev et al. 1998; Gong and Peersen 2010). This requirement does not appear to hold for DnaG. Moreover, our data suggest that DnaG binds the incoming nucleotide in a manner more akin to a lock-and-key mechanism, whereas other polymerases bind the incoming nucleotide using induced fit. A functional consequence of this work is the development of several novel and improved primer synthesis assays. Details on one such assay are described in **Chapter 6**.

The next task facing the field is to understand how the N-terminal Zinc Binding Domain directs primase to appropriate start sites on DNA, and then regulates the length of the subsequent primer product. The model for primer synthesis presented in **Chapter 2** implies that the ZBD could act analogously to the thumb domain of canonical polymerases. However, recent evidence indicates that the ZBD interacts with a nucleotide in the active site (Lee, Zhu et al. 2010), an observation that would be more consistent with a combined fingers-thumb role. To differentiate between these two models, it will be necessary to obtain an atomically detailed model of the ZBD bound to a template or primer•template complex. Ideally, the atomic structure of a full quaternary complex with the RPD, the ZBD, a primer•template substrate, and an incoming nucleotide would be solved. Towards accomplishing this goal, preliminary data are presented in **Appendix III** for covalently and productively tethering the DnaG RPD bound to a primer•template substrate that could serve as a substrate for ZBD binding and structural studies. Clarification of the role of the ZBD in substrate recognition and catalysis would complete our understanding of the structural and functional differences between DnaG and the rest of the members of the DNA and RNA polymerase superfamily.

In conclusion, the present work establishes that DnaG is not only structurally, but also functionally divergent from other polymerases. As DnaG is found only in bacteria, these insights could motivate future efforts to develop a primase-targeted antimicrobial agent. An initial step in this direction would be adapting an improved primer synthesis assay to high-throughput screening platforms to search for lead inhibitors. In combination with additional research into inhibitors of bacterial DNA replication (Dallmann, Fackelmayer et al. 2010), such an effort could lead to a truly universal antibiotic.

References

Adams, P. D., P. V. Afonine, et al. (2011). "The Phenix software for automated determination of macromolecular structures." Methods **55**(1): 94-106.

Agarwal, A., S. Louise-May, et al. (2007). "Small molecule inhibitors of E. coli primase, a novel bacterial target." Bioorg Med Chem Lett **17**(10): 2807-2810.

Akabayov, B., S. J. Lee, et al. (2009). "DNA recognition by the DNA primase of bacteriophage T7: a structure-function study of the zinc-binding domain." Biochemistry **48**(8): 1763-1773.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-410.

Aravind, L., D. D. Leipe, et al. (1998). "Toprim - a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins." Nucleic Acids Research **26**(18): 4205-4213.

Bailey, S., W. K. Eliason, et al. (2007). "Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase." Science **318**(5849): 459-463.

Bax, B. D., P. F. Chan, et al. (2010). "Type IIA topoisomerase inhibition by a new class of anti-bacterial agents." Nature 466(7309): 935-940.

Bernstein, J. A. and C. C. Richardson (1988). "A 7-kDa region of the bacteriophage T7 gene 4 protein is required for primase but not for helicase activity." Proc Natl Acad Sci U S A **85**(2): 396-400.

Bhattacharyya, S. and M. A. Griep (2000). "DnaB helicase affects the initiation specificity of Escherichia coli primase on single-stranded DNA templates." Biochemistry **39**(4): 745-752.

Bird, L. E., H. Pan, et al. (2000). "Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from Bacillus stearothermophilus." Biochemistry **39**(1): 171-182.

Bouche, J. P., K. Zechel, et al. (1975). "dnaG gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form." J Biol Chem **250**(15): 5995-6001.

Brautigam, C. A. and T. A. Steitz (1998). "Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes." Curr Opin Struct Biol **8**(1): 54-63.

Cameron, C. E., I. M. Moustafa, et al. (2009). "Dynamics: the missing link between structure and function of the viral RNA-dependent RNA polymerase?" Curr Opin Struct Biol **19**(6): 768-774.

Carl, P. L. (1970). "Escherichia coli mutants with temperature-sensitive synthesis of DNA." Mol Gen Genet **109**(2): 107-122.

Chen, V. B., W. B. Arendall, 3rd, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Chintakayala, K., M. A. Larson, et al. (2007). "Domain swapping reveals that the C- and N-terminal domains of DnaG and DnaB, respectively, are functional homologues." Mol Microbiol **63**(6): 1629-1639.

Chintakayala, K., C. Machon, et al. (2009). "Allosteric regulation of the primase (DnaG) activity by the clamp-loader (tau) in vitro." Mol Microbiol **72**(2): 537-549.

Chu, M., R. Mierzwa, et al. (2003). "Isolation and structure elucidation of Sch 642305, a novel bacterial DNA primase inhibitor produced by Penicillium verrucosum." J Nat Prod **66**(12): 1527-1530.

Copeland, R. A. (2005). "Evaluation of enzyme inhibitors in drug discovery. A guide for medicinal chemists and pharmacologists." Methods Biochem Anal **46**: 1-265.

Corn, J. E. and J. M. Berger (2006). "Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions." Nucleic Acids Research **34**(15): 4082-4088.

Corn, J. E. and J. M. Berger (2007). "FASTDXL: a generalized screen to trap disulfide-stabilized complexes for use in structural studies." Structure **15**(7): 773-780.

Corn, J. E., P. J. Pease, et al. (2005). "Crosstalk between primase subunits can act to regulate primer synthesis in trans." Molecular Cell **20**(3): 391-401.

Corn, J. E., J. G. Pelton, et al. (2008). "Identification of a DNA primase template tracking site redefines the geometry of primer synthesis." Nature Structural & Molecular Biology **15**: 163-169.

Cramer, P. (2002). "Multisubunit RNA polymerases." Curr Opin Struct Biol **12**(1): 89-97.

Cramer, P., K. J. Armache, et al. (2008). "Structure of eukaryotic RNA polymerases." Annu Rev Biophys **37**: 337-352.

Dallmann, H. G., O. J. Fackelmayer, et al. (2010). "Parallel multiplicative target screening against divergent bacterial replicases: identification of specific inhibitors with broad spectrum potential." Biochemistry **49**(11): 2551-2562.

Doublie, S. and T. Ellenberger (1998). "The mechanism of action of T7 DNA polymerase." Curr Opin Struct Biol **8**(6): 704-712.

Emsley, P., B. Lohkamp, et al. (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Frick, D. N., S. Kumar, et al. (1999). "Interaction of ribonucleoside triphosphates with the gene 4 primase of bacteriophage T7." Journal of Biological Chemistry **274**(50): 35899-35907.

Frick, D. N. and C. C. Richardson (1999). "Interaction of bacteriophage T7 gene 4 primase with its template recognition site." Journal of Biological Chemistry **274**(50): 35889-35898.

Frick, D. N. and C. C. Richardson (2001). "DNA primases." Annu Rev Biochem **70**: 39-80.

Godson, G. N., J. Schoenich, et al. (2000). "Identification of the magnesium ion binding site in the catalytic center of Escherichia coli primase by iron cleavage." Biochemistry **39**(2): 332-

339.

Gong, P. and O. B. Peersen (2010). "Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase." Proc Natl Acad Sci U S A **107**(52): 22505-22510.

Griep, M. A., B. J. Adkins, et al. (1997). "The tyrosine photophysics of a primase-derived peptide are sensitive to the peptide's zinc-bound state: proof that the bacterial primase hypothetical zinc finger sequence binds zinc." Biochemistry **36**(3): 544-553.

Griep, M. A. and E. R. Lokey (1996). "The role of zinc and the reactivity of cysteines in Escherichia coli primase." Biochemistry **35**(25): 8260-8267.

Hamdan, S. M. and A. M. van Oijen (2010). "Timing, coordination, and rhythm: acrobatics at the DNA replication fork." J Biol Chem **285**(25): 18979-18983.

Heitz, F., P. Harter, et al. (2010). "Poly(ADP-ribosyl)ation polymerases: mechanism and new target of anticancer therapy." Expert Rev Anticancer Ther **10**(7): 1125-1136.

Heller, R. C. and K. J. Marians (2006). "Replication fork reactivation downstream of a blocked nascent leading strand." Nature **439**(7076): 557-562.

Hine, A. V. and C. C. Richardson (1994). "A functional chimeric DNA primase: the Cys4 zinc-binding domain of bacteriophage T3 primase fused to the helicase of bacteriophage T7." Proc Natl Acad Sci U S A **91**(25): 12327-12331.

Jain, V., M. Kumar, et al. (2006). "ppGpp: stringent response and survival." J Microbiol **44**(1): 1-10.

Johnson, S. K., S. Bhattacharyya, et al. (2000). "DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates." Biochemistry **39**(4): 736-744.

Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Kato, M., D. N. Frick, et al. (2001). "A complex of the bacteriophage T7 primase-helicase and DNA polymerase directs primer utilization." J Biol Chem **276**(24): 21809-21820.

Kato, M., T. Ito, et al. (2004). "A molecular handoff between bacteriophage T7 DNA primase and T7 DNA polymerase initiates DNA synthesis." Journal of Biological Chemistry **279**(29): 30554-30562.

Kato, M., T. Ito, et al. (2003). "Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis." Mol Cell **11**(5): 1349-1360.

Katoh, K., K. Misawa, et al. (2002). "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." Nucleic Acids Res **30**(14): 3059-3066.

Keck, J. L., D. D. Roche, et al. (2000). "Structure of the RNA polymerase domain of E. coli primase." Science **287**(5462): 2482-2486.

Khopde, S., E. E. Biswas, et al. (2002). "Affinity and sequence specificity of DNA binding and site selection for primer synthesis by Escherichia coli primase." Biochemistry **41**(50): 14820-14830.

Kiefer, J. R., C. Mao, et al. (1998). "Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal." Nature **391**(6664): 304-307.

Kitani, T., K. Y. Yoda, et al. (1985). "EVIDENCE THAT DISCONTINUOUS DNA-REPLICATION IN ESCHERICHIA-COLI IS PRIMED BY APPROXIMATELY 10 TO 12 RESIDUES OF RNA STARTING WITH A PURINE." Journal of Molecular Biology **184**(1): 45-52.

Koepsell, S., D. Bastola, et al. (2004). "Thermally denaturing high-performance liquid chromatography analysis of primase activity." Anal Biochem **332**(2): 330-336.

Koepsell, S. A., S. Hanson, et al. (2005). "Fluorometric assay for bacterial primases." Anal Biochem **339**(2): 353-355.

Koepsell, S. A., M. A. Larson, et al. (2008). "Staphylococcus aureus primase has higher initiation specificity, interacts with single-stranded DNA stronger, but is less stimulated by its helicase than Escherichia coli primase." Mol Microbiol **68**(6): 1570-1582.

Koepsell, S. A., M. A. Larson, et al. (2006). "Staphylococcus aureus helicase but not Escherichia coli helicase stimulates S. aureus primase activity and maintains initiation specificity." J Bacteriol **188**(13): 4673-4680.

Kohanski, M. A., D. J. Dwyer, et al. (2010). "How antibiotics kill bacteria: from targets to networks." Nat Rev Microbiol **8**(6): 423-435.

Kolodner, R., Y. Masamune, et al. (1978). "Gene 4 protein of bacteriophage T7. Purification physical properties, and stimulation of T7 DNA polymerase during the elongation of polynucleotide chains." J Biol Chem **253**(2): 566-573.

Kolodner, R. and C. C. Richardson (1977). "Replication of duplex DNA by bacteriophage T7 DNA polymerase and gene 4 protein is accompanied by hydrolysis of nucleoside 5'-triphosphates." Proc Natl Acad Sci U S A **74**(4): 1525-1529.

Kornberg, A. B., Tania A (1992). DNA Replication. New York, W. H. Freeman and Company.

Kricka, L. J. (2002). "Stains, labels and detection strategies for nucleic acids assays." Ann Clin Biochem **39**(Pt 2): 114-129.

Kuchta, R. D. and G. Stengel (2010). "Mechanism and evolution of DNA primases." Biochim Biophys Acta **1804**(5): 1180-1189.

Kusakabe, T., K. Baradaran, et al. (1998). "Roles of the helicase and primase domain of the gene 4 protein of bacteriophage T7 in accessing the primase recognition site." EMBO J **17**(5): 1542-1552.

Kusakabe, T., A. V. Hine, et al. (1999). "The Cys4 zinc finger of bacteriophage T7 primase in sequence-specific single-stranded DNA recognition." Proc Natl Acad Sci U S A **96**(8): 4295-4300.

Kusakabe, T. and C. C. Richardson (1996). "The role of the zinc motif in sequence recognition by DNA primases." Journal of Biological Chemistry **271**(32): 19563-19570.

Laponogov, I., X. S. Pan, et al. (2010). "Structural basis of gate-DNA breakage and resealing by type II topoisomerases." PLoS One 5(6): e11338.

Larson, M. A., R. Bressani, et al. (2008). "Hyperthermophilic Aquifex aeolicus initiates primer synthesis on a limited set of trinucleotides comprised of cytosines and guanines." Nucleic Acids Res **36**(16): 5260-5269.

Larson, M. A., M. A. Griep, et al. (2010). "Class-specific restrictions define primase interactions with DNA template and replicative helicase." Nucleic Acids Res **38**(20): 7167-7178.

Lee, J. B., R. K. Hite, et al. (2006). "DNA primase acts as a molecular brake in DNA replication." Nature **439**(7076): 621-624.

Lee, S. J. and C. C. Richardson (2001). "Essential lysine residues in the RNA polymerase domain of the gene 4 primase-helicase of bacteriophage T7." J Biol Chem **276**(52): 49419-49426.

Lee, S. J. and C. C. Richardson (2002). "Interaction of adjacent primase domains within the hexameric gene 4 helicase-primase of bacteriophage T7." Proc Natl Acad Sci U S A **99**(20): 12703-12708.

Lee, S. J. and C. C. Richardson (2005). "Acidic residues in the nucleotide-binding site of the bacteriophage T7 DNA primase." J Biol Chem **280**(29): 26984-26991.

Lee, S. J., B. Zhu, et al. (2010). "Mechanism of sequence-specific template binding by the DNA primase of bacteriophage T7." Nucleic Acids Res **38**(13): 4372-4383.

Levin, B. R. and D. E. Rozen (2006). "Non-inherited antibiotic resistance." Nat Rev Microbiol **4**(7): 556-562.

Li, Y., Y. Kong, et al. (1998). "Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates." Protein Sci **7**(5): 1116-1123.

Li, Y., S. Korolev, et al. (1998). "Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation." EMBO J **17**(24): 7514-7525.

Liu, C. C. and B. M. Alberts (1980). "Pentaribonucleotides of mixed sequence are synthesized and efficiently prime de novo DNA chain starts in the T4 bacteriophage DNA replication system." Proc Natl Acad Sci U S A **77**(10): 5698-5702.

Lu, Y. B., P. V. Ratnakar, et al. (1996). "Direct physical interaction between DnaG primase and DnaB helicase of Escherichia coli is necessary for optimal synthesis of primer RNA." Proc Natl Acad Sci U S A **93**(23): 12902-12907.

MacDowell, A. A., R. S. Celestre, et al. (2004). "Suite of three protein crystallography beamlines with single superconducting bend magnet as the source." J Synchrotron Radiat **11**(Pt 6): 447-455.

Maciag, M., M. Kochanowska, et al. (2010). "ppGpp inhibits the activity of Escherichia coli DnaG primase." Plasmid **63**(1): 61-67.

Makowska-Grzyska, M. and J. M. Kaguni (2010). "Primase directs the release of DnaC from DnaB." Mol Cell **37**(1): 90-101.

Manosas, M., M. M. Spiering, et al. (2009). "Coupling DNA unwinding activity with primer synthesis in the bacteriophage T4 primosome." Nat Chem Biol **5**(12): 904-912.

Marceau, A. H., S. Bahng, et al. (2011). "Structure of the SSB-DNA polymerase III interface and its role in DNA replication." EMBO J **30**(20): 4236-4247.

Marians, K. J. (1992). "PROKARYOTIC DNA-REPLICATION." Annual Review of Biochemistry **61**: 673-719.

Matson, S. W., S. Tabor, et al. (1983). "The gene 4 protein of bacteriophage T7. Characterization of helicase activity." J Biol Chem **258**(22): 14017-14024.

Mechold, U., H. Murphy, et al. (2002). "Intramolecular regulation of the opposing (p)ppGpp catalytic activities of Rel(Seq), the Rel/Spo enzyme from Streptococcus equisimilis." J Bacteriol **184**(11): 2878-2888.

Mehta, G. and H. M. Shinde (2005). "Enantioselective total synthesis of bioactive natural product (+)-Sch 642305: a structurally novel inhibitor of bacterial DNA primase and HIV-1 Tat transactivation." Chem Commun (Camb)(29): 3703-3705.

Mendelman, L. V., B. B. Beauchamp, et al. (1994). "Requirement for a zinc motif for template recognition by the bacteriophage T7 primase." EMBO J **13**(16): 3909-3916.

Mendelman, L. V., R. G. Kuimelis, et al. (1995). "Effects of base analog substitutions in the noncoding dC of the 3'-d(CTG)-5' template recognition site of the bacteriophage T7 primase." Biochemistry **34**(32): 10187-10193.

Mendelman, L. V. and C. C. Richardson (1991). "Requirements for primer synthesis by bacteriophage T7 63-kDa gene 4 protein. Roles of template sequence and T7 56-kDa gene 4 protein." J Biol Chem **266**(34): 23240-23250.

Mitkova, A. V., S. M. Khopde, et al. (2003). "Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of Escherichia coli in RNA primer synthesis." J Biol Chem **278**(52): 52253-52261.

Mustaev, A. A. and G. N. Godson (1995). "Studies of the functional topography of the catalytic center of Escherichia coli primase." J Biol Chem **270**(26): 15711-15718.

Neidhardt, F. C. (1987). Escherichia coli and Salmonella typhimurium : cellular and molecular biology. Washington, D.C., American Society for Microbiology.

Nelson, S. W., R. Kumar, et al. (2008). "RNA primer handoff in bacteriophage T4 DNA replication: the role of single-stranded DNA-binding protein and polymerase accessory proteins." J Biol Chem **283**(33): 22838-22846.

Nicoletti, R., M. P. Lopez-Gresa, et al. (2007). "Production and fungitoxic activity of Sch 642305, a secondary metabolite of Penicillium canescens." Mycopathologia **163**(5): 295-301.

Nossal, N. G. (1980). "RNA priming of DNA replication by bacteriophage T4 proteins." J Biol Chem **255**(5): 2176-2182.

Oakley, A. J., K. V. Loscha, et al. (2005). "Crystal and solution structures of the helicase-binding domain of Escherichia coli primase." J Biol Chem **280**(12): 11495-11504.

Pan, H. and D. B. Wigley (2000). "Structure of the zinc-binding domain of Bacillus stearothermophilus DNA primase." Structure **8**(3): 231-239.

Podobnik, M., P. McInerney, et al. (2000). "A TOPRIM domain in the crystal structure of the catalytic core of Escherichia coli primase confirms a structural link to DNA topoisomeras-

es." <u>J Mol Biol</u> **300**(2): 353-362.

Potrykus, K. and M. Cashel (2008). "(p)ppGpp: still magical?" <u>Annu Rev Microbiol</u> **62**: 35-51.

Powers, L. and M. A. Griep (1999). "Escherichia coli primase zinc is sensitive to substrate and cofactor binding." <u>Biochemistry</u> **38**(23): 7413-7420.

Qimron, U., S. J. Lee, et al. (2006). "Primer initiation and extension by T7 DNA primase." <u>Embo Journal</u> **25**(10): 2199-2208.

Robinson, A., R. J. Causer, et al. (2012). "Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target." <u>Curr Drug Targets</u> **13**(3): 352-372.

Rodina, A. and G. N. Godson (2006). "Role of conserved amino acids in the catalytic activity of Escherichia coli primase." <u>J Bacteriol</u> **188**(10): 3614-3621.

Romano, L. J. and C. C. Richardson (1979). "Characterization of the ribonucleic acid primers and the deoxyribonucleic acid product synthesized by the DNA polymerase and gene 4 protein of bacteriophage T7." <u>J Biol Chem</u> **254**(20): 10483-10489.

Rothwell, P. J. and G. Waksman (2005). "Structure and mechanism of DNA polymerases." <u>Adv Protein Chem</u> **71**: 401-440.

Rowen, L. and A. Kornberg (1978). "Primase, the dnaG protein of Escherichia coli. An enzyme which starts DNA chains." <u>J Biol Chem</u> **253**(3): 758-764.

Rowen, L. and A. Kornberg (1978). "A ribo-deoxyribonucleotide primer synthesized by primase." <u>J Biol Chem</u> **253**(3): 770-774.

Salzberg, S. L., A. J. Salzberg, et al. (1998). "Skewed oligomers and origins of replication." <u>Gene</u> **217**(1-2): 57-67.

Schekman, R., J. H. Weiner, et al. (1975). "Ten proteins required for conversion of phiX174 single-stranded DNA to duplex form in vitro. Resolution and reconstitution." <u>J Biol Chem</u> **250**(15): 5859-5865.

Scherzinger, E., E. Lanka, et al. (1977). "Role of bacteriophage T7 DNA primase in the initiation of DNA strand synthesis." <u>Nucleic Acids Res</u> **4**(12): 4151-4163.

Scherzinger, E., E. Lanka, et al. (1977). "Bacteriophage-T7-induced DNA-priming protein. A novel enzyme involved in DNA replication." <u>Eur J Biochem</u> **72**(3): 543-558.

Schmidt, B. H., A. B. Burgin, et al. (2010). "A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases." <u>Nature</u> **465**(7298): 641-644.

Schrodinger, L. (2010). "The PyMOL Molecular Graphics System, Version 1.3."

Segel, I. H. (1993). <u>Enzyme kinetics : behavior and analysis of rapid equilibrium and steady-state enzyme systems</u>. New York etc., John Wiley & Sons.

Silver, L. L. and N. G. Nossal (1979). "DNA replication by bacteriophage T4 proteins: role of the DNA-delay gene 61 in the chain-initiation reaction." <u>Cold Spring Harb Symp Quant Biol</u> **43 Pt 1**: 489-494.

Silver, L. L. and N. G. Nossal (1982). "Purification of bacteriophage T4 gene 61 protein. A

protein essential for synthesis of RNA primers in the T4 in vitro DNA replication system." J Biol Chem **257**(19): 11696-11705.

Snider, B. B. and J. Zhou (2006). "Synthesis of (+)-sch 642305 by a biomimetic transannular Michael reaction." Org Lett **8**(7): 1283-1286.

Spiering, M. M., S. W. Nelson, et al. (2008). "Repetitive lagging strand DNA synthesis by the bacteriophage T4 replisome." Mol Biosyst **4**(11): 1070-1074.

Srivatsan, A. and J. D. Wang (2008). "Control of bacterial transcription, translation and replication by (p)ppGpp." Curr Opin Microbiol **11**(2): 100-105.

Stein, R. A., J. C. Wilkinson, et al. (2001). "An analytical approach to the measurement of equilibrium binding constants: application to EGF binding to EGF receptors in intact cells measured by flow cytometry." Biochemistry **40**(20): 6142-6154.

Steitz, T. A. (2009). "The structural changes of T7 RNA polymerase from transcription initiation to elongation." Curr Opin Struct Biol **19**(6): 683-690.

Steitz, T. A., S. J. Smerdon, et al. (1994). "A unified polymerase mechanism for nonhomologous DNA and RNA polymerases." Science **266**(5193): 2022-2025.

Sun, W. and G. N. Godson (1993). "Binding and phasing of Escherichia coli single-stranded DNA-binding protein by the secondary structure of phage G4 origin of complementary DNA strand synthesis (G4oric)." J Biol Chem **268**(11): 8026-8039.

Sun, W., J. Schoneich, et al. (1999). "A mutant Escherichia coli primase defective in elongation of primer RNA chains." J Bacteriol **181**(12): 3761-3767.

Sun, W., J. Tormo, et al. (1994). "Domains of Escherichia coli primase: functional activity of a 47-kDa N-terminal proteolytic fragment." Proc Natl Acad Sci U S A **91**(24): 11462-11466.

Swart, J. R. and M. A. Griep (1993). "Primase from Escherichia coli primes single-stranded templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins. Template sequence requirements based on the bacteriophage G4 complementary strand origin and Okazaki fragment initiation sites." J Biol Chem **268**(17): 12970-12976.

Swart, J. R. and M. A. Griep (1995). "Primer synthesis kinetics by Escherichia coli primase on single-stranded DNA templates." Biochemistry **34**(49): 16097-16106.

Syson, K., J. Thirlway, et al. (2005). "Solution structure of the helicase-interaction domain of the primase DnaG: a model for helicase activation." Structure **13**(4): 609-616.

Tabor, S. and C. C. Richardson (1981). "Template recognition sequence for RNA primer synthesis by gene 4 protein of bacteriophage T7." Proc Natl Acad Sci U S A **78**(1): 205-209.

Tanner, N. A., S. M. Hamdan, et al. (2008). "Single-molecule studies of fork dynamics in Escherichia coli DNA replication." Nat Struct Mol Biol **15**(2): 170-176.

Thirlway, J. and P. Soultanas (2006). "In the Bacillus stearothermophilus DnaB-DnaG complex, the activities of the two proteins are modulated by distinct but overlapping networks of residues." J Bacteriol **188**(4): 1534-1539.

Thirlway, J., I. J. Turner, et al. (2004). "DnaG interacts with a linker region that joins the N-

and C-domains of DnaB and induces the formation of 3-fold symmetric rings." Nucleic Acids Res **32**(10): 2977-2986.

Thompson, A. A., R. A. Albertini, et al. (2007). "Stabilization of poliovirus polymerase by NTP binding and fingers-thumb interactions." J Mol Biol **366**(5): 1459-1474.

Tougu, K. and K. J. Marians (1996). "The extreme C terminus of primase is required for interaction with DnaB at the replication fork." J Biol Chem **271**(35): 21391-21397.

Tougu, K. and K. J. Marians (1996). "The interaction between helicase and primase sets the replication fork clock." J Biol Chem **271**(35): 21398-21405.

Tougu, K., H. Peng, et al. (1994). "Identification of a domain of Escherichia coli primase required for functional interaction with the DnaB helicase at the replication fork." J Biol Chem **269**(6): 4675-4682.

Traxler, M. F., S. M. Summers, et al. (2008). "The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli." Mol Microbiol **68**(5): 1128-1148.

Urlacher, T. M. and M. A. Griep (1995). "Magnesium acetate induces a conformational change in Escherichia coli primase." Biochemistry **34**(51): 16708-16714.

Valentine, A. M., F. T. Ishmael, et al. (2001). "A zinc ribbon protein in DNA replication: primer synthesis and macromolecular interactions by the bacteriophage T4 primase." Biochemistry **40**(50): 15074-15085.

van Oijen, A. M. and J. J. Loparo (2010). "Single-molecule studies of the replisome." Annu Rev Biophys **39**: 429-448.

VanLoock, M. S., Y. J. Chen, et al. (2001). "The primase active site is on the outside of the hexameric bacteriophage T7 gene 4 helicase-primase ring." J Mol Biol **311**(5): 951-956.

Vogt, S. L., C. Green, et al. (2011). "The stringent response is essential for Pseudomonas aeruginosa virulence in the rat lung agar bead and Drosophila melanogaster feeding models of infection." Infect Immun **79**(10): 4094-4104.

Wang, J. D., G. M. Sanders, et al. (2007). "Nutritional control of elongation of DNA replication by (p)ppGpp." Cell **128**(5): 865-875.

Waterhouse, A. M., J. B. Procter, et al. (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." Bioinformatics **25**(9): 1189-1191.

Wilson, E. M. and D. Trauner (2007). "Concise synthesis of the bacterial DNA primase inhibitor (+) -Sch 642305." Org Lett **9**(7): 1327-1329.

Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Wu, C. A., E. L. Zechner, et al. (1992). "Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size." J Biol Chem **267**(6): 4030-4044.

Wu, C. A., E. L. Zechner, et al. (1992). "Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. V. Primase action regulates the cycle of Okazaki

fragment synthesis." <u>J Biol Chem</u> **267**(6): 4074-4083.

Yang, W., J. Y. Lee, et al. (2006). "Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity." <u>Mol Cell</u> **22**(1): 5-13.

Yuzhakov, A., Z. Kelman, et al. (1999). "Trading places on DNA--a three-point switch underlies primer handoff from primase to the replicative DNA polymerase." <u>Cell</u> **96**(1): 153-163.

Zechner, E. L., C. A. Wu, et al. (1992). "Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size." <u>J Biol Chem</u> **267**(6): 4045-4053.

Zhu, B., S. J. Lee, et al. (2010). "Direct role for the RNA polymerase domain of T7 primase in primer delivery." <u>Proc Natl Acad Sci U S A</u> **107**(20): 9099-9104.

Appendix I – Design and implementation of fluorometric *de novo* primer synthesis assays

DEVELOPMENT OF A POLYMERASE COUPLED FLUOROMETRIC PRIMER SYNTHESIS ASSAY

To develop an improved microplate based and highly sensitive primer synthesis assay, we chose to develop a helicase-stimulated and polymerase coupled version of the published fluorometric primer synthesis assay using a double-strand specific dye, PicoGreen (**Fig. AI.1**, {Dragan, 2010 #1617}). Such an approach provides high sensitivity and low material consumption, without the need for any specialized equipment beyond a microplate fluorometer. Primer extension is the actual readout in this approach, which is presumably entirely dependent on primer synthesis, but we nonetheless conducted a series of experiments designed to demonstrate the feasibility of this approach. In our first proof-of-principle experiment, we measured the fluorescent signal sensitivity to primer synthesis by DnaG. As shown in **Fig. AI.2a**, an increase in fluorescence is only observed with wild-type DnaG and not with a mutant of DnaG with an alanine substitution to Asp269 that renders the enzyme incapable of primer synthesis (Godson, Schoenich et al. 2000). Primer synthesis was measurable at DnaG concentrations as low as 10 nM with 900 nM DnaB, and these results indicated that a protein concentration of 100 nM DnaG provided a robust signal. We performed this experiment in a number of plate formats (not shown), and ultimately settled on a flat-bottom, half-area 96-well black plate (Corning). A titration of the polymerase indicated that a per-reaction concentration of 0.1 U provided ample signal amplification (**not shown**). We also searched for an ideal helicase concentration, and found that an equimolar concentration of helicase produced the most stimulation (**not shown**).

We next endeavored to establish whether or not the primer extension measured in this assay provided an unbiased readout of primer synthesis by DnaG, as well as to develop a means of determining if an inhibitor is specific to DnaG, or if it also interferes with DNA synthesis by Klenow. To this end, we examined the kinetic characteristics of primer synthesis by DnaG reported by this assay. A time course of the reaction with 100 nM DnaG and 100 nM DnaB (**Fig. AI.1b**) revealed a complex kinetic profile with substantial sigmoidicity; a common feature for coupled reactions where at least one of the substrates is non-saturating {Segel, 1993 #1511}. Despite these complex kinetics, analysis of the dependence of primer synthesis on NTP (**Fig. AI.1c**) returns an average K_m for ribonucleotides of ~40 μM, a value in good agreement with those reported previously (Swart and Griep 1995). A titration of the m13 ssDNA template (**Fig. AI.1d**) displays more pronounced substrate inhibition at high concentrations than what has been observed previously (Johnson, Bhattacharyya et al. 2000). We fit the data to a canonical substrate inhibition model (Copeland 2005) to address the possibility that measuring primer extension as a proxy for primer synthesis had altered the apparent template dependent kinetics for helicase stimulated primer synthesis by DnaG. The K_m obtained from this fit (10 nM) is in good agreement with the previously published K_m of the primase-helicase complex for a ssDNA template (13±2 nm (Johnson, Bhattacharyya et al. 2000)). Finally, we addressed the possibility that inhibitors could in-

terfere with the action of the DNA polymerase by investigating the inhibition of primase by the nucleotide analogue cordycepin triphosphate (3′-deoxy-ATP), which is a specific inhibitor of several RNA polymerases {Kuchta, 2010 #1529}. A titration of 3DA into reactions containing primase and the polymerase (**Fig. AI.1e**, black squares), or reactions containing performed primers and the polymerase (**Fig. AI.1e**, gray circles) reveals a pattern of inhibition highly indicative of an inhibitor specific to RNA polymerases (primase in this instance). Thus the dependencies of the reaction on the substrates of DnaG are largely unaltered, and measuring primer extension by Klenow does provide an unbiased readout on primer synthesis.

IMPLEMENTATION OF AN UNCOUPLED FLUOROMETRIC PRIMER SYNTHESIS ASSAY

Despite the utility of this approach, subsequent efforts (**Chapter 3**) highlighted a need for an uncoupled method. As developing this assay would require retitrating all reaction components, we began by looking for an ideal m13 ssDNA template concentration using an elevated concentration of DnaG/DnaB (500 nM). This experiment showed relatively robust signal (although far reduced from the same experiment with polymerase coupling), and demonstrated that relatively few condition changes were necessary. Additional PicoGreen was used initially to stain reactions, but we ultimately found this was unnecessary, and that the standard 1:200 final stain dilution does not result in signal saturation.

In the same vein, we also exchanged the m13 ssDNA for a short oligonucleotide template used in the standard primer synthesis assay. We found that using a 3′-phosphorylated template avoids so-called "overlong" products, but only marginally increases the cost of the template. A titration with this template showed that a template concentration of 600 nM provided robust signal without substrate inhibition. This final method proved to be extremely versatile. It can be run at virtually any volume, and tolerates inclusion of a surprisingly high concentration of other solvents (e. g., MeOH, EtOH, DMSO, and IPA). For experiments in which reagents are not limiting, a 50 μL reaction volume and a 100 μL final volume is convenient, and tends to generate narrower error bars. For experiments in which reagents are limiting, such as with the (p)ppGpp titrations (**Chapter 4**), a 20 μL reaction volume and a 30 μL final volume is more appropriate. To perform an experiment with such small volumes, a specialized 384-well plate is used (BioRad HSP-3866). This is technically a PCR plate. It has a polysterene skirt and polypropylene concical-bottom wells that are raised above the skirt to allow the plate to be sealed.

Finally, many different buffers were tried for all of the fluorometric *de novo* primer synthesis assays described here. A standard reaction buffer was identified that worked for all of these assays, and functioned in the radiolabeled- and fluorescent-nucleotide incorporation assays as well. It works equally well as a binding buffer, and if the BSA is replaced with glycerol to 10%, functions as a low salt size exclusion buffer, though in general DTT and Mg-acetate (see below) were not included in SEC buffer versions.

Standard procedure for a helicase-stimulated *de novo* primer synthesis assay, coupled or uncoupled

MATERIALS

All components were brought up in Rnase free water (purchased, or DMPC treated), or were purchased as certified Rnase/Dnase free.

Individual solutions of ATP, CTP, GTP, UTP, dATP, dCTP, dGTP, and dTTP at 100 mM in TE. dNTP solutions were typically purchased. rNTP solutions were made from powder, and brought up in TE. The pH was adjusted to ~ 7.5 with 100 mM NaOH.

> 5x Reaction Buffer:
> 500 mM potassium glutamate (stock should be @ pH 7.5)
> 100 mM HEPES pH 7.5
> ≥50 mM MgOAc$_2$
> 5 mM DTT
> 1 mg/mL BSA

1 mM template of various designs, though the "standard" template had the following sequence:
5´-*CACACACACACACA***CTG**AAAGCCAAAAG-3´-phosphate (CTGCA$_{16}$P)

The constant region is underlined. The start site is bolded, and the variable region is in italics. An appropriate template design consists of a 3´-constant region, a start site, and a variable region of some dinucleotide repeat, such as *AT* or *GA*. Avoid *GC* repeats.

m13 ssDNA was purchased from USB corp. as a 1 µg/µL solution

Klenow exo⁻ was purchased from epicenter biotechnologies as a 50 U/µL solution.

DnaB was obtained from Jacob E. Corn Ph. D. as a 96 µM (hexamer) stock solution in 500 mM NaCl, 20 mM HEPES pH 7.5, 10% glycerol.

DnaG was purified as described (**Chapter 2**), and was stored in 100 mM potassium glutamate, 20 mM HEPES pH 7.5, 10% glycerol at 500 µM.

PicoGreen was purchased from Invitrogen

50 mM EDTA, 50 mM HEPES pH 7.5 (stop solution)

A 96-well, 200 µL/well PCR plate

A reaction plate (see above)

METHODS

A GTP titration for the uncoupled, low-volume assay will be used as an example.

1. *Prepare "constant solution" mix, which will contain all constant components*:
        Combine in the following order:
        1 μL Full Length –wild type DnaG
        5 μL DnaB
        0.6 μL template
        1 μL of each of ATP and UTP
        96 μL 5x buffer (add slowly)
        Water to 480 μL (376 μL – best to add in 2-3 steps and slowly)
(NOTE: almost never were these volumes actually pipetted, usually multiple experiments were conducted at once, so some positive integer multiple of each value was actually pipetted)
        -Mix thoroughly, but gently by slow pipetting.
        -Distribute 35 μL into each well of a row of the PCR plate. Incubate at RT for at least 10 min.
        -If running a polymerase-coupled assay, it should be included here at 0.175 U/rxn.
2. *In the mean time, prepare GTP dilutions* (For a 2 mM – 62.5 nM 1:1 titration):
        -Combine:
        3.2 μL GTP
        16 μL 5x buffer
        60.8 μL water
        -Serial dilute 35 μL of solution into 35 μL of 1x buffer 10 times in the same PCR plate used earlier. First well is undiluted, last well has no GTP.
3. *Assemble reactions*:
        -Using a 12-channel 20 μL pipettor, distribute 10 μL of GTP solution into each well of 3 rows of a Bio-Rad HSP-3866 PCR plate, there is no need to switch tips unless severe bubbles are introduced.
        -Switch tips. Distribute 10 μL of constant solution into each of these wells, mix by pipetting 3 times. There is no need to switch tips between rows.
4. *Seal and Spin plate*:
        -Seal plate with a transparent, pressure sensitive plate seal, do not use aluminum or sticky seals, and do not thermoseal.
        -Spin plate at 3500 RPM in Sorvall Table Top centrifuge for at least 1 min.
5. *Incubate at 37 ℃ for 1 h.*
6. *In the mean time, prepare stop/stain solution*:
        -Dilute 6 μL PicoGreen solution into 414 μL of stop solution.
        -Distribute 35 μL to each well of a row of the same PCR plate.
        -Cover, and keep in dark until ready to stop reaction.
7. *Stop/stain reaction*:
        -Remove reaction from incubator, and remove seal.
        -Distribute 10 μL of stop/stain solution into each well using a 12-channel pipettor. It is a good idea to switch tips between rows. Mix by pipetting 3 times for each row.

8. *Reseal and incubate plate in dark for 5 min*.
9. *Measure fluorescence*:
> -Used a PerkinElmer Victor 3V multilabel plate reader and the following protocol:
>> • Shake plate for 2 s
>> • Wait 20 s
>> • Excitation filter: 485 nm ± 14 nm
>> • Emission filter: 535 nm ± 25 nm
>> • Lamp set to "Stabilized Energy" at level 12,032.
>> • Counting time of 1 s
>> • 4 mm measurement height

10. *Data processing:*
> -Export data to an Excel file.
> -Open it up in MS Excel, and go to "plate" tab.
> -Remove unused columns.
> -To calculate the background corrected average of each data point:
>> •Enter the following formula (R stands for Replicate, S for Sample, and B for Background, so SR1 corresponds to replicate 1 of a sample reaction, whereas BR3 corresponds replicate 3 of a background reaction):
>> = AVERAGE(SR1:SR3) – AVERAGE($BR$1:$BR$3)
>> •Copy this formula to the remaining data points.
> -To calculate the background corrected error of each data point:
>> •Enter the following formula to obtain the standard error of the mean
>> = (SQRT(STDEV(SR1:SR3)^2 + STDEV($BR$1:$BR$3)^2))/SQRT(3)
>> •Copy this formula to the remaining data points.
> -Copy resulting background corrected values into Kaleidagraph, along with the titration series concentrations.
> -Plot as X-Y scatter, and fit to Eqn. 2.1.
> -Copy out fitted m1 value (V_{max}) back to Excel.
> -Divide all average, background corrected values by this number to obtain Relative Primer Synthesis. This includes errors and all the mutants tested, as everything within a single experiment must be normalized to a single wild-type control!
> -Copy new values (including error) back into Kaleidagraph, and refit with Eqn 2.1, but m1 starting value should now be 1.


FINAL REMARKS

Design of background corrections appropriately
PicoGreen stains everything, but only double-stranded DNA and RNA•DNA heteroduplexes result in an increase in fluorescence. Thus, an appropriate background reaction takes into account available ssDNA. A template titration, for example, requires a point-by-point background titration series, as each point of the titration has a different amount of ssDNA. Moreover, it is possible to deplete the PicoGreen during staining, resulting in aberrant signal saturation. It is therefore necessary to determine empirically the amount of PicoGreen required to obtain a nonsaturating signal with the highest concentration data point. This amount should be used for all other points in the titration.

Quantification

It is theoretically possible to calculate the concentration of primers synthesized by running a standard curve. This requires using a product mimic that recapitulates as closely as possible the actual reaction product. For the single-stranded oligo templates, this requires a titration of a corresponding heteroduplex with a 3´-overhang into a solution containing everything but DnaG. The fluorescence intensity of actual reactions can then be converted to moles of primers synthesized using the standard curve. This method was not commonly employed here, but it was tried, and it does work.
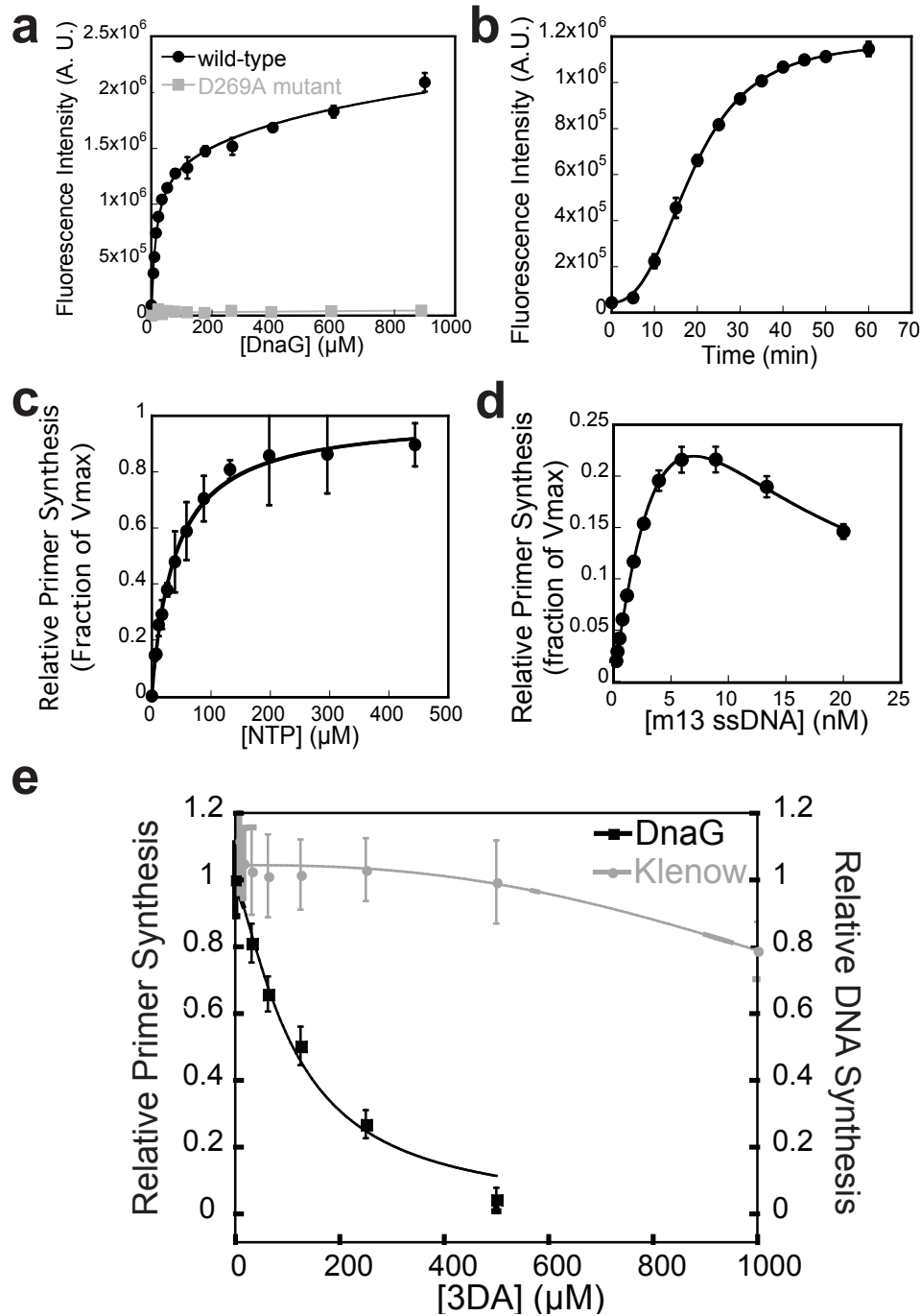
Figure AI.1: Development of a Fluorometric, Helicase-Stimulated, Polymerase-Coupled primer synthesis assay.

a) Specificity of staining for synthesis by DnaG. Concentration dependent primer extension signal (as determined by staining with PicoGreen) in the presence of 900 nM DnaB with wild-type DnaG (black circles) and DnaG containing a Asp269Ala substitution (gray squares) that disrupts catalysis by the enzyme. The trend in the increase in fluorescence intensity is illustrated by a smoothed fit, and is intended for visualization purposes only. A. U., arbitrary units. Error bars represent ±SD (**Methods**). Continued on next page.

b) Primer synthesis by DnaG over time in the HSPC primer synthesis assay. The plot shows the background corrected, time-dependent increase in fluorescence from reactions containing 100 nM DnaG and 100 nM DnaB. A. U., arbitrary units. Error bars represent ±SD (**Methods**). c) NTP dependent activity of DnaG in the HSPC primer synthesis assay. All four NTPs were titrated into reactions containing 100 nM DnaG, 100 nM DnaB, and 100 μM of all 4 dNTPs. Data are fit to a standard Michaelis-Menten enzyme kinetics model (**Methods**), yielding an average Km for NTPs of 50 μM. Error bars represent ±SD (**Methods**). d) Template dependent activity of DnaG in the HSPC primer synthesis assay. M13 ssDNA was titrated into reactions containing 100 nM DnaG, 100 nM DnaB. Data are fit to a substrate-inhibition model (**Eqn. 3.2**, **Methods**). Error bars represent ±SD (**Methods**). e) Impact of an RNA polymerase specific inhibitor, 3'-dATP, on primer synthesis by DnaG, and DNA synthesis by Klenow in the HSPC primer synthesis assay. 3DA was titrated from 30 μM to 0.5 mM into reactions containing 100 nM DnaG and 100 nM DnaB (black circles) or from 8 μM to 1 mM into reactions containing 10 nM of an m13 complementary DNA primer and no DnaG (gray circles). Curves are fit to an IC_{50} model (**Methods**). Error bars represent ±SD (**Methods**). Synthesis is reported as a fraction of the synthesis extent for the 0 μM 3DA data point.

Appendix II – Attempts to develop a fluorescent nucleotide binding assay for DnaG

INTRODUCTION

Nucleotide binding is a distinct kinetic step in nucleic acid polymerization. As such, most polymerases possess a suite of residues that specifically interact with individual nucleoside triphosphates. During our work to determine the structural basis of nucleotide binding by DnaG (**Chapter 2**), we attempted to develop a fluorescent nucleotide-binding assay for DnaG. Initially this work seemed to be quite successful, but subsequent data clouded the validity of these results. Here, this work is summarized, and detailed protocols are described.

A series of naturally fluorescent and fluorescent-labeled nucleotide analogues were considered for developing a nucleotide-binding assay for the *E. coli* DnaG RPD. Of these, Fluorescein-12-ATP (**Fig. AII.1a**, Fl.12.ATP) and adenosine 5'-O-(3-thiotriphosphate)-BODIPY-fluorescein-thioester (**Fig. AII.1d**, 5'Fl.ATP) were selected based on availability, compatibility with available instrumentation, and unit price. The *E. coli* RPD bound to both of these analogues, with Fl.12.ATP displaying the greatest affinity ($\sim$40 μM) and 5'Fl.ATP the worst ($\sim$500 μM) in the presence of Mg^{2+}.

EXPERIMENTS WITH FLUORESCEIN-12-ATP

Several key residues were identified for involvement in nucleotide binding (**Chapters 2,5**). These were tested for their ability to bind to Fl.12.ATP in the presence of 20 mM Mg^{2+} (**Fig. AII.1b**). Several strong effects on nucleotide binding were observed, such as with Arg146Ala (gold squares) and Arg221Ala (cyan diamonds), as well as more modest reductions in affinity, as with Asn232Ala (magenta inverted-triangles) and Lys229Ala (black triangles). Substituting glutamine for alanine at position 313 in *E. coli* only modestly reduced affinity. All of these results are consistent with the interactions observed in the nucleotide bound crystal structures of the *Sa*DnaG RPD. However, many other results were not. For example, disrupting metal binding by replacing Asp269 with alanine also only modestly reduces affinity, which is in direct contradition with the results obtained for the basic ridge residues. Additionally, substituting either Gly266 or Gly286 with alanine likewise does not impact nucleotide binding materially, which does not agree with the effect caused by mutating Asn232.

To address these discrepancies, we investigated the metal dependence of binding, and the stoichiometry of binding. We observed that omission of Mg^{2+} from the reactions had no effect on the binding isotherms for any of the mutant proteins discussed above (not shown). A titration of the RPD into reactions containing free fluorescein (no nucleotide) showed only a linear increase consistent with an increase in viscosity caused by the high concentrations of protein utilized (not shown). Correction of the binding isotherms with these data (which can be thought of as background) does not impact the apparent dissociation constants, or the overall shape of the curves (not shown). As this experiment seemed to rule

out the possibility of a nonspecific interaction with fluorescein, we next investigated a non-specific interaction with the whole labeled nucleotide. To do so, we examined the ability of unlabeled ATP to compete for binding to the labeled ATP. As can be seen in **Fig. AII.1c**, ATP does compete for binding, and this competition is Mg^{2+}-independent. Attempts to determine the stoichiometry of binding were unsuccessful (not shown), as were efforts to eliminate binding through the introduction of severe or multiple mutations (not shown). Finally, we established that Fl.12.ATP can be incorporated into a primer in a sequence specific manner (**Chapter 6**), and must therefore be binding to the incoming nucleotide-binding site at certain times. Additional efforts to establish conditions under which Fl.12.ATP binds competently to the RPD were not fruitful, and we eventually ceased work on developing and understanding the binding of Fl.12.ATP to the RPD.

EXPERIMENTS WITH ADENOSINE 5'-O-(3-THIOTRIPHOSPHATE)-BODIPY-FLUORESCE-IN-THIOESTER

We used 5'Fl.ATP to investigate whether (p)ppGpp compete for binding to the incoming nucleotide binding site. In a control experiment, we found that the *Ec*DnaG RPD associated with 5'Fl.ATP with or without Mg^{2+} with a Kd-app of 500 μM, although the use of Mn^{2+} results in approximately 5-fold tighter binding (**Fig. AII.1e**). We next assayed for the ability of (p)ppGpp and GDP to displace the labeled nucleotide in the presence of Mn^{2+}. As can be seen in **Fig. AII.1f**, all three inhibitors displace the nucleotide from the RPD in a concentration dependent manner, with G5P being the most, and GDP the least potent, a result consistent with dose-response data obtained in *de novo* primer synthesis experiments. No additional experiments were conducted using 5'Fl.ATP.

Detailed protocols for nucleotide binding experiments

MATERIALS

The *E. coli* DnaG RPD and mutants thereof were cloned, expressed, and purified as described in **Chapter 2, Methods**, and stored at ≥2 mM.

Fluorescein-12-ATP was purchased from PerkinElmer as a 1 mM solution in TE. The purchased stock was aliquoted and stored at -20 ˚C. Individual aliquots were thawed, and diluted to 10 µM, realiquited, flash frozen, and stored at -80 ˚C until just prior to use.

Adenosine 5'-O-(3-thiotriphosphate)-BODIPY-fluorescein-thioester was purchased from Invitrogen as a 1 mM stock solution in TE. The purchased stock was aliquoted and stored at -20 ˚C. Individual aliquots were thawed, and diluted to 10 µM, realiquited, flash frozen, and stored at -80 ˚C until just prior to use.

Inhibitor solution in TE at a concentration ~ 20 fold above the expected IC_{50}.

5x Reaction Buffer:
      500 mM potassium glutamate (stock should be @ pH 7.5)
      100 mM HEPES pH 7.5
      ≥50 mM $MgOAc_2$
      5 mM DTT
      1 mg/mL BSA

E. coli DnaG RPD sizing buffer:
      100 mM potassium glutamate, pH 7.5
      20 mM HEPES, pH 7.5
      10% glycerol

A 96-well, 200 µL/well PCR plate (prep plate)

A BioRad HSP-3866 plate (reaction plate)

METHODS

PROTOCOL FOR A STANDARD PROTEIN TITRATION
1. *Prepare "constant solution" mix, which will contain all constant components, total volume is 420 µL*:
      -Combine in the following order:
       4.2 µL labeled nucleotide
       96 µL 5x buffer (final concentration will be 0.5x, with the exception of potassium glutamate)
       Water to 420 µL (320 µL)
      -Mix thoroughly, but gently by slow pipetting.

-Distribute 35 µL into each well of a row of the prep plate.

2. *Prepare protein dilutions* (For a 2 mM – 62.5 nM 1:1 titration):

-In a single row of the prep plate, serial dilute 35 µL of concentrated protein solution (wild type was stored at 4 mM) into 35 µL of sizing buffer 10 times. First well is undiluted, last well has no protein.

3. *Assemble reactions*:

-Using a 12-channel 10 or 20 µL pipettor, distribute 10 µL of constant solution into each well of 3 rows of a reaction plate, there is no need to switch tips unless severe bubbles are introduced.

-Switch tips. Distribute 10 µL of protein dilution into each of these wells, mix by pipetting 3 times. There is no need to switch tips between rows.

4. *Seal and Spin plate*:

-Seal plate with a transparent, pressure sensitive plate seal, do not use aluminum or sticky seals, and do not thermoseal.

-Spin plate at 3500 RPM in Sorvall Table Top centrifuge for at least 1 min.

5. *Incubate at room temperature for 30 min.*

9. *Measure fluorescence polarization*:

-Used a PerkinElmer Victor 3V multilabel plate reader and the following protocol:

- Excitation filter: 485 nm ± 14 nm
- Emission filter: 535 nm ± 25 nm
- Lamp set to maximum stable energy level
- Counting time of 0.1 s
- 4 mm measurement height

10. *Data processing*

-Export data to an Excel file.

-Open it up in MS Excel, and go to "plate" tab.

-Remove unused columns.

-Convert all data points from FP to FA using Eqn. AII.1.

Equation AII.1 FA =

-Average the FA three reactions for each concentration.

-Convert to ΔFA by subtracting the no protein control from each concentration point:

•Enter the following formula for each concentration (R stands for Replicate, S for Sample, and B for Background, so SR1 corresponds to replicate 1 of a sample reaction, whereas BR3 corresponds replicate 3 of a background reaction):

= AVERAGE(SR1:SR3) – AVERAGE($BR$1:$BR$3)

-Calculate the standard deviation for each concentration.

•Enter the following formula for each concentration

= SQRT(STDEV(SR1:SR3)^2 + STDEV($BR$1:$BR$3)^2)

-Copy ΔFA values, errors, and protein concentrations into Kaleidagraph.

-Plot as X-Y scatter, and fit to Eqn. 2.2.

PROTOCOL FOR A COMPETITION EXPERIMENT

1. *Prepare "constant solution" mix, which will contain all constant components, total volume is 420 µL*:

      -Combine in the following order:

       52.5 µL DnaG @ 4 mM

       4.2 µL labeled nucleotide

       168 µL 5x buffer (final concentration will be 0.5x, with the exception of potassium glutamate)

       1 M MgOAc must be added to 2x the maximum inhibitor concentration

       Water to 420 µL (320 µL)

      -Mix thoroughly, but gently by slow pipetting.

      -Distribute 35 µL into elevens wells of a row of the prep plate.

2. *Prepare competitor dilutions*:

      -In a single row of the prep plate, serial dilute 35 µL of concentrated inhibitor solution (ATP was used at 100 mM for example) into 35 µL of TE 9 times. First well is undiluted, last two wells have no inhibitor.

3. *Assemble reactions*:

      -Using a 12-channel 10 or 20 µL pipettor, distribute 10 µL of constant solution into eleven wells of 3 rows of a reaction plate, there is no need to switch tips unless severe bubbles are introduced.

      -Prepare a control reaction with no protein: 1 µL labeled nucleotide, 40 µL 5x buffer, water to 100 µL (59 µL). Distribution 10 µL into the last 3 wells.

      -Switch tips. Distribute 10 µL of inhibitor dilution into each well, mix by pipetting 3 times. There is no need to switch tips between rows.

4. *Seal and Spin plate*:

      -Seal plate with a transparent, pressure sensitive plate seal, do not use aluminum or sticky seals, and do not thermoseal.

      -Spin plate at 3500 RPM in Sorvall Table Top centrifuge for at least 1 min.

5. *Incubate at room temperature for 30 min.*

9. *Measure fluorescence polarization*:

      -Used a PerkinElmer Victor 3V multilabel plate reader and the following protocol:

          • Excitation filter: 485 nm ± 14 nm

          • Emission filter: 535 nm ± 25 nm

          • Lamp set to maximum stable energy level

          • Counting time of 0.1 s

          • 4 mm measurement height

10. *Data processing*:

      -Export data to an Excel file.

      -Open it up in MS Excel, and go to "plate" tab.

      -Remove unused columns.

      -Convert all data points from FP to FA using Eqn. AII.1.

          Equation AII.1 FA =

      -Average the FA three reactions for each concentration.

      -Convert to ΔFA by subtracting the no protein control from each concentration point:

          •Enter the following formula for each concentration (R stands for Replicate,

S for Sample, and B for Background, so SR1 corresponds to replicate 1 of a sample reaction, whereas BR3 corresponds replicate 3 of a background reaction):

= AVERAGE(SR1:SR3) – AVERAGE($BR$1:$BR$3)

-Calculate the standard deviation for each concentration.

•Enter the following formula for each concentration

= SQRT(STDEV(SR1:SR3)^2 + STDEV($BR$1:$BR$3)^2)

-Copy ΔFA values, errors, and inhibitor concentrations into Kaleidagraph.

-Plot as X-Y scatter, and fit to Eqn. 4.1.

FINAL REMARKS

A method for directly studying the binding of DnaG to its substrates is required to fully understand how DnaG synthesizes a primer. The experiments presented here reprsent a first step toward developing an assay for measuring nucleotide binding to DnaG. Although many caveats accompany the results of the nucleotide binding methods employed in the experiments described above, the fact that a crystal structure was captured of the DnaG RPD bound to an individual nucleotide suggests that the principle of the approach is sound. In the future, intrinsically fluorescent nucleotide analogues should be employed exclusively, and initial binding screens should be conducted in the presence of Mn^{2+} or 100 mM or greater Mg^{2+}.
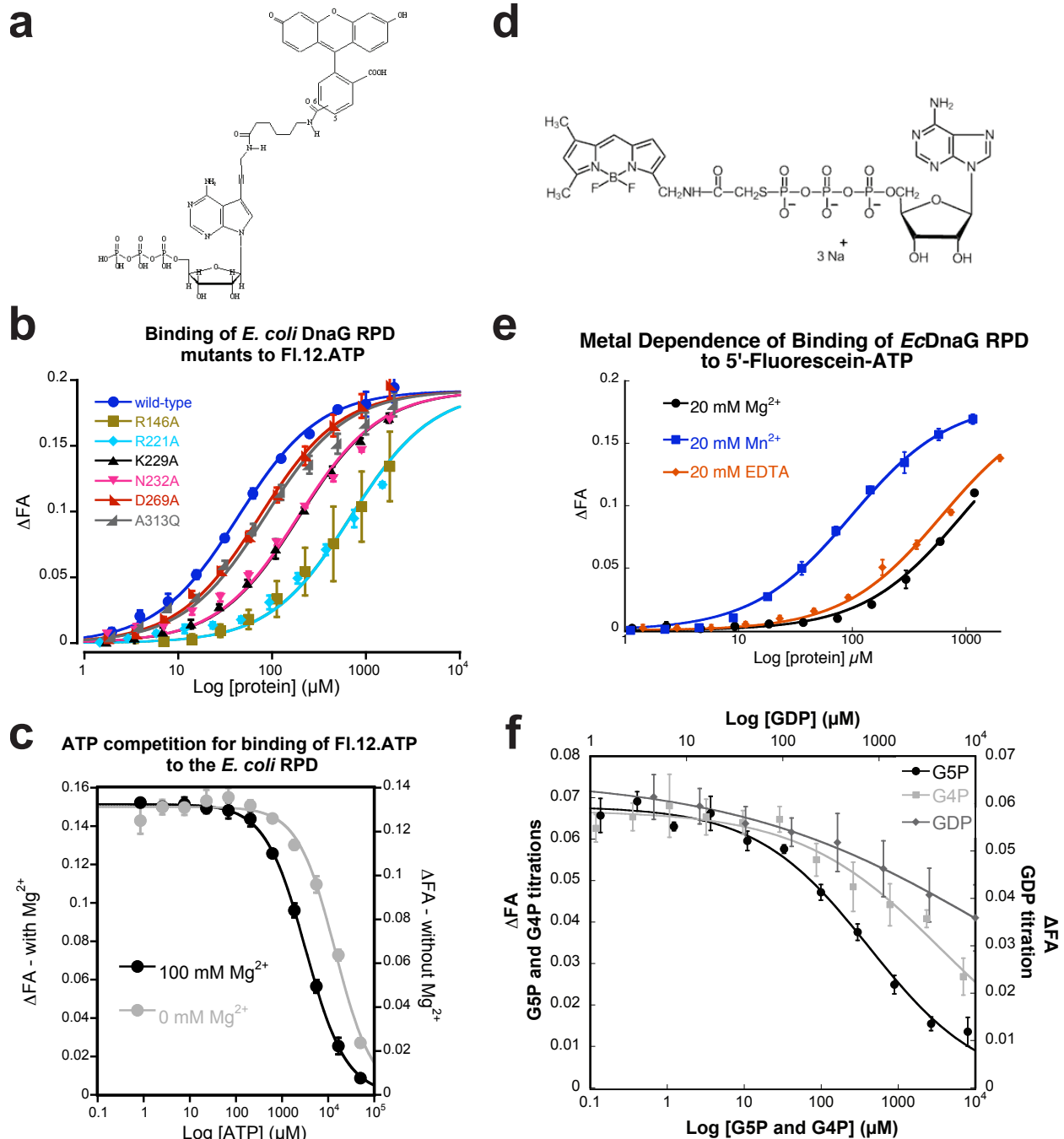
Figure AII.1: Fluorescently labeled nucleotide binding by the *Ec*DnaG RPD.

a) Structure of base-conjugated fluorescently labeled ATP analogue, Fl.12.ATP. b) Change in Fluorescence Anisotropy (ΔFA) of Fl.12.ATP in the presence of increasing concentrations of wild-type and mutant *Ec*DnaG RPDs to Fl.12.ATP. Curves were fit to the explicit solution to single-site binding (**Methds**). Error bars represent ±SD c) ΔFA of Fl.12.ATP in the presence of 250 μM wild-type *Ec*DnaG RPD and increasing coentrations of cold (unlabled) ATP with (black circles), or without (gray circles) of Mg²⁺. Curves were fit to a standard IC_{50} model (**Methods**). Error bars represent ±SD d) Structure of the γ-phosphate-conjugated fluorescently labeled ATP analogue, 5'Fl.ATP. e) Metal dependence of the binding of the *Ec*DnaG

RPD to 5'Fl.ATP. *Ec*DnaG RPD was titrated into reactions containing the labeled nucleotide, and 20 mM Mg^{2+} (blacks circles), 20 mM EDTA (orange diamonds), or 20 mM $MnCl_2$ (blue squares), and binding was determined by measuring the ΔFA of 5'Fl.ATP. Curves were fit to the explicit solution to single-site binding (**Methods**), and were assumed to reach the same plateau value, which was taken from the Mn^{2+} curve fit. Error bars represent ±SD. f) Competition for binding of the *Ec*DnaG RPD to 5'Fl.ATP by G5P (dark-gray circles), G4P (black squares), and GDP (light-gray diamonds). Inhibitors were titrated into reactions containing the fluorescently labeled ATP, and 250 µM RPD. Binding extent was determined by measuring the ΔFA of the labeled ATP. Curves were fit to a standard IC_{50} model (**Methods**). Error bars represent ±SD.

Appendex III – Synthesis and purification of catalytically competent hairpin-crosslinked DnaG RPD complexes

The most immediate need in the study of DnaG enzymology is the crystal structure of a DnaG-type primase stalled in a true ternary complex. A line of research we do not discuss here, but which we spent a great deal of time on, involved an attempt to capture an RPD ternary complex using an engineered disulfide crosslinking site strategy. The crosslinking site used in this effort was identified by Shellie Weisfield and Jacon Corn in 2006 by screening crosslinking mutants for their ability to extend a preformed, crosslinkable hairpin. The most capable of these sites, corresponding to position 166 in *E. coli* DnaG, was located on the extreme edge of the template-tracking groove (**Fig. AIII.1a**). We attempted to capture a crystal structure of the ternary complex using a series of hairpin substrates exemplified by the hairpin shown in **Fig. AIII.1b**. We were able to successfully form and purify (**Fig. AIII.1c/d**) these complexes, while at the same time preserving their catalytic abilities (**Fig. AIII.1e**). However, no diffracting crystals, and only a few initial crystal hits were obtained.

After discovery of the incoming nucleotide-binding site (**Chapter 2**), we redesigned the substrate hairpins to be much shorter (**Fig. AIII.1f**). I was again able to form and purify these new complexes (**Fig. AIII.1g/h**), and we demonstrated that they are also catalytically competent (**Fig. AIII.1i**). Again, no diffracting crystals were obtained, and very few hits. The new complexes were also considerably more difficult to purify, and were less stable.

Two factors likely hampered the progress of this project. First, the extreme solvent exposure of the position-166 crosslinking site made it excessively susceptible to reduction, leading dissociation of the complex, and severely interfering with the purity of the crosslinked complexes. Second, more recent modeling suggests that this crosslinking position pulls the substrate into an unfavorable conformation, leading to, we think, constant undocking of the substrate from the enzyme. In support of this second claim, we found that the position corresponding to Asp166 in *S. aureus* DnaG (His166), does not crosslink to any of the crosslinkable hairpins used successfully with *E. coli*. Notably, *Sa*DnaG lacks Trp165, which may be required to position the 5´-end of the hairpin near the position-166 crosslinking site. Based on this analysis, we propose a modified strategy for obtaining a ternary complex with the DnaG RPD. First, *S. aureus* DnaG should be used instead of *E. coli*, due to the fact that crystallization of the *Sa*DnaG RPD can be driven by benzamidine. Also, new crosslinking positions are needed. To this end, crosslinking site modeled to a more central location on the RPD (**Fig. AIII.1j**). These positions will consistently be across from the incoming nucleotide binding site, and therefore are less likely to distort substrate binding. However, they necessitate using redesigned substrates with internal crosslinking sites, in which the convertible base is placed in the template strand 1 nucleotide 5´ to the template base that pairs with the incoming nucleotide.

Standard protocol for the production and purification of DnaG RPD hairpin-crosslinked complexes

MATERIALS

Convertible oligonucleotides were purchased from the Midland Reagent Company, and prepared according to the protocols described in Jacob E. Corn's doctoral thesis (2007).

E. coli DnaG RPD D166C was generated by site directed mutagenesis, expressed, and purified as described in **Chapter 2**.

Crosslinking Buffer:

>50 mM potassium glutamate pH 8.5

>20 mM HEPES pH 8.5

>10% Glycerol

1 M Cystamine, prepared according to the protocol described in Jacob E. Corn's doctoral thesis (2007).

Purification Buffers:
>Low Salt Buffer
>50 mM sodium chloride
>20 mM HEPES pH 7.5
>5 mM oxidized glutathione
>10% glycerol

>High Salt Buffer
>1 M sodium chloride
>20 mM HEPES pH 7.5
>5 mM oxidized glutathione
>10% glycerol

>Size Exlcusion Buffer
>100 mM potassium glutamate pH 7.5
>20 mM HEPES pH 7.5
>5 mM ozidized glutathione
>10% glycerol

12 mL Uno Q column (Bio-Rad)

GE HiPrep 16/60 Sephacryl S-200 column

METHODS

1. Mix 1 μmol of EcDnaG RPD D166C with 500 nmol of converted oligonucleotide. Dilute > 10x in crosslinking buffer. Incubate on ice for 1-2 days.

2. Follow reaction progress by removing 9 μL aliquots and quenching with 1 μL of MMTS. Dilute samples into non-reducing SDS loading buffer, and analyze by SDS PAGE (12% gels work well).

3. When crosslinking has reached completion (corresponding to a 1:1 ratio of crosslinked protein to uncrosslinked protein), concentrate reaction to 1 mL, and adjust pH to 7.5.

4. Load sample onto Uno Q column equilibrated in Low Salt buffer.

5. Wash with 1 CV of Low Salt buffer, or until free protein has finished eluting. Free protein is easily distinguished by high A230, and an A260:A280 ratio of ~0.5.

6. Elute complex, dimer, and free DNA with a gradient from Low Salt buffer to High Salt buffer over 10 CVs. Collect fractions throughout.

7. Complex can be distinguished from free DNA as a peak with high A230, and an A260:A280 ratio of 1.5 or higher.

8. Combine all non-DNA fractions, and concentrate to 1 mL.

9. Rerun Uno Q column.

10. Combine fractions containing complex, and concentrate to 1 mL.

11. Load onto S200 equilibrated 1x with water, and 2x with Size Exclusion buffer. This column ideally is never used with reducing buffers.

12. Run column. Identify fractions containing complex, and no dimer by non-reducing SDS-PAGE. Combine, and concentrate to 20 mg/mL or higher. Yields are normally in the 5-10% of input DNA range.

13. Complexes can be stored in Size Exclusion buffer at -80 °C.
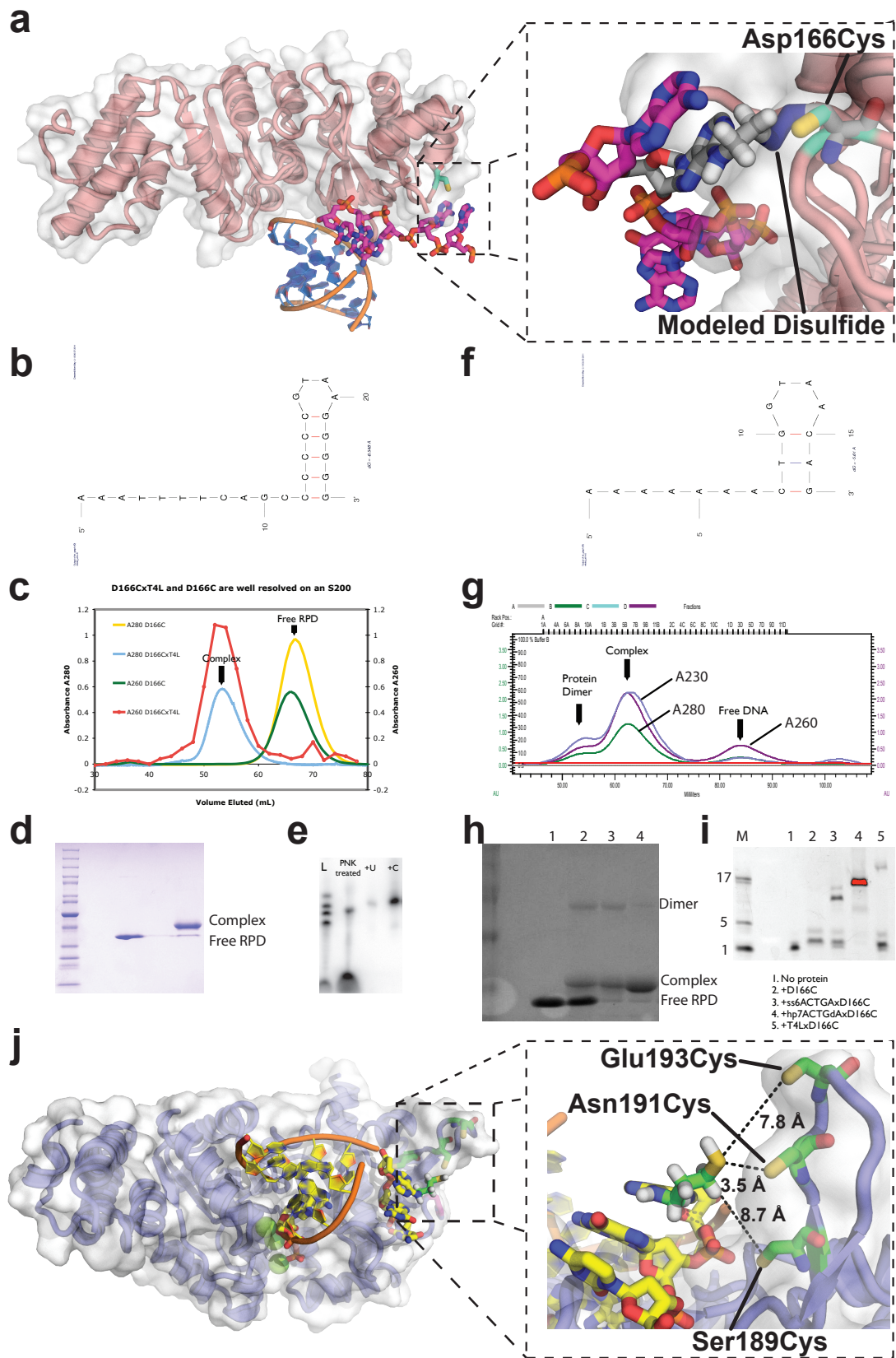
Figure AIII.1: Design, production, and purification of catalytically-active, crosslinked DnaG RPD•DNA complexes.

Figure AIII.1 Legend
a) Schematic of the modeling leading to selection of Asp166 as a possible crosslinking site. The *E. coli* DnaG RPD ssDNA bound structure (PDB code: 3B39) is shown as a cartoon representation in salmon, and a surface representation in white. The bound DNA is shown as sticks with carbon colored magenta. A modeled duplex is shown as a cartoon representation. Inset panel: Expanded view of the crosslinking site. The engineered cysteine is shown in its free form as a stick representation with carbon colored turquoise. A model of the disulfide that forms is shown, with sulfur colored blue, and the crosslinkable residues (DNA and protein) shown with carbon colored gray. Elements of the ssDNA are not involved in crosslinking are shown with carbon colored magenta. b) Schematic of the long hairpin DNAs used initially. c) Chromatograms of gel fitration runs with the indicated proteins. d) SDS-PAGE gel image showing protein before and after crosslinking and purification. e) Gel image showing incorporation of a radiolabeled nucleotide into the crosslinked hairpin. f) Schematic of the short hairpin DNAs used after solution of the NTP bound structures of the *Sa*DnaG RPD. g) Representative chromatogram of the gel filtration purification step with the short hairpin complexes. h) Gel image showing results of crosslinking and purification steps. Lanes: 1) before crosslinking, 2) After 48 h incubation with hairpin, 3) Pooled material from second Q-column, 4) Pooled material from gel filtration. i) Gel image showing incorporation of a fluorescently labeled nucleotide (Fl.12.ATP, see **Appendix II**) into the complexes with the shorter hairpins. j) Schematic of the modeling for new crosslinking sites of the *Sa*DnaG RPD. A model of the *S. aureus* DnaG RPD bound to a hemiduplex product is shown as a cartoon representation in slate, and a surface representation in white. modeled duplex is shown as a cartoon representation. Inset panel: Expanded view of a possible crosslinking site near the incoming nucleotide binding site. Possible locations for placement of engineered cysteines are shown as stick representations with carbon colored green, and are labeled according to *Sa*DnaG. A model of the crosslinkable DNA is shown as a stick representation with carbon colored yellow, and crosslinking moiety shown with carbon colored green. Sulfur is colored yellow throughout the panel.