2002

Structural Studies of DNA Replication Proteins by X-Ray Crystallography and Molecular Dynamics

Yanxiang Zhao

Structural Studies of DNA Replication Proteins by X-Ray Crystallography and Molecular Dynamics

A thesis presented to the faculty of

The Rockefeller University

in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

by

Yanxiang Zhao

To my family

Acknowledgements

I would like to first thank my advisor, John Kuriyan, for providing me the opportunity to work on interesting and educational projects that encompassed both protein crystallography and computational structural biology. By doing so, John taught me the importance of critical thinking to address biological questions across the borders of specific experimental or theoretical approaches. His insightful evaluation of the relevance of my many experimental results to the core biological question being addressed has been essential in guiding my research on the right course. His advice on how to become a better biologist by utilizing my background in physics has been especially helpful for me to define my future plans.

I would like to also thank members of the Kuriyan lab, present and past, who have made my time here very scientifically rewarding and personally happy. I would like to especially acknowledge David Jeruzalmi, who taught me protein crystallography starting from the ground level and provided enormous help at various stages of my project. Matt Young is gratefully acknowledged for teaching me about molecular dynamics and helpful discussions about the free energy calculations. Special thanks go to Morgan Huse, Xiaomin Chen, Elena Conti for their helpful discussion about protein crystallography and assistance in conducting experiments. Thanks also go to Steven Jacques and Lore Leighton for technical help, to Jeff Bonanno for skillful assistance at home X-ray source and synchrotron data collection, and to Ronni Gulli, Anne Roche, and Vicki Willet for administrative support.

I would like to thank Benoit Roux for introducing me to the modern science of computational structural biology, and advising me on the MD project in my thesis research. I am grateful to him for accommodating me in his lab at the final stage of my thesis preparation after John's lab moved to Berkeley. Thanks also to members of the Roux lab for their many helpful discussions about computational biology and great patience in explaining the theories of computation to an experimental crystallographer like me.

Seth Darst and Andrej Sali are gratefully acknowledged for serving on my thesis committee, and for doing his best to make my committee meetings and my defense into very educational experience. Andrej Sali is specially thanked for serving on my committee in the final stage of my thesis due to the departure of Stephen Burley, my former committee member, from Rockefeller. Thanks also to Aneel Aggarwal for serving as my outside examiner, and also for the many helpful suggestions on my thesis.

Finally, I would like to thank my family and friends for their support and kindness over the past six years. Special thanks to my parents, who inspired me to pursue a successful life through hard work and provided great care for my two children so that I can focus on my work, and my husband, who has been the most encouraging and supportive of anything I do.

Table of Contents:

List of Figures

List of Tables

Abstract

Organisms that span evolutionary space utilize an assembly of proteins (the replicase) in a coordinated effort to faithfully replicate their genomes. This chromosomal replicase consists of three functionally distinct subassemblies. The first of these is the polymerase/exonuclease complex, which harbors DNA synthesis and proofreading activities. The second functional complex is the sliding clamp which adopts a ring-shaped structure, composed of either two or three protomers. It confers processivity onto the polymerase subunit by tethering it to the template. The third complex is the clamp loader complex, which loads the sliding clamp onto DNA using energy from ATP binding and hydrolysis.

This thesis analyses the structures of two important components of the chromosomal replicase assembly through X-ray crystallography and molecular dynamics simulations. First, a crystal structure of the DNA polymerase from archaebacterium *Desulfurococcus Tok* (D. Tok Pol) was solved at 2.4 Å. The structure revealed its similarity to that of the DNA polymerase from bacteriophage RB69 in spite of a low sequence identity between these two members of the Pol II family of DNA polymerases. Secondly, a series of molecular dynamics simulations were performed on the sliding clamps from *Escherichia coli* and *Streptococcus pyogenes*. The studies demonstrated that one subunit of the dimeric clamp, when the other subunit is absent, would relax to a structure of reduced curvature ("open" state) when compared to its structure in the dimer ("closed" state). Free energy calculations suggest that this spontaneous structural change is driven by higher angle and dihedral energies in

the "closed" state. This finding led to the hypothesis that sliding clamps are spring-loaded rings that relax during the loading reaction when one of their oligomeric interfaces is disrupted by the clamp loader complex. Lastly, deconvolution of X-ray diffraction data from a perfectly merohedrally twinned crystal was used to improve the structure of the human sliding clamp in complex with a peptide derived from $p21^{WAF1/CIP1}$, a DNA replication inhibitor, from a resolution of 2.6 Å to a resolution of 2.3 Å.

Introduction

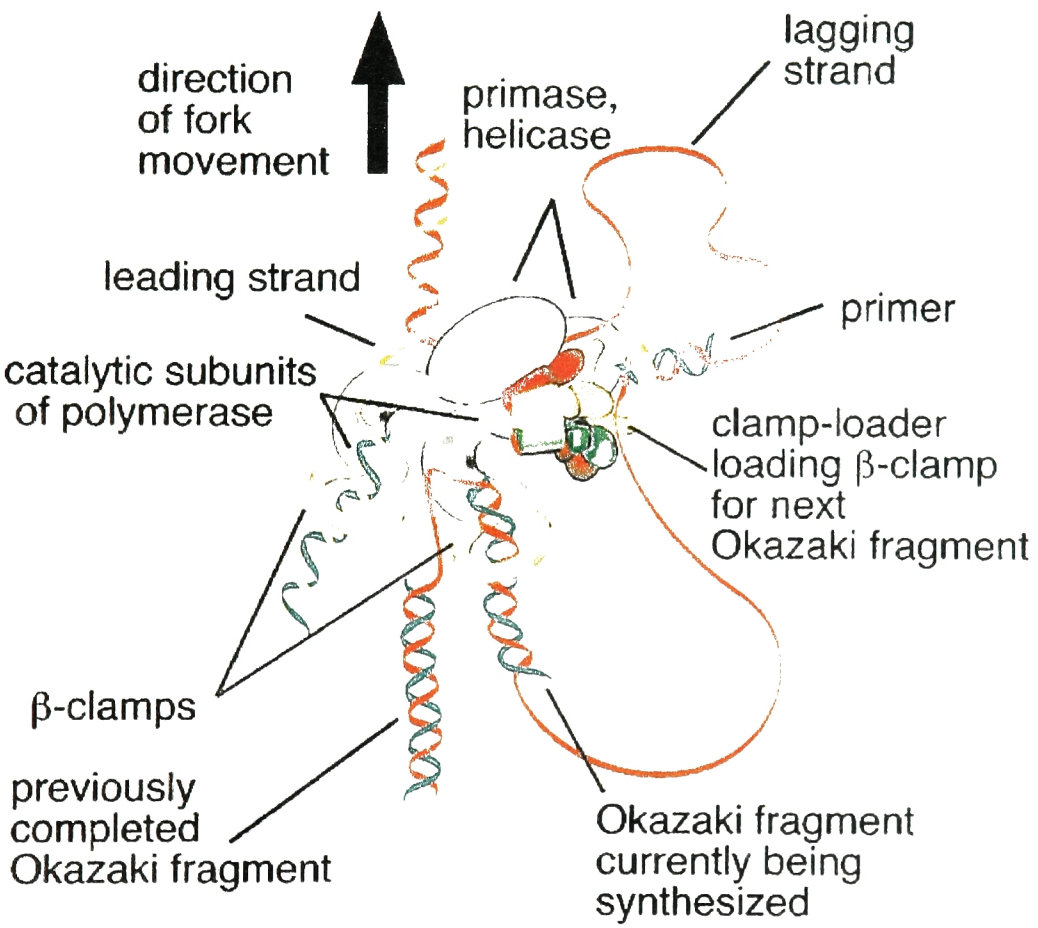
DNA replication is a cellular event of central importance in cell proliferation. During each cell division, chromosomal DNA, the repository of genetic material, is duplicated and then equally divided between the two daughter cells. This process ensures that the new cells produced will carry the same genetic information as their parents, thus maintaining genetic stability from one generation to next. The duplication of DNA is carried out by a DNA replication machinery composed of multiple proteins working in a collaborative fashion. Accuracy in this replication process is of crucial importance. This process is also subject to regulation from other cellular components for accurate timing in the cell cycle.

DNA molecules are made up by two anti-parallel strands of deoxyribonucleotides wound into a right-handed double helix (Watson and Crick, 1953b). The nucleotide sequence on one strand is complementary to that on the other according to the A-T, G-C pairing rule, favored energetically by hydrogen bonding between two nucleotide bases (Watson and Crick, 1953b). This complementarity of the double helical structure suggested a simple and elegant mechanism for replication (Watson and Crick, 1953a). Since each strand of DNA contains the nucleotide sequence that is exactly complementary to its double helical partner, the two strands can be separated and each of them can serve as a template for synthesizing a new strand, whose nucleotide sequence will be dictated through base-pairing with the template nucleotides, resulting in a product that is the same as the template's original complementary partner. At the end of

such an operation, two copies of the original DNA strands are generated, each containing a strand from the original DNA molecule and a newly synthesized complementary strand. The original nucleotide sequence, i.e. the genetic code of the cell, is thus faithfully duplicated.

In reality the actual implementation of this elegant mechanism is a complicated process involving many players (Kornberg and Baker, 1991). The copying process is a well-coordinated effort of a consortium of several protein complexes. First, when cell cycle timing is right, multiple proteins involved in initiating DNA replication assemble at the origins of DNA replication and trigger this process. Then the double helical structure of DNA is unwound by helicases, which are multi-subunit proteins operating in an ATP-powered motor-like fashion to unwind the DNA double strands. The unwound single strands of DNA are quickly coated by single-strand DNA-binding (SSB) proteins to prevent them from tangling or forming secondary structure. Primases lay down short stretches of RNA along the single-stranded DNA to serve as the initiation segments for the DNA polymerases. These preparatory steps set the stage for DNA replication core complex to engage productive DNA replication. The Y-shaped junction of the double helical unwound parent DNA and its two separated single strands is popularly referred to as "the replication fork". As DNA replication progresses, the fork moves along the double stranded DNA to expose fresh segments for polymerase replication (Kornberg and Baker, 1991) (Figure 1.1).

The DNA polymerase complexes (the replicases) that carry out chromosomal DNA replication in different organisms all comprise of three


direction
of fork
movement
primase,
helicase
lagging
strand
leading strand
primer
catalytic subunits
of polymerase
clamp-loader
loading β-clamp
for next
Okazaki fragment
β-clamps
previously
completed
Okazaki fragment
Okazaki fragment
currently being
synthesized

functionally distinct components. These are the DNA polymerase/exonuclease subunits, the sliding clamps, and the clamp loader complex, which loads the sliding clamp processivity factor onto DNA at the beginning of DNA replication or off DNA at the end of this process (Table 1.1). The sliding clamps and the clamp loader complexes are collectively known as the processivity factors.

The DNA polymerase/exonuclease subunits consist of the key catalytic enzymes, the DNA polymerase and exonuclease. All DNA polymerases, whether participating in chromosomal replication or in other activities like DNA repair, are unidirectional enzymes that catalyze nucleotide addition only at the 3' end of the elongating product strand. This means that only one strand, the leading strand, can be replicated continuously in the same direction as the replication fork. On the other strand, the lagging strand, new strand synthesis is also carried out in a 5'-3' direction, but discontinuously, in segments of 1000-2000 bases. This results in long stretches of new product DNA, called Okazaki fragments after the discoverer (Sugino et al., 1972). These Okazaki fragments are later ligated together by DNA ligase to form a single continuous strand. The exonuclease activity in the DNA polymerase/exonuclease subunit acts like a proof-reading step to remove mistakes in the newly-synthesized strands (Kunkel, 1992).

The DNA polymerase subunits of the chromosomal replicases are usually distributive enzymes when acting alone (i.e., removed from the rest of the replicase assembly) because they dissociate from their DNA substrates after

Table 1.1 Families of Processive DNA Replicases

E. coli	Bacteriophage T4	Eukaryotes	Archaebacteria	Function
Pol III core ($\alpha\theta\epsilon$)	gp 43	Pol δ/ϵ	DNA polymerase of Pol II family	DNA polymerase/exonuclease
β	gp 45	PCNA	PCNA	Sliding clamp, attaches to DNA polymerase, stimulates DNA-dependent ATPase activity of clamp loader complex
γ complex ($\gamma_3\delta\delta'\chi\psi$)	gp 44/62	RFC 1-5	RFC L/S	Clamp loader complex, DNA-dependent ATPase, primer-template binding, stimulates loading of DNA polymerase

Pol, polymerase; gp, gene protein; PCNA, proliferating cell nuclear antigen; RFC, replication factor C

very short runs of nucleotide incorporation. The processivity factors in the DNA replicase assembly convert the distributive DNA polymerase into a highly processive enzyme so that chromosomal DNA replication can proceed for thousands of nucleotides without the polymerase dissociating from the DNA template (Kornberg and Baker, 1991). The name "sliding clamp" for the protein that tethers the polymerase to DNA originated from biochemical studies that showed this protein to slide freely along double-stranded DNA in ATP-independent fashion with a long half-life on circular DNA (Huang et al., 1981; Stukenberg et al., 1991). Crystal structures of sliding clamps from different organisms illustrated a common ring-shaped topology composed of two or three crescent-shaped monomers that form a central channel that is large enough to encircle double stranded DNA (Kong et al., 1992; Krishna et al., 1994). DNA polymerases achieve high processivity by attaching to the stable yet free-to-move platform on DNA provided by the sliding clamps via C-terminal segments that bind to the sliding clamps. The DNA polymerase and its associated sliding clamp then move along the DNA strands as one entity during the replication process (Stukenberg et al., 1991).

The circular shape of the clamp requires that it be opened up by a "clamp loader" complex and loaded onto DNA before the polymerase can attach itself to it and start processive replication. The clamp loader complex is a multi-protein assembly driven by energy from ATP binding and hydrolysis to temporarily separate one interface between protomers in the sliding clamps, creating a gap

in the circular ring for DNA to pass through (Onrust et al., 1991; Turner et al., 1999).

In addition to the major players listed above, the DNA polymerase complex interacts with several other proteins to ensure the smooth and controlled operation of this process. For example, the initiation of DNA replication is regulated by cell cycle proteins for accurate timing. One such player in eukaryotic cells, $p21^{WAF1/CIP1}$, binds to the sliding clamp in a manner that blocks chain elongation by DNA polymerase, thus stalling the replication process (Peter and Herskowitz, 1994). Interestingly, another segment of $p21^{WAF1/CIP1}$ binds to and inhibits cyclin dependent protein kinases (CDKs) for cell cycle regulation (reviewed by (Peter and Herskowitz, 1994)).

The general features of DNA replication and the components of the replication machinery are conserved in prokaryotes, eukaryotes, archaebacteria as well as certain bacteriophages such as T4 (and its relative RB69). Insight gained from studying the replication systems from *E. coli*, T4 bacteriophage, yeast, and humans has provided a unified view of the replication mechanism, and has added species-specific details that enrich our understanding of replication. Structural biology has made important contributions to a detailed understanding of the functional mechanism of several key proteins in the replication process. Key proteins in the DNA replicase, including the DNA polymerase, the processivity factor, and its loading complex, have been most extensively studied structurally (Ellison and Stillman, 2001; Jeruzalmi et al., 2002; Steitz, 1999).

Two kinds of proteins in the DNA replication assembly have been the focus of my thesis studies. The first one is a DNA polymerase/exonuclease (sometimes referred to as DNA polymerase or DNA Pol for brevity in this thesis), the enzyme in the replication machinery that is responsible for the key function of DNA synthesis. I chose a DNA polymerase from the archaebacterium *Desulfurococcus Tok* (*D. Tok*) as my subject for structural analysis by X-ray crystallography. *D. Tok* DNA polymerase (D. Tok DNA Pol) is part of the archaebacterial replicase complex and belongs to the so-called Pol II family of DNA polymerases, based on sequence similarity to *E. coli* DNA polymerase II (Braithwaite and Ito, 1993). No structure of a Pol II family DNA polymerase was known at the time of our initiating this project. D. Tok DNA Pol also possesses high sequence identity to the Pol II family DNA polymerases involved in eukaryotic chromosomal replication, whose structures are still unknown. Although there are now several crystal structures available for *E. coli* DNA polymerase I and its homologous DNA polymerases from other eubacteria, either with or without DNA substrate bound, they are not related at the sequence level to the Pol II family of DNA polymerases. Sequence alignment indicates no overall similarity between these two classes of polymerases except for a few residues at the polymerase or exonuclease catalytic sites. Functional studies indicate that *E. coli* DNA polymerase I and its homologs, also called Pol I family DNA polymerases, are generally involved in synthesis of relatively short stretches of DNA in processes such as DNA repair, which contrast with the Pol II family's functional role of chromosomal DNA replication (Braithwaite and Ito, 1993). The

structure of D. Tok DNA polymerase represents a topology that is likely applicable to all Pol II family archaebacterial and eukaryotic DNA polymerases. Although the structure I have determined is without DNA, the mode of binding of the DNA template and exit of product duplex DNA can be inferred from structural alignment to other DNA/Polymerase complexes. We also identified a RNA-binding domain in the structure that likely exists in human homologs.

The second topic of my thesis involves the sliding clamps, which are indispensable for speedy replication of long stretches of chromosomal DNA. The main question we tried to address concerns the mechanism by which the sliding clamp is opened before it can load onto DNA. Through molecular dynamics simulations of sliding clamps from *E. coli*, and *S. pyogenes*, we discovered that while these sliding clamps are very stable in their closed-ring states, there is a large internal strain stored in the structure so that weakening of one of the interfaces between two monomer components of the clamp by the clamp loader complex suffices to trigger the clamp to spring to an open form.

The third and last part of my thesis involves trying to obtain high-resolution structural details of a complex between the human proliferating cell nuclear antigen (PCNA), which is the sliding clamp for human DNA replicase, and a 22-residue C-terminal peptide of cell-cycle checkpoint protein $p21^{WAF1/CIP1}$ by molecular replacement to better understand the nature of the interaction between $p21^{WAF1/CIP1}$ and the clamp. A crystal form of this complex diffracting to 2.0 Å suffered from perfect merohedral twinning, which arises from the crystal containing equal proportions of two identical crystals lattices of P3 space group

related by 60° rotations. The diffraction patterns from these two crystal lattices are identical in space due to their special arrangement relative to each other, i.e. P3 symmetry and 60° rotation. Thus the intensity of each diffraction spot (h,k,l) measured by the detector is a sum of contributions from these two lattices. Using a previously-solved 2.6 Å structure of the same complex as the starting model (Gulbis et al., 1996b), an iterative detwinning procedure using the algorithm of Yeates (Redinbo and Yeates, 1993) was carried out to separate the intensities from these two twinned diffraction patterns and to refine a model at 2.3 Å resolution. This detwinning process brought out new structural information for previously disordered regions in the lower resolution structure. It also provided a valuable training in understanding the mathematical and physical principles behind protein crystallization and structure solution.

Chapter 1. Crystal Structure of an Archaebacterial DNA Polymerase

1.1 Introduction

1.1.1 DNA Polymerase Families

DNA polymerases involved in chromosomal DNA replication belong to a broadly distributed family of template-dependent polymerases that catalyze the addition of an energetically favored nucleotide to the 3' end of the elongating primer chain as dictated by the Watson-Crick pairing rule according to the nucleotide sequence on the template strand. These enzymes are divided into two big families of DNA polymerases or RNA polymerases, depending on the products generated. The family of DNA polymerases can again be subdivided depending on the template used, including RNA-dependent DNA polymerases, such as reverse transcriptases in viruses, or DNA-dependent DNA polymerases (referred to as DNA Polymerases or DNA Pols in this thesis for brevity), such as the ones in DNA replication. The DNA-dependent DNA polymerases so far discovered can be categorized into five large families, based on their sequence similarity to *E. coli* DNA Polymerases I, II, III, IV, and V (Braithwaite and Ito, 1993; Sutton et al., 2000). The Pol I family of DNA Polymerases includes the *E. coli* Pol I and its homologs from eubacteria, such as those from *Thermus aquatics*, *Streptococcus pheumoniae*, which are mainly responsible for synthesis of short stretches of DNA in DNA replication and repair. It also contains the T7 bacteriophage DNA polymerase, which is responsible for replication of the whole

T7 plasmid as an exceptional case for Pol I type of DNA polymerases. The Pol II family is the largest, and contains *E. coli* Pol II, many archaebacterial DNA Polymerases such as those from *Thermococcus litoralis* (Vent), *Pyrococcus furiosus* (Pfu), *Desulfurococcus Tok* (*D. Tok*) and T4 bacteriophage DNA Pol. Eukaryotic DNA polymerases α, δ, and ϵ are all type II DNA polymerases. Many viral DNA polymerases are categorized into this family too.

Except for *E. coli* Pol II, all the listed members of Pol II family DNA polymerases are involved in chromosomal or plasmid (for bacteriophages and viruses) replication. They perform this synthesis in a processive manner together with their processivity factors and other accessory proteins. D. Tok DNA Pol from this family is the subject of our study because of its relevance in chromosomal DNA replication.

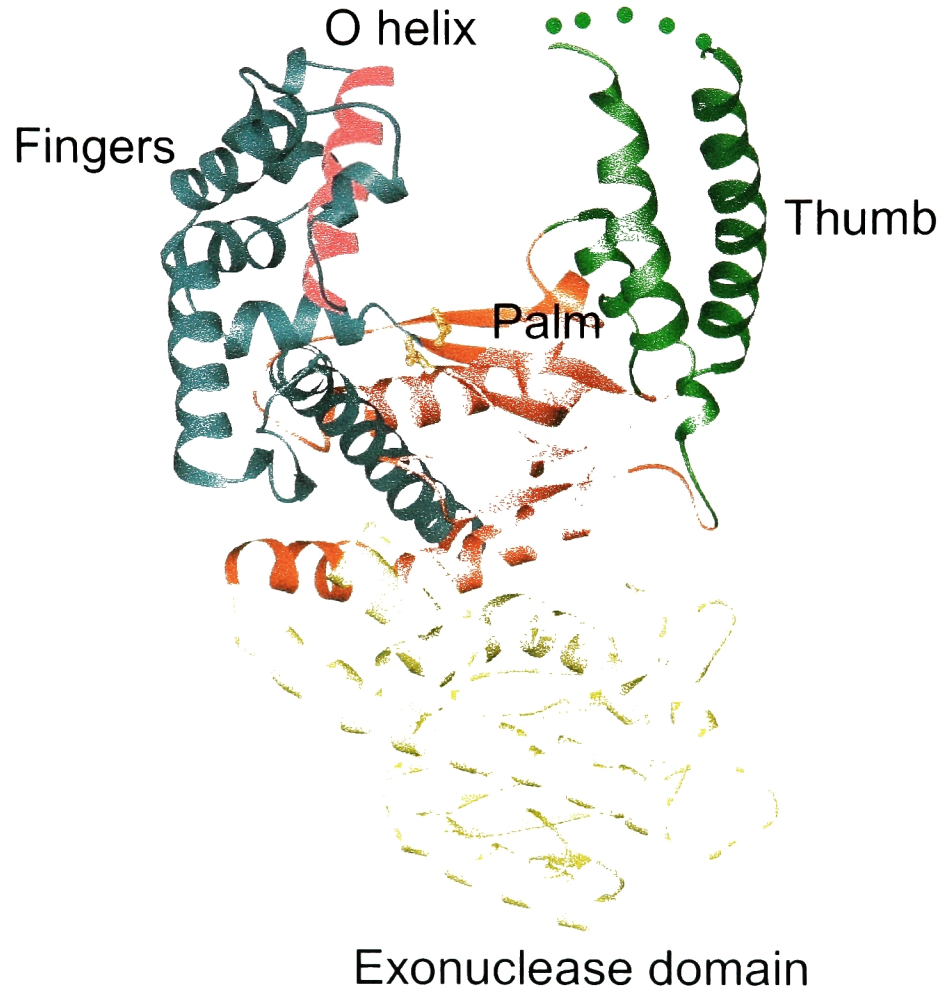
The Pol III family is rather small and so far only consists of *E. coli* Pol III and its relatives in eubacteria. In *E. coli*, the Pol III holoenzyme, an assembly of proteins including Pol III (polymerase α, exonuclease ϵ, and the θ subunit), the sliding clamp (β dimer), the clamp loader complex $[(\gamma/\tau)_3\delta\delta'\chi\Psi]$, is the functional entity for chromosomal replication. This system has been extensively studied, both biochemically and structurally. Findings from this system have been found to be generally applicable to other replicase systems, including archaebacteria and eukaryotes. The Pol IV and V family are recently discovered emerging families of DNA polymerases, which have the distinct ability of replicating imperfect DNA templates, such as mismatched primer-template junctions, or

chemically altered bases. They are mainly involved in the cellular response to DNA damage. They include *E. coli* Pol IV (UmuC), *E. coli* V (DinB), several eukaryotic homologs, and are sometimes referred to as the UmuC superfamily because UmuC was the first discovered and best studied (Sutton et al., 2000).

Sequence alignment and biochemical studies confirmed that two well conserved sequence motifs A and C (Delarue, 1990) harbor the two key aspartic acid residues for the catalysis of nucleotide addition in all DNA polymerases (Kornberg and Baker, 1991). For DNA polymerases that contain 3'-5' exonuclease activity to improve fidelity, the active site for exonuclease activity contains a well conserved DxE sequence motif. There is significant sequence and functional conservation among members of DNA polymerases within one family to suggest a common mechanism for reaction. But across different families there is little sequence similarity, except for the presence of the conserved motifs mentioned above, suggesting different structural scaffoldings for different families.

Crystal structures of DNA polymerases, either with or without DNA/nucleotide substrate, have been solved for four of the five families, yielding significant information on the catalytic mechanism and substrate specificity of DNA polymerases. The Pol III family is so far the only one lacking a representative crystal structure. All the DNA polymerase structures revealed a common architecture in the core of the polymerase domain that can be likened to a right hand, consisting of the palm, thumb, and fingers subdomain (Joyce and Steitz, 1994) (Figure 1.2). The palm domain is positioned at a central solvent-

Figure 1.2 Structure of Klenow Fragment of *E. coli* DNA Pol I. Ribbons representation of the structure of Klenow Fragment of *E. coli* DNA polymerase I shows the four domains (Ollis et al. 1985). The domains are illustrated as Exo (yellow), palm (red), fingers (blue), and thumb (green). The O helix important for nucleotide incorporation was colored magenta and labeled. Two carboxylates at the polymerase active site are shown.


O helix
Fingers
Thumb
Palm
Exonuclease domain

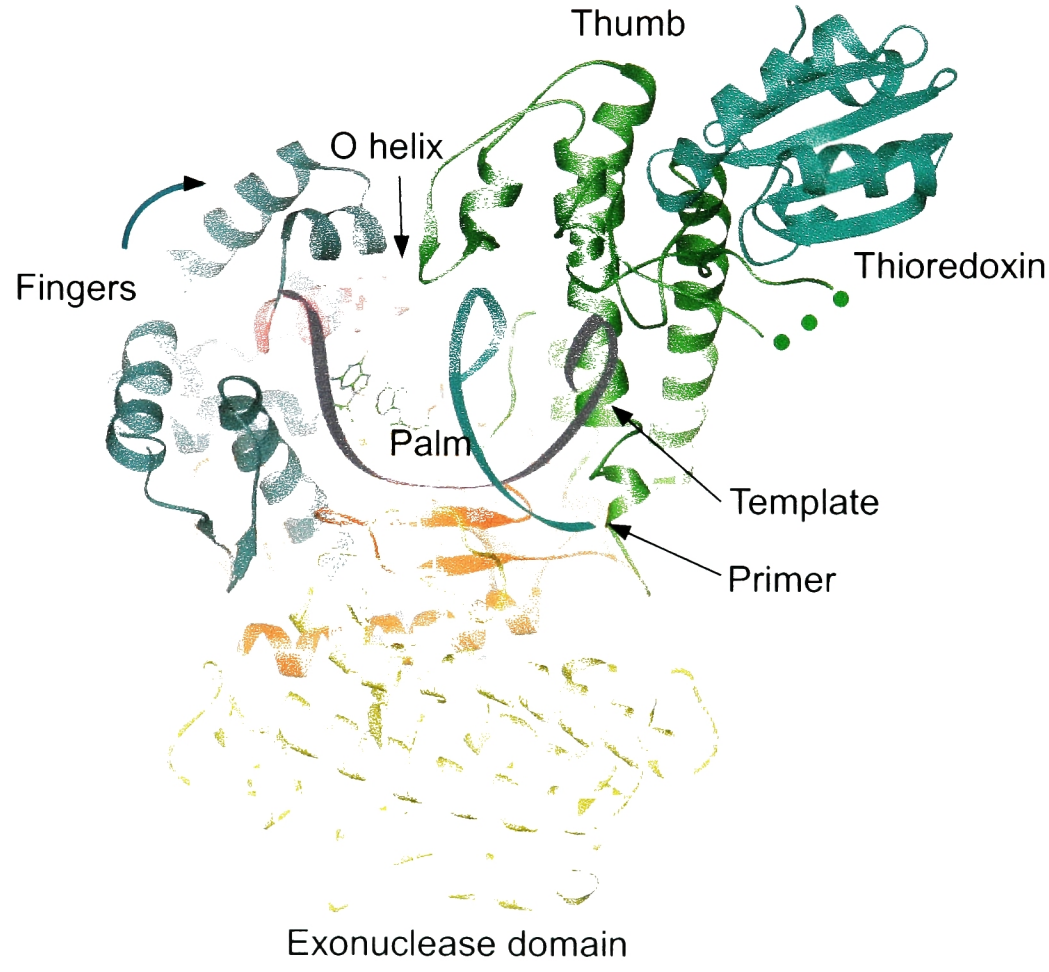
accessible cleft formed by the surrounding fingers and thumb domain. The two key acidic residues, together with other well conserved residues in their sequence vicinity, reside on two neighboring β strands to form the polymerase active site (Figure 1.2). The fingers domain contains an important helix (O helix) with well conserved residues, which is responsible for interacting with the incoming nucleoside triphosphate and in increasing the fidelity of nucleotide incorporation at the polymerase active site (Doublie et al., 1998) (Figure 1.3). The thumb domain contributes to interactions with the duplex product DNA and the DNA polymerase's sliding clamps, as well as to product translocation (Figure 1.3). While the fold of the palm domain is similar in its central features among different DNA Pol families with known structure, the fingers and thumb domains from different families have divergent structural details (reviewed by (Steitz, 1999)). In addition to these polymerases involved in replication, considerable insight into DNA synthesis has been obtained by structural studies on the DNA repair enzyme Pol β (Pelletier et al., 1994). This enzyme is not directly related to the other polymerases, but is instead a member of a distinct family of nucleotide transferases (Steitz, 1999).

1.1.2 Extensive Structural Information for the Pol I family

The Pol I family of DNA polymerases is the first and best studied structurally among the five DNA Pol families, providing significant insight into the common functional mechanism of polymerase catalysis and substrate fidelity that is presumably shared by all DNA polymerases. There is a long list of crystal structures solved in this family. The first DNA polymerase structure obtained was

for the Klenow fragment of *E. coli* Pol I, which is a proteolytic fragment derived from the full-length protein, retaining the polymerase and 3'-5' exonuclease domains. The right-hand analogy and "palm, thumb, fingers" terminology were developed from this first structure and later adopted by other structures (Ollis et al., 1985) (Figure 1.2). The structure of the complex with deoxynucleoside monophosphate product and a single-stranded DNA, or a thymidine tetranucleotide, at its exonuclease active site, revealed a detailed picture of the editing mode of this enzyme, which led to the proposal of a two metal ion mechanism for phosphoryl transfer (Beese and Steitz, 1991). A 2.2 Å crystal structure of T7 bacteriophage DNA polymerase in complex with a primer-template DNA substrate and a nucleoside triphosphate in the polymerase active site (Doublie et al., 1998) was the first structure to provide a complete picture of how the polymerase binds to its DNA substrate, catalyzes the phosphoryl transfer reaction using the same two metal ion mechanism as seen in exonuclease active site, and achieves substrate specificity (Figure 1.3). Thioredoxin, the processivity factor for T7 DNA replication, is also present in the complex crystallized. But how it confers processivity to the T7 DNA polymerase is not clear from the structure because it binds at the tip of the thumb domain of T7 DNA Pol, far away from the palm domain or DNA substrate (Doublie et al., 1998) (Figure 1.3). There are also crystal structures of other Pol I type DNA polymerases, with or without DNA substrate, such as those from thermophilic eubacteria like *Thermus aquaticus* (Eom et al., 1996; Kim et al., 1995; Li Y et al., 1998), *Bacillus stearothermophilus* (Kiefer et al., 1998; Kiefer et al., 1997), which

Figure 1.3 A ternary complex of the T7 DNA Pol with DNA and ddNTP. Ribbons representation of the structure of the ternary complex of T7 DNA polymerase with primer-template DNA, incoming dideoxynucleotides and thioredoxin (Doublie et al. 1998). The protein domains are colored according to *E. coli* structure. The O helix important for nucleotide incorporation was colored magenta and labeled. Two carboxylates at the polymerase active site are shown. The incoming nucleotide are shown in bond-and-stick representation and colored according to the atom. The two Mg^{2+} ions are the polymerase active site are shown as silver spheres.


Thumb
O helix
Fingers
Thioredoxin
Palm
Template
Primer
Exonuclease domain

added to our knowledge of polymerase catalysis and replication fidelity for this family.

Summarizing all the information provided by crystal structures of different species of Pol I family DNA polymerases and their complexes with various substrates, either at the exonuclease or polymerase domain, we have obtained an almost complete picture of the structural topology and functional mechanism of Pol I family DNA polymerases (Doublie et al., 1999; Steitz, 1999). All Pol I family DNA polymerases share a highly conserved chain fold and similar domain organization. The polymerase and exonuclease domains are two distinct subunits in the structure. The palm, thumb, and fingers subdomains in the polymerase subunit are arranged in a U shape, resembling a right hand, with the palm domain guarded on each side by the fingers and thumb domain, both mostly helical (Figure 1.2). The palm domain, which contains the polymerase active site, consists of β strands flanked by two long α helices. The catalytic site contains the two absolutely conserved aspartic acid residues, which reside on two antiparallel β strands, coordinating two divalent metal ions (Mg^{2+}) that facilitate the nucleophilic attack on the nucleoside triphosphate by the 3'-hydroxyl of a DNA primer (Figure 1.2, 1.3).

The polymerase domain undergoes significant conformational changes to form a more closed structure upon primer-template and substrate binding, as compared to its unliganded state (Figure 1.3). This change is most drastic for the fingers domain, which binds the incoming nucleoside triphosphate and brings it to the polymerase active site for incorporation into primer DNA strand (Doublie et al.,

1998; Li et al., 1998). The O helix in the fingers domain, which contains several highly conserved residues responsible for nucleoside triphosphate binding and substrate specificity, rotates about 40° inward to close upon the active site (Figure 1.3). The thumb domain also closes in toward to the palm domain when primer-template DNA is bound, so as to accommodate the duplex DNA which lies at a groove formed by the palm and thumb domains (Figure 1.3).

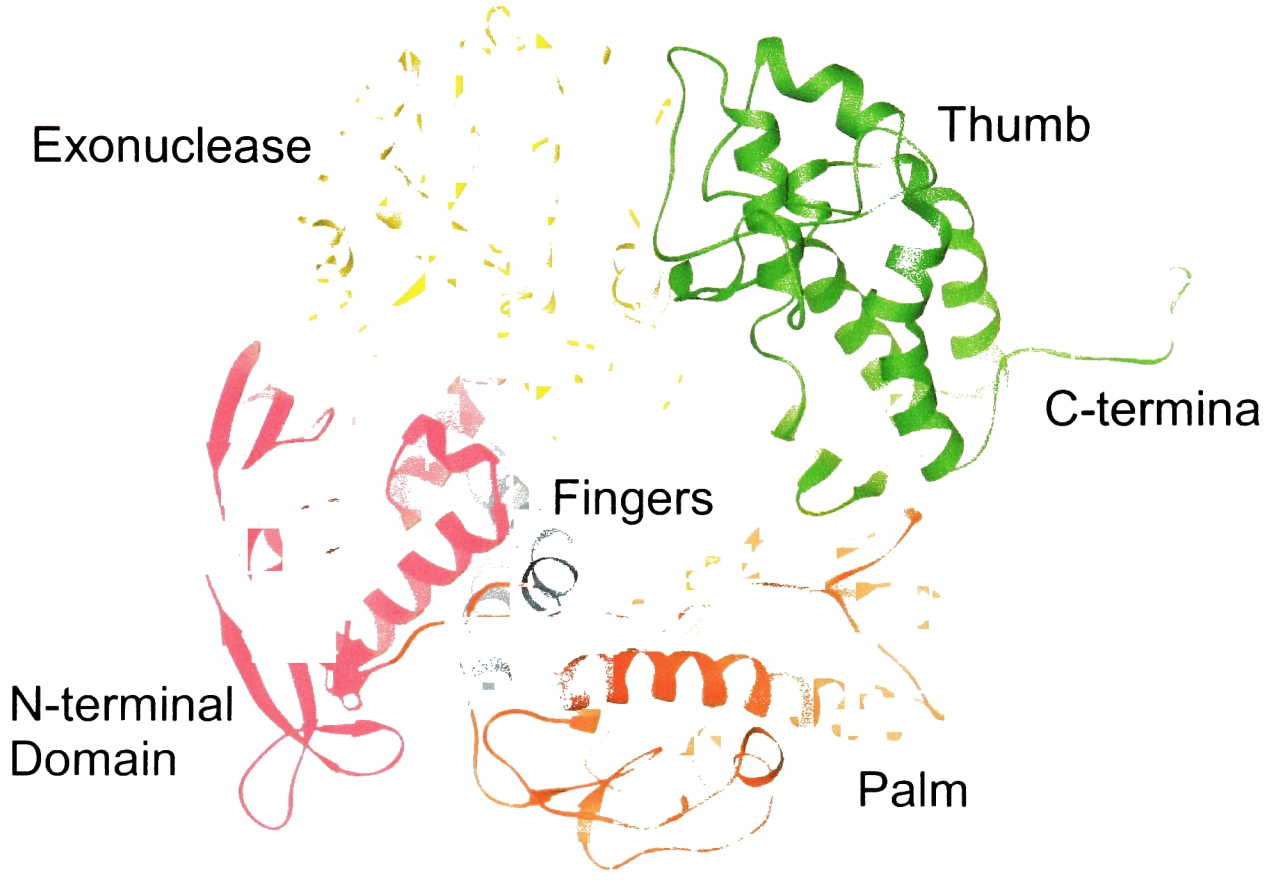
It is hypothesized that at each step of DNA synthesis, the polymerase alternates between the closed conformation as seen in the substrate-bound structures for nucleotide incorporation and the opened conformation as seen in unliganded structures for product translocation. The exonuclease domain, which is structurally separated from the polymerase domain, serves an important role in a proof-reading step to improve the accuracy of DNA synthesis by trimming mistakenly incorporated nucleotides on the growing DNA strand. Its active site, containing two carboxylates (Asp and Glu), utilizes the same two metal ion mechanism as the polymerase active site, but for phosphodiester bond cleavage. It is still unclear how the enzyme switches between polymerizing mode and editing mode, i.e., how it detects misincorporation at polymerization steps and subsequently directs the DNA to the exonuclease active site for error correction, and how it brings the DNA back for further synthesis at the polymerase active site once all the mistakes are removed. It is hypothesized that the different shapes of mismatched bases, distinct from correct base-pairs, weaken their steric fit within the polymerase active site, prompting the dissociation from the polymerase active site to migrate to the exonuclease site (Doublie et al., 1998).

1.1.3 Pol II family of DNA polymerases

Members of the Pol II family of DNA polymerases are responsible for chromosomal replication in eukaryotes, and they carry out highly processive DNA replication when attached to ring-shaped sliding clamps. They achieve high fidelity through an internal error-correcting 3'-5' exonuclease activity. Archaebacterial DNA polymerases and eukaryotic DNA polymerase α, δ, and ϵ are members of the Pol II family (Braithwaite and Ito, 1993). Given their importance in eukaryotic DNA replication, there is interest in further extending our knowledge of structures and mechanisms for this family of DNA polymerases.

The first structure of a Pol II family DNA polymerase to be determined was that of bacteriophage RB69, a close relative of T4 (Wang et al., 1997b). The structure of RB69 DNA polymerase revealed that the general architecture of the catalytic core of the polymerase of the Pol II family is strikingly similar to that of the polymerases of the Pol I family. Using the right hand analogy, the polymerase domain in the RB69 structure consists of a central palm domain surrounded by fingers and thumb domains (Figure 1.4). The palm domain is similar to that of DNA Pol I in its fold and in its arrangement of two conserved acidic residues to construct the polymerase catalytic center. The fingers and thumb domains of RB69 are functionally similar to the corresponding elements of Pol I, but the underlying polypeptide chain fold is quite different. The exonuclease domain of the Pol I and Pol II DNA polymerases are closely related in sequence and, not surprisingly, the structure of the exonuclease domain of RB69 resembles that of the Pol I type polymerase. Given the general similarity in the polymerase

Figure 1.4 Structure of Pol II family DNA Pol from RB69. Ribbons representation of the RB69 DNA polymerase structure shows its five domains with the C-terminus extending to the right (Wang et al. 1998). The domains are NH_2-terminal (yellow), Exo (magenta), palm (red), fingers (blue), and thumb (green). The three carboxylates in the polymerase active site are shown.


Exonuclease
Thumb
C-terminal
Fingers
N-terminal
Domain
Palm

domains of the Pol I polymerases and RB69 despite lack of sequence similarity, the location of the exonuclease domain in RB69 turned out to be a surprise. In RB69 the 3'-5' exonuclease domain is located above the fingers and opposite the thumb sub-domains, suggesting that the shuttling of DNA between the polymerization and proofreading sites must occur in different directions in polymerases of these two families.

The DNA polymerase from the archaebacterium *Desulfurococcus* strain *Tok* (D. Tok Pol) is a member of the Pol II family with both thermostable DNA polymerase and 3'-5' exonuclease activities (Lasken et al., 1996). D. Tok Pol sustains undiminished DNA polymerase activity after incubation at 95°C for one hour (R. Lasken, unpublished). The sequence of D. Tok DNA polymerase is very closely related (>75% identity) to that of other archaebacterial DNA polymerases, such as those from *Pyrococcus furiosus* (Uemori et al., 1993) and *Thermococcus littoralis* (Kong et al., 1993). D. Tok DNA polymerase is also related to eukaryotic DNA polymerases α, δ and ϵ (34% sequence identity over 196 residues of the DNA polymerase core for the human δ sequences (Braithwaite and Ito, 1993)). The archaebacterial genomes also contain genes coding for proteins with clear homology to proliferating cell nuclear antigen (PCNA), the DNA polymerase clamp in eukaryotes, as well as subunits of the clamp-loader complex RF-C (replication factor C). It is likely that archaebacterial DNA polymerases achieve processivity by attachment to the ring-shaped PCNA, although direct evidence for such a mechanism is lacking. The structure of an archaebacterial PCNA has

been determined recently, confirming the structural correspondence to eukaryotic PCNA (Matsumiya et al., 2001).

We have determined the structure of the D. Tok DNA polymerase at 2.4 Å resolution. D. Tok Pol shares less than 20% sequence identity with RB69 Pol, but the structures of the two enzymes resemble each other closely. The structure reported here has been determined in the absence of DNA. Nevertheless, the close structural correspondence between the active sites of pol I and pol II DNA polymerases allows inferences to be made about the mode of DNA recognition by D. Tok Pol. The very N-terminal region of D. Tok Pol contains a domain (residues 1 to 132) which is closely related in structure to single stranded RNA binding domains (RBDs), also known as RNA recognition modules (RRMS) (Varani and Nagai, 1998). The structure of the 3'-5' proofreading exonuclease domain of D. Tok Pol is similar to that of the Pol I type polymerases. However, its location relative to the palm sub-domain resembles that seen in RB69 (Wang et al., 1997b) rather than the Pol I-type polymerases (Kim et al., 1995; Korolev et al., 1995; Ollis et al., 1985). The structure of D. Tok polymerase reported here provides further evidence that the mode of DNA template recognition and the distinct editing channel established for the Pol II family by the structure of RB69 Pol is valid for the entire Pol II family.

1.2 Methods and Materials

1.2.1 Protein Expression and Purification

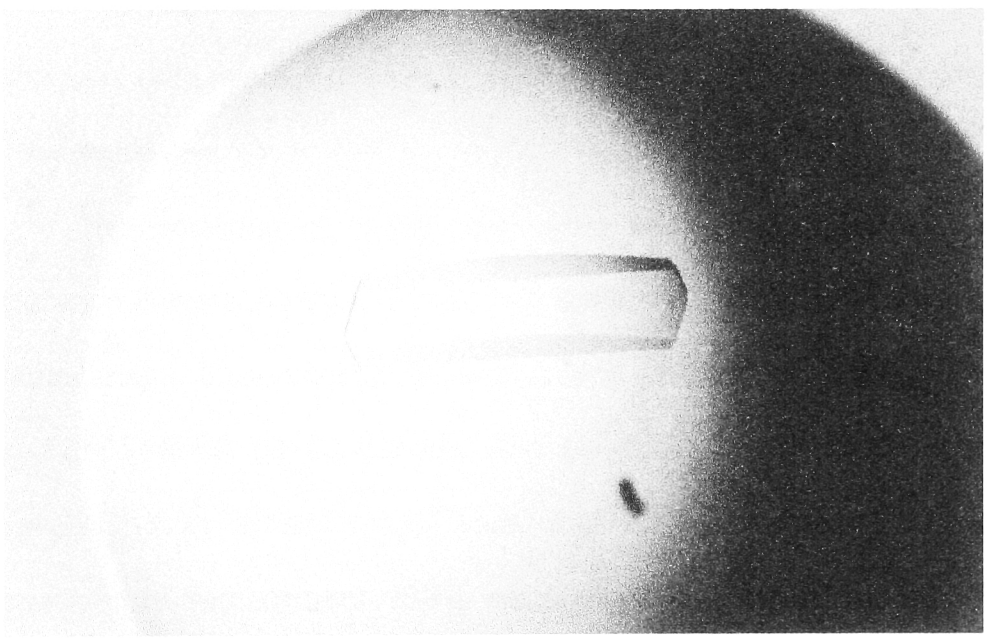
The D. Tok DNA polymerase bacterial expression vector and partial amino acid sequence were generous gifts of Life Technology Corporation. Convenient and reproducible protein expression was achieved by cloning the D.Tok Pol gene into the pET30 expression vector. Determination of the amino acid sequence of the polymerase was completed using this construct. Bacterial expression was performed using BL21(DE3) *E. coli* strain. For expression, bacteria were grown at 37°C and induced with 1mM isopropyl-thio-β-D-galactoside (IPTG) at an optical density (OD) of 2.0 (absorbance measured at 600 nm). Induction was generally allowed to proceed for about 12 hours overnight. This induction at high OD and prolonged time period was found to be critical in improving yield. *E. coli* cells from 8 liters of culture prepared from the above expression system were thawed and resuspended in 200 mL of 50 mM Tris pH 7.5, 50 mM KCl, 10 mM MgCl2, 10% glycerol v/v, 5 mM dithiothretol (DTT), 1 mM phenyl-methyl sulphonyl fluoride (PMSF). The cells were lysed in a French pressure cell (Avestin) and the resultant lysate subjected to centrifugation at 16,500 rpm in a midspeed Beckman J2-MI centrifuge for 45 minutes. The soluble fraction of the lysate was incubated at 80°C for 30 min and spun down again at 16,500 rpm for 45 minutes. The soluble sample was then applied to a Heparin column (5 mL bed volume, Pharmacia) using a peristaltic pump. The Heparin column was then washed with a buffer containing 50 mM Tris pH 7.5, 5 mM DTT, 10% glycerol v/v, and then eluted with a linear gradient into a buffer containing 50 mM Tris 7.5, 1 M

KCl, 5 mM DTT, 10% glycerol v/v. Fractions containing D. Tok Pol were confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using standard protocols and pooled. This semi-pure D. Tok Pol protein was then dialysed using membrane of 50 kD molecular weight cutoff in a buffer of 50 mM Tris pH 7.5, 10% glycerol v/v in 40X volume. This sample was loaded onto an ion-exchange High-Q column (5 mL bed volume, Pharmacia) and eluted with the same gradient and buffer as for the Heparin column. Pooled fractions of D. Tok Pol, as confirmed by SDS-PAGE, were concentrated using a 15mL Biomax ultrafiltration device (Millipore) with a 50 kD cutoff of molecule weight to a final volume of less than 5 mL. The concentrated D. Tok Pol was then applied to a Hiload Superdex 200 sizing column (Pharmacia) equilibrated in a buffer containing 40mM Tris pH 7.5, 50 mM $(NH_4)_2SO_4$. The elution volume of D. Tok Pol is consistent with a single species with molecule weight of about 90 kD. Fractions containing D. Tok Pol were confirmed by SDS-PAGE and pooled. The protein was then concentrated by ultrafiltration using a Biomax cell (Millipore) again with 90 kD cutoff. D. Tok Pol can easily be concentrated to ~ 200mg/mL without aggregation, as measured by absorbance at 280 nm using the theoretical extinction coefficient calculated from amino acid composition in D. Tok Pol sequence. The final product was flash frozen in ~50 μL aliquots for storage at 80°C. D. Tok Pol can be stored indefinitely in this manner. 8 L of *E. coli* culture typically yields 60 mg of purified D. Tok Pol.

1.2.2 Crystallization, Cryo-stabilization, and Heavy Metal Derivatization

Prior to crystallization, the D. Tok Pol sample is diluted to 20 mg/mL in the gel filtration buffer described above. Crystals of D.Tok Pol were obtained by standard vapor diffusion method at 20°C using a reservoir solution of 100 mM Tris-HCl (pH=8.6), 10 mM $MgSO_4$, 200 mM $(NH_4)_2SO_4$, 20% (v/v) 2, 4 methyl pentane diol (MPD), 11% (w/v) polyethylene glycol-4000 (PEG4K), 10 mM DTT. Equal volumes of protein and reservoir solutions were combined in the crystallization drop. Crystals grow readily in a couple of days, reaching a maximum dimensions of 200 μm x 150 μm x 100 μm (Figure 1.5). These crystals were cryo-stabilized in 100mM Tris 8.6, 10mM $MgSO_4$, 200mM Li_2SO_4 , 20% v/v MPD, 13% w/v PEG4K for 30 minutes and when shock-cooled in freshly thawed liquid propane (-180°C), diffracted synchrotron wiggler radiation (A1 beamline, Cornell High Energy Synchrotron Source) to Bragg spacings of 2.4 Å. D.Tok Pol crystallized in space group $P2_12_12_1$ with cell parameters (Native I crystal form: a=64.8 Å b=107.6 Å c=153.2 Å α=90° β=90° γ=90°). V_M calculations suggest one molecule per asymmetric unit with high solvent content (~65%). Native data sets recorded under these conditions resulted in unacceptably high non-isomorphism between frozen samples. Substitution of polyethylene glycol-400 (PEG400) for MPD in the crystallization and stabilization media resolved this problem and allowed structure determination by multiple isomorphous replacement (MIR) (Native II crystal form: a=66.1 Å b=107.6 Å c=155.9 Å α=90° β=90° γ=90°). Heavy metal derivatives were obtained by soaking Native II crystals in stabilizing solution containing 10mM heavy-atom compound for 24 hours. Two derivatives

Figure 1.5 Crystal of D. Tok Pol in a hanging drop.



of high quality were obtained (K_2PtCl_4 and tri-methyl-lead-acetate, respectively) this way, yielding peaks of 7 σ and higher in difference Patterson maps.

1.2.3 Data Collection and Phase Determination

X-ray diffraction data sets from a set of shock-cooled native and isomorphous heavy atom derivatives were recorded using synchrotron wiggler radiation (λ=0.908). Data from Native I crystals (prepared with MPD) extended to Bragg spacing of 2.4 Å with an R_{sym} = 4.6%. MIR analysis was conducted on Native II crystals (prepared with PEG400) which yielded data to ~2.6 Å. X-ray diffraction data were indexed, integrated and scaled using the HKL package (Otwinowski and Minor, 1997).

The positions of heavy atom were located manually by inspection of difference Patterson maps (Figure 2.6) and checked by cross-phased difference Fourier maps. Experimental phases were calculated using these sites with the program SHARP (De-La-Fortelle and Bricogne, 1997). In our hands, higher quality electron density maps were obtained by performing individual single isomorphous replacement (SIR) calculations in SHARP and combining the individual SIR phase sets using the program SIGMAA (CCP4., 1994; Read, 1986). The experimental phases were improved and extended by solvent flipping and negative density truncation as implemented in SOLOMON (CCP4., 1994). This procedure (SHARP/ SOLOMON) yielded electron density maps of sufficient quality to allow the entire D. Tok DNAP polypeptide to be traced unambiguously. This map was dramatically improved over a map calculated with MLPHARE/ SOLOMON (CCP4., 1994).

1.2.4 Model Building and Refinement

The initial molecular model was built into a 3.0 Å electron density map using the interactive molecular graphics program O (Jones et al., 1991). Model refinement was carried by conjugate gradient minimization, torsion angle dynamics, and tightly constrained atomic temperature factor refinement in the program CNS (Brünger et al., 1998). Refinement against the 2.6 Å Native II data set was interspersed with manual re-building of the model against σ_A weighted $2F_o$-F_c and F_c-F_c maps calculated by averaging structure factors of 10 models resulting from multi-start torsion angle dynamics (Brünger et al., 1998). The original electron density map remained a useful guide throughout the re-building process. Progress of the refinement was monitored by reductions in R_{free} (10% of the recorded reflections) (Brünger et al., 1998). Against the Native II data set, the model was refined to an R_{free}=29.5% and $R_{working}$=24.2%. The refinement was continued against the 2.4 Å Native I data set. A rigid body search in CNS with the 2.6 Å model yielded a clear solution that was refined as above. The final model was refined to an R_{free}=29.9% and $R_{working}$=25.3%, and the final model contains residues 1 to 756 with three disordered regions (residues 386-389, 665-676, 757, 772). The Native II model contains 6030 non-solvent protein atoms, 4 sulfate ions, 2 magnesium ions, and 116 water molecules. The Native I model contains 5,992 non-solvent protein atoms, 9 sulfate ions, 2 magnesium ions, and 106 water molecules. Model geometry was analyzed using the program PROCHECK (Laskowski et al., 1993). Both models have no outliers in the Ramachandran plot, with over 80% of the residues in the most favored region

(Figure 1.6). Coordinates of the final structure, refined using either the Native I data set (2.4 Å resolution) or Native II data set (2.6 Å resolution), have been deposited with the Research Collaboratory for Structural Bioinformatics (RCSB) under the accession code 1D5A (Native I) and 1QQC (Native II).

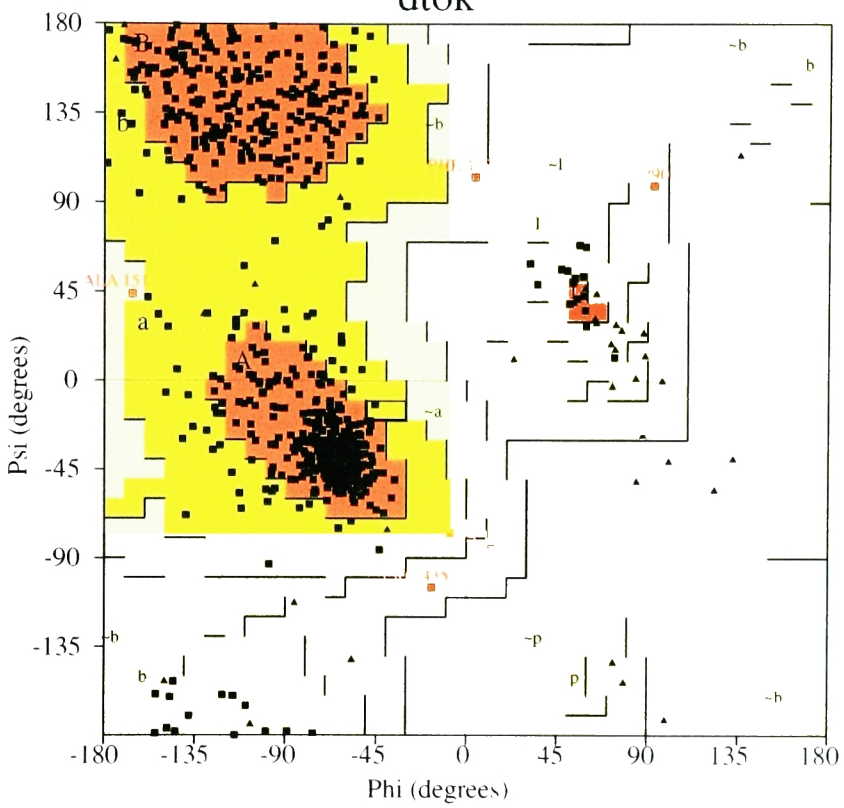
The R values for the final refined structures, although higher than 20% (R_{free} ~ 29.5% and $R_{working}$ ~ 25%), are not a necessary indication of any gross errors in the structure. In fact there are always certain effects in the process of structure solution that are not accounted for appropriately yet in the refinement process, such as disordered region with ambiguous electron density, multiple conformations of side chain, anisotropy etc, which doesn't respond to the crystallographer's effort of vigorous refinement and model rebuilding. While an R_{free} value higher than 35% is surely alarming, it is often acceptable to think that structures of large biomolecules and medium resolution with R_{free} value lower than or close to 30% are correct. In fact, R factors for the crystal structures of several other thermostable Pol II type DNA polymerases, which has high sequence identity to D. Tok (~70%), have displayed similar R values ($R_{free}/R_{working}$ = 31.3%/25% for DNA polymerase from *Pyrococcus kodakaraensis*, PDB code 1GCX and resolution 3.0 Å; $R_{free}/R_{working}$ = 28%/21% for DNA polymerase from RB69 bacteriophage, PDB code 1WAJ and resolution 2.8 Å; $R_{free}/R_{working}$ = 27%/21% for DNA polymerase from *Thermococcus gorgonarius*, PDB code 1TGO and resolution 2.5 Å).

Figure 1.6 Ramachandran plot of the refined D. Tok models. Both the Native I model and Native II model are checked. These plots are generated using the program PROCHECK (Laskowski et al. 1993).

PROCHECK

Ramachandran Plot

dtok



Plot statistics

Residues in most favoured regions [A,B,L]	558	85.3%
Residues in additional allowed regions [a,b,l,p]	92	14.1%
Residues in generously allowed regions [~a,~b,~l,~p]	4	0.6%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	654	100.0%
Number of end-residues (excl. Gly and Pro)	5	
Number of glycine residues (shown as triangles)	44	
Number of proline residues	37	
Total number of residues	740	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

dtok_01.ps

1.3 Results and Discussion

1.3.1 Structure determination

Crystals of D. Tok Pol have been obtained from MPD (Native I) and PEG 400 (Native II). Both crystal forms are orthorhombic ($P2_12_12_1$; a=64.8Å, b=107.6Å, c=153.2Å for Native I and a=66.1Å, b=107.6Å, c=155.9Å for Native II). Experimental phases (Table 1.2) to 3.0 Å were obtained from four isomorphous heavy atom derivatives, using the Native II crystal form and the program SHARP (De-La-Fortelle and Bricogne, 1997). Phases were improved by iterative cycles of real space density modification, consisting of solvent flipping and negative density truncation, using SOLOMON (Abrahams and Leslie, 1996; CCP4., 1994). The resulting electron density map allowed the chain to be traced unambiguously, with ready determination of sequence register. The model was refined to 2.6 Å against data for Native II (R value = 24.2%, R_{free} =29.5%,) and subsequently to 2.4 Å against data for Native I (R value = 25.3%, R_{free}=29.9%,), using CNS (Brünger et al., 1998). The model for Native II is somewhat more complete (see Methods) and is used for most of the discussion. This model includes 740 residues from 1 to 756 in Native II. Amino acids 386-390 and 665-676 are not visible in our electron density maps and are not included in the model.

1.3.1 General Description of the Structure

D. Tok Pol (Figure 1.7) is composed of a polymerase domain (residues 390 to 773) and an exonuclease domain (residues 133 to 385), as well as an N-

Table 1.2 Data Collection, Structure Determination, and Refinement Statistics

Table 1. Data Collection, Structure Determination, and Refinement Statistics

	Resolution (Å)	Number of Reflections (unique)	Completeness(%)[g]	R_{sym}[a,g](%)	R_{iso}[b](%)	Sites	Phasing Power[c,g]	Figure of Merit[d]
Native Data								
Native 1	50.0-2.6	32,909	93.9(57.8)	5.2(15.6)				
Native 2	50.0-2.4	40,540	92.2(53.8)	4.6(31.9)	57.4			
MIRAS Analysis								0.367
Pt	50.0-3.0	40,316	97.9(93.0)	8.4(22.9)	19.1	4	1.34(0.99)	0.214
Pb	50.0-3.0	34,905	84.2(70.4)	5.9(18.6)	13.6	1	1.16(0.98)	0.195
Pt/Pb	50.0-3.0	35,107	80.8(57.4)	9.9(21.9)	18.3	5	1.59(0.80)	0.221

Refinement		Number of Reflections ($\|F\| > 2\varsigma$)	$R_{working}$[e] / R_{free}[f] (%)	Total Number of Atoms	Rmsd. For bonds (Å)	Rmsd. For angles (Deg.)	Rmsd for B values (Å^2)
			24.2/29.5	6,167	0.008273	1.61591	1.691
Native 1	50.0-2.6	31,591	25.3/29.9	6,145	0.008273	1.50479	1.409
Native 2	50.0-2.4	37,229					

[a] $R_{sym}\% = 100 \times 6|\mathbf{I} - \langle \mathbf{I} \rangle| / 6\mathbf{I}$, where $\mathbf{I}$ is the integrated intensity of a given reflection.

[b] $R_{iso}\% = 100 \times 6|\mathbf{F}_{PH} - \mathbf{F}_P| / 6\mathbf{F}_P$, where $\mathbf{F}_{PH}$ and $\mathbf{F}_P$ are the derivative and native structure factor amplitudes, respectively.

[c] Phasing power $= 6|\mathbf{F}_H| / 6||\mathbf{F}_{PH}(obs)| - |\mathbf{F}_{PH}(calc)||$, where $\mathbf{F}_H$ is the calculated heavy atom structure factor amplitude.

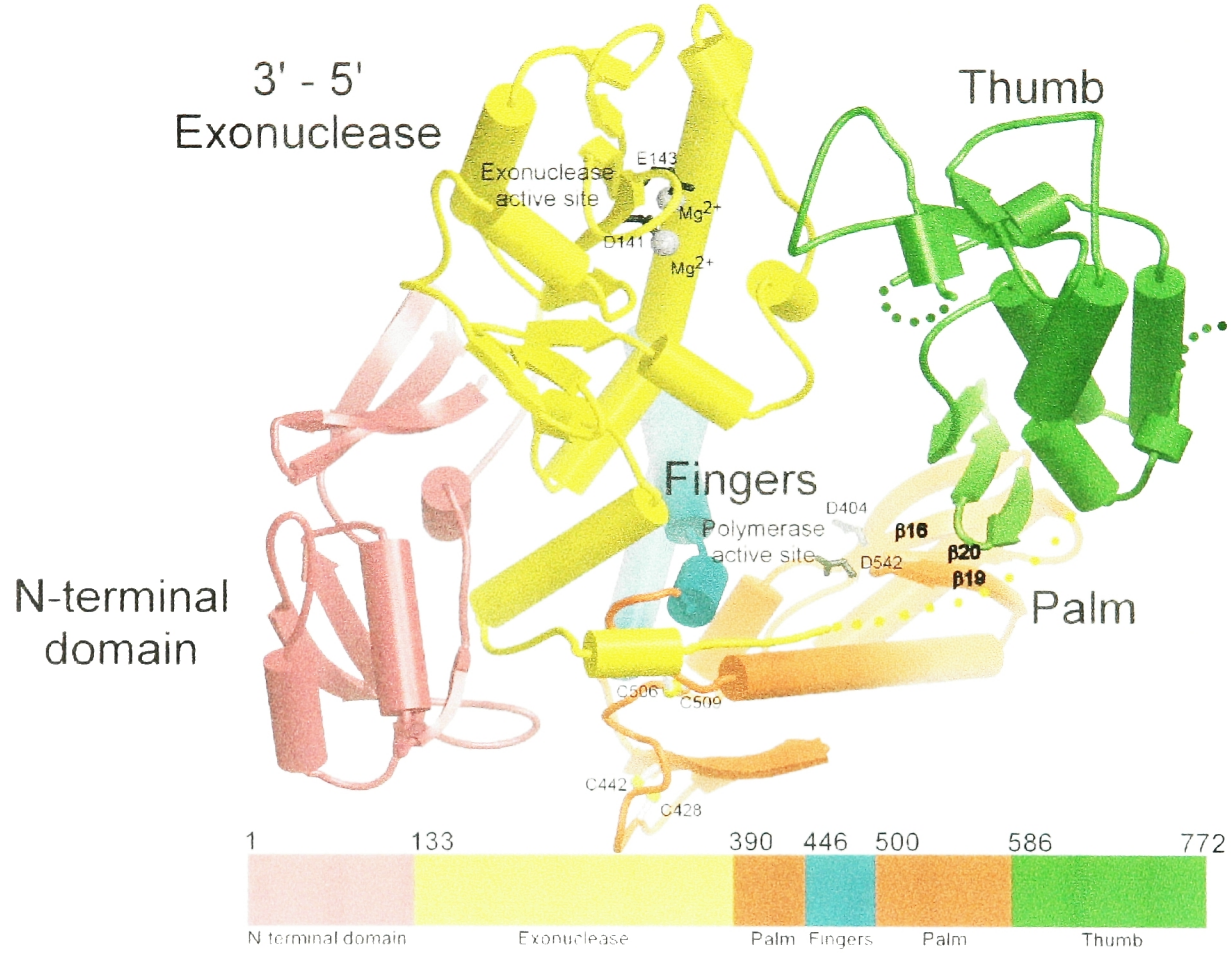
[d] Figure of merit $= \langle |6\mathbf{P}(\Delta)e^{i\Delta}/6|\mathbf{P}(\Delta)| \rangle$ where Δ is the phase and $\mathbf{P}(\Delta)$ is the phase probability distribution.

[e] R working $= 6||\mathbf{F}(obs)| - |\mathbf{F}(calc)|| / 6\mathbf{F}(obs)$

[f] R free $= 6|\mathbf{F}(obs) - \mathbf{F}(calc)| / 6\mathbf{F}(obs)$, calculated using 10% of the data.

[g] Number in parenthesis apply to the highest resolution shell

Figure 1.7 Structure of D.Tok Pol. The structure is represented by cylinders for helices, arrows for strands, and a thin worm for other secondary structural elements. Two gray spheres represent metal ions (presumed to be Mg^{2++}) observed bound to the exonuclease domain. The active site of the polymerase is marked by the location of two aspartate residues D404 and D542. The two disulphide bonds are indicated. Regions of the polypeptide chain that could not be modeled in the palm sub-domain due to disorder are indicated by dotted lines. The various domains and sub-domains and their boundaries are indicated in the bar.


3' - 5'
Exonuclease
Thumb
Exonuclease
active site
E143
Mg2+
D141
Mg2+
Fingers
D404
Polymerase
active site
D542
β16
β20
β19
N-terminal
domain
Palm
C506
C509
C442
C428
1
133
390
446
500
586
772
N terminal domain
Exonuclease
Palm
Fingers
Palm
Thumb

terminal domain (residues 1 to 131) that is not found in Pol I-type DNA polymerases (Brautigam and Steitz, 1998a). The polymerase domain is further comprised of three smaller sub-domains, termed thumb (residues 607 to 756), palm (residues 390 to 445 and 500 to 606), and fingers (residues 446 to 499). The structures of the MPD and PEG400 crystal forms of D. Tok Pol are very similar in terms of the individual subunits. The major difference between the two structures is a rotation of ~8-10^{o} in the orientation of the exonuclease domain with respect to the thumb sub-domain.

The domains of D. Tok Pol are arranged as an irregularly shaped flattened ring with a central cavity located near the polymerase active site. The mostly α helical thumb sub-domain forms one side of the active site cleft and makes contacts with the exonuclease domain (Figure 1.7). The structures of the thumb domains of various polymerases are often unrelated in structure. However, in all cases where structures are available the thumb domain is seen to fulfill an important role by forming contacts with duplex DNA as it exits the polymerase active site (Brautigam and Steitz, 1998a). The D. Tok Pol structure has been determined in the absence of DNA, and a portion of the thumb sub-domain which is likely to contact DNA (residues 665 to 676) is disordered. This is commonly observed for the corresponding regions of other polymerases in the absence of substrate (Jacobo-Molina et al., 1993; Jeruzalmi, 1998; Kim et al., 1995; Ollis et al., 1985; Sousa et al., 1993). In the DNA polymerases from bacteriophage T4 and RB69, the thumb sub-domains also provide a C-terminal element that

interacts with the processivity clamp (Berdis, 1996; Goodrich, 1997). In D. Tok Pol, the corresponding region (residues 757 to 773) is disordered.

The central region of the active site cleft is occupied by the palm sub-domain and includes residues important for substrate discrimination and the catalysis of the polymerase reaction. In D. Tok Pol, the palm is organized around three β strands (β16, β19, β20) flanked by an α helix (αQ) (Figures 1.7, 1.8A, 1.9A). It contains two disulphide bonds (Cys 428-Cys 442, Cys 506-Cys 509) that have not been previously observed in palm sub-domains and which may be important for thermostability (Figure 1.7).

The central elements of the palm sub-domains from polymerases belonging to the Pol I and Pol II families can be aligned closely (the RMS deviation in Cα positions for strands β16, β19, β20 and helix αQ is in the range of 0.9 2.0 Å), indicating a potential conservation in function. There are two residues in the palm domains of Pol I polymerases that are crucial for enzymatic activity because they coordinate two metal ions (Doublie et al., 1998; Joyce and Steitz, 1994; Li Y et al., 1998; Steitz et al., 1994). The corresponding residues in D. Tok Pol are Asp 404 and Asp 542 (Figure 1.7). No metal ions are, however, visible in our electron density maps.

The fingers sub-domain in D. Tok Pol consists of a set of anti-parallel α helices (Figure 1.8, αN, αO, αP). These helices are shorter in length than the corresponding elements of RB69 Pol, and a helical segment which connects helices O and N in RB69 Pol is missing altogether (Figure 1.9). The fingers domain of D. Tok Pol is unrelated in overall structure to that of Pol I-type

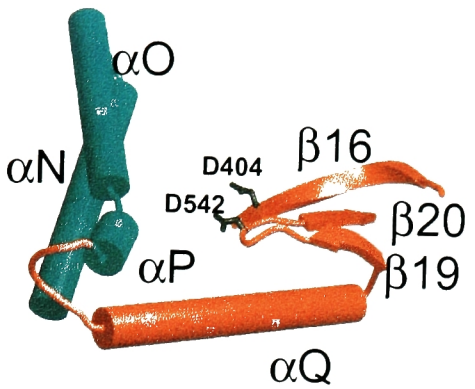
polymerases (Figure 1.8). However, helix αP in D. Tok Pol is positioned similarly to helix O in Pol I polymerases (Figure 1.8), and is likely to play an analogous and crucial role in recognition of the incoming nucleotide (Doublie et al., 1998; Kiefer et al., 1998; Li Y et al., 1998; Ollis et al., 1985)

The 3'-5' exonuclease domain in D. Tok Pol is located opposite the thumb sub-domain and above the fingers sub-domain, as noted for RB69 Pol. It contains two metal ions (presumably Mg^{++}) ligated to Asp 141 and Glu 143 (Figures 1.10). The position of this domain relative to the polymerase active site is distinct from the arrangement seen in Pol I-type polymerases. The conservation between RB69 and D. Tok Pol of the location of the exonuclease domain suggests that this is a characteristic feature of Pol II-type polymerases. The structure of the D. Tok 3'-5' exonuclease domain resembles those associated with other DNA polymerases (Brautigam and Steitz, 1998b; Kiefer, 1997). The 3'-5' exonuclease domains from the Pol I (*E. coli*, *T aquaticus*, *B. subtilis*, bacteriophage T7) or Pol II (RB69) polymerase families can be aligned onto each other closely (RMS deviation in Cα positions for strands β10, β11, β12, β14 and helices αE and αI is in the range of 1.0 2.8 Å). This alignment superimposes in a satisfactory manner residues associated with substrate binding and catalysis and metal binding (Figure 1.10) (Brautigam and Steitz, 1998b).

The arrangement of the N-terminal, exonuclease, and polymerase domains creates two deep grooves leading into and out of the polymerase active site. The D groove (for duplex DNA binding), following the nomenclature of

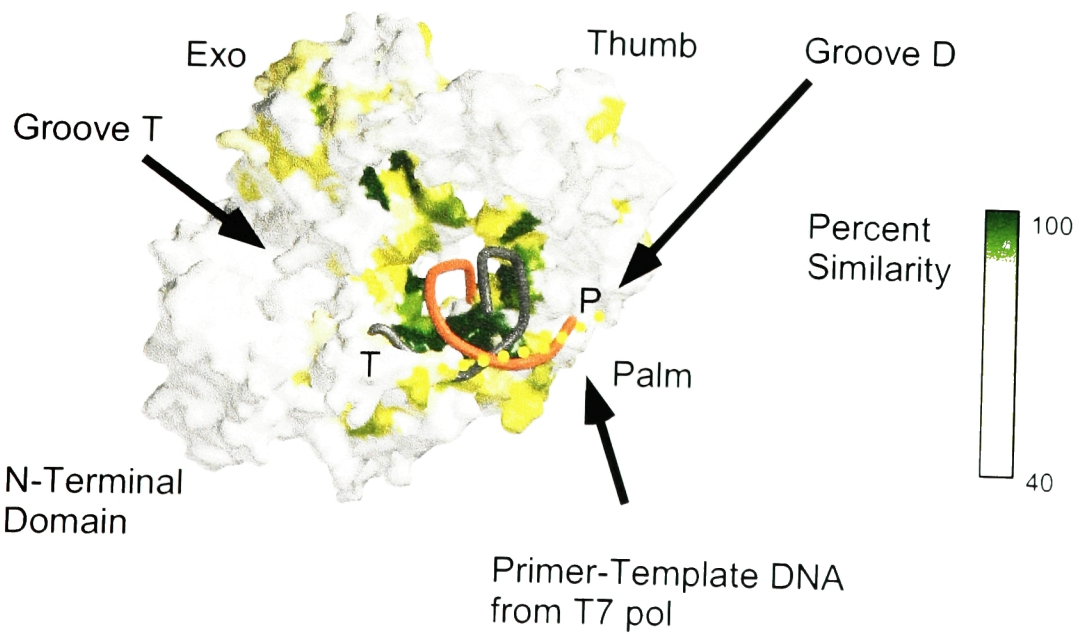
Figure 1.8 Comparison of D. Tok Pol and T7 DNA Pol Structures. (A) A view of the secondary structural elements of polymerase active site region (palm and fingers sub-domains) of D. Tok pol, colored as in Figure 1.9. (B) The corresponding region of T7 DNA polymerase including the primer template duplex from the crystal structure (PDB entry=1T7P, (Doublie et al., 1998)). The orientation of T7 Pol was derived by superposition onto strands β16, β19, β20 of D. Tok pol. D. Tok pol helix αP is seen to be in an analogous position relative to the active site aspartates as T7 pol αO. (C) A GRASP surface representation of D. Tok pol with modelled primer template duplex from the T7 DNA polymerase – DNA complex (PDB entry=1T7P, (Doublie et al., 1998)). The surface is colored according to sequence similarity (40% to 100%) calculated as in Figure 1.11. The primer strand is an orange worm representing phosphate positions, and the template strand is in gray.

A) D.Tok Pol



B) T7 DNA Pol

C) D.Tok Pol



(Wang et al., 1997b), is located immediately below the thumb sub-domain and includes a region of positive electrostatic potential. The T groove (for template DNA binding) leads away from the active site in the opposite direction and is located below the fingers sub-domain. A small channel (the editing channel) leads from the polymerase domain to the exonuclease active site (Figures 1.8C).

At the time this structure was determined there was no structural information available for DNA complexes of Pol II family polymerases. We have used the structure of T7 Pol (a Pol I family member) bound to primer-template DNA to model DNA onto D. Tok Pol (Figure 1.8). Superposition of the palm sub-domains of the two polymerases shows that remarkably few bad contacts are formed between the DNA (from T7 Pol) and atoms in the D. Tok Pol model. The one region that does collide with the DNA is the segment connecting the exonuclease and polymerase domains. This region (residues 377 to 390) is partially disordered in the D. Tok Pol structures, and is likely to reorganize upon binding DNA. This superposition allows 5 base pairs of DNA to be accommodated in the D. Tok active site with the formation of DNA-protein contacts. The formation of contacts with additional base pairs would require a change in the position of the thumb sub-domain in the region of the D groove. A change in the conformation of the fingers sub-domain (helices αO and αP) is also required to position residues Lys 487 and Tyr 493 (or Tyr 494) of D. Tok Pol (Figure 1.8) for interaction with the incoming nucleotide, by analogy with the T7 Pol structure (Doublie et al., 1998). Finally, the superimposed primer-template DNA is well positioned so that the incoming template strand will likely reside in

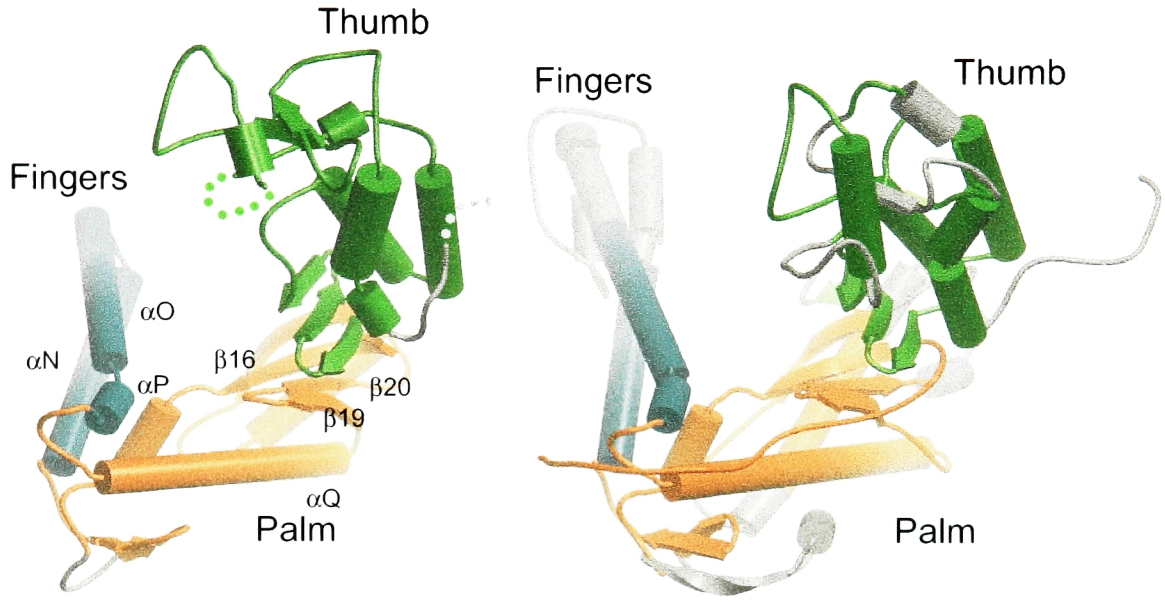
the T groove. Superposition of the DNA molecule derived from the structure of HIV-1 reverse transcriptase complexed to DNA (Huang et al., 1998) leads to similar conclusions.

1.3.2 Comparison between D. Tok Pol and RB69 Pol

Although the DNA polymerases from D. Tok. and bacteriophage RB69 share less that 20% primary sequence identity (Figure 1.11), their structures resemble each other closely (Figure 1.9). Not surprisingly, the regions of highest sequence similarity are concentrated in and around the exonuclease and polymerase active sites (Figures 1.8C, Figure 1.11). Despite the low overall sequence identity, the individual sub-domains in the two structures superimpose well (the RMS deviation in Cα positions in the fingers, thumb and palm sub-domains is in the range of 0.8 to 1.5 Å). Moreover, the overall arrangement of domains and sub-domains with respect to each other is preserved in the two polymerases, strengthening the proposal that Pol II DNA polymerases share a common architecture (Figure 1.9).

One difference between the overall structures of D. Tok Pol and RB69 Pol concerns the orientation of the exonuclease domain with respect to the rest of the structure. When the two polymerases are superimposed on their respective palm sub-domains it is seen that the exonuclease domain of RB69 is rotated inwards by ~8°, burying the active site in a solvent inaccessible configuration (Wang et al., 1997b). In contrast, the exonuclease domain in D. Tok Pol has its active site essentially exposed to solvent. It is possible that conformational changes between open and closed configurations of the exonuclease domain are

Figure 1.9 Comparison of RB69 and D. Tok Pol structures. The structures of A) D. Tok pol and B) RB69 Pol are presented in the same orientation after superposition of their respective palm sub-domains. Structural elements that are in common between the structures are represented and colored as Figure 1.9. Elements that are unique to RB69 pol are colored in gray. Disordered segments are indicated by dotted lines. The N-terminal and the exonuclease domains are not shown.



DTOK

RB69

Figure 1.10 Structural Alignment of Exonuclease Domains. Structures of exonuclease domains from KF, 1WAJ, 1T7P, 1BDF, 1TAQ have been aligned by superimposing residues 137:145, 158:164, 167:172, 205:220, 257:260, 303:313 which represent strands β8, β9, β10, β12, β15 and helix αE αI. A color gradient is used to depict the average RMS deviation for the family of superimposed structures ranging from blue (1.0 Å - 1.5 Å) to white (> 4.0 Å). Residues conserved amongst exonuclease sequences and implicated in catalysis are drawn in a ball and stick representation in green. Two gray spheres represent two metal ions bound at the active site. The active site is also indicated by a tetranucleotide (in gold) derived from superposition of the exonuclease domain from the RB69 Pol structure.

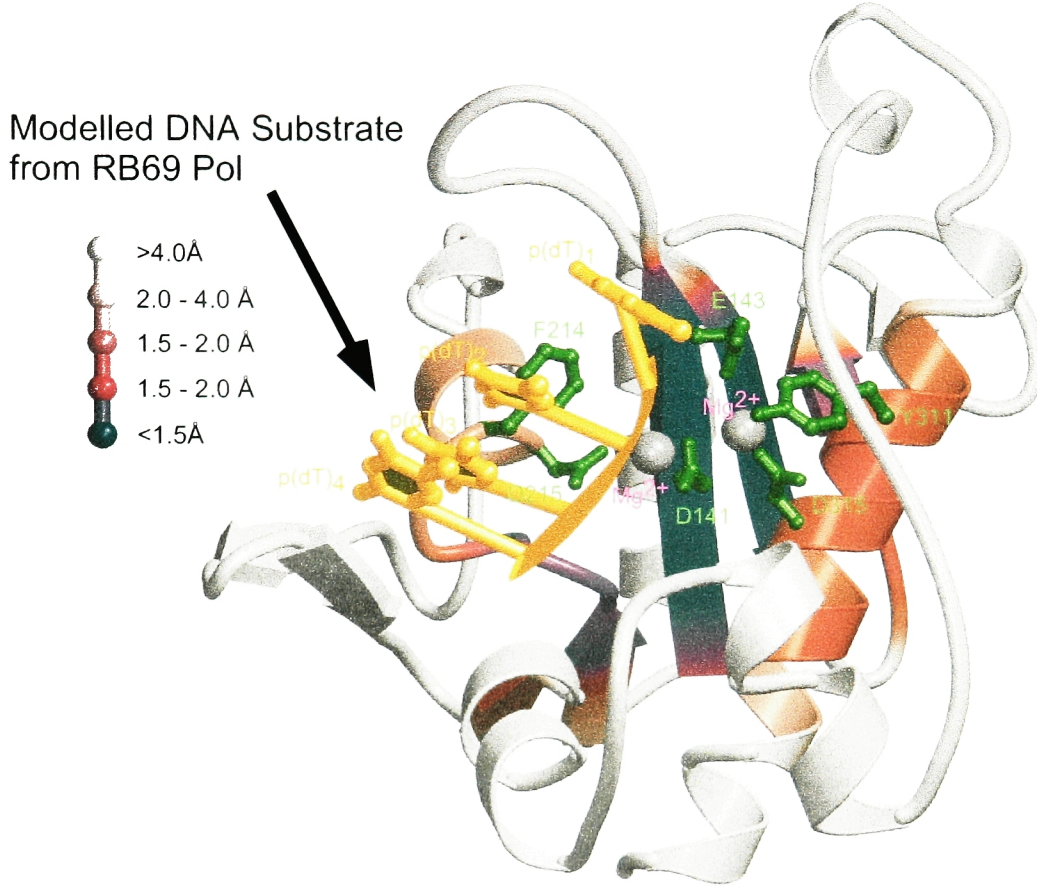
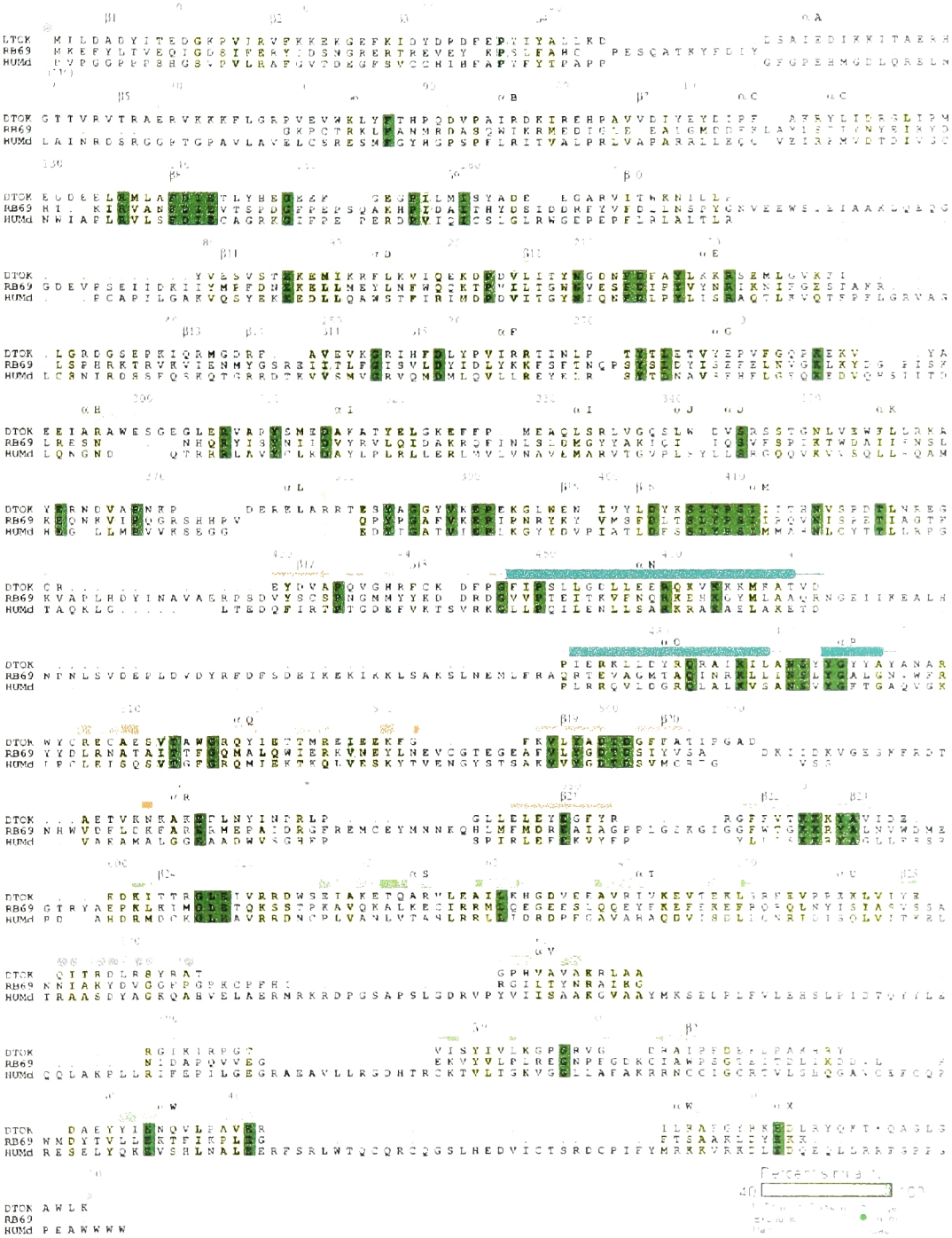

Modelled DNA Substrate
from RB69 Pol
>4.0Å
2.0 - 4.0 Å
1.5 - 2.0 Å
1.5 - 2.0 Å
<1.5Å
p(dT)1
E143
F214
Mg2+
p(dT)4
D141

Figure 1.11 Structure-based sequence alignment for D. Tok Pol, RB69 Pol and Human Pol *δ*. The alignment is colored by sequence similarity (40%, white to 100%, green) calculated as described in Figure 1.10(C). Shown here is a small subset of a larger set of sequences that were used to generate the alignment. The respective secondary structural elements colored as in figure 1.9 are represented with helices as cylinders, strands as arrows, and other as thin lines. Gray circles represent portions of the polypeptide chain that could not be modeled.

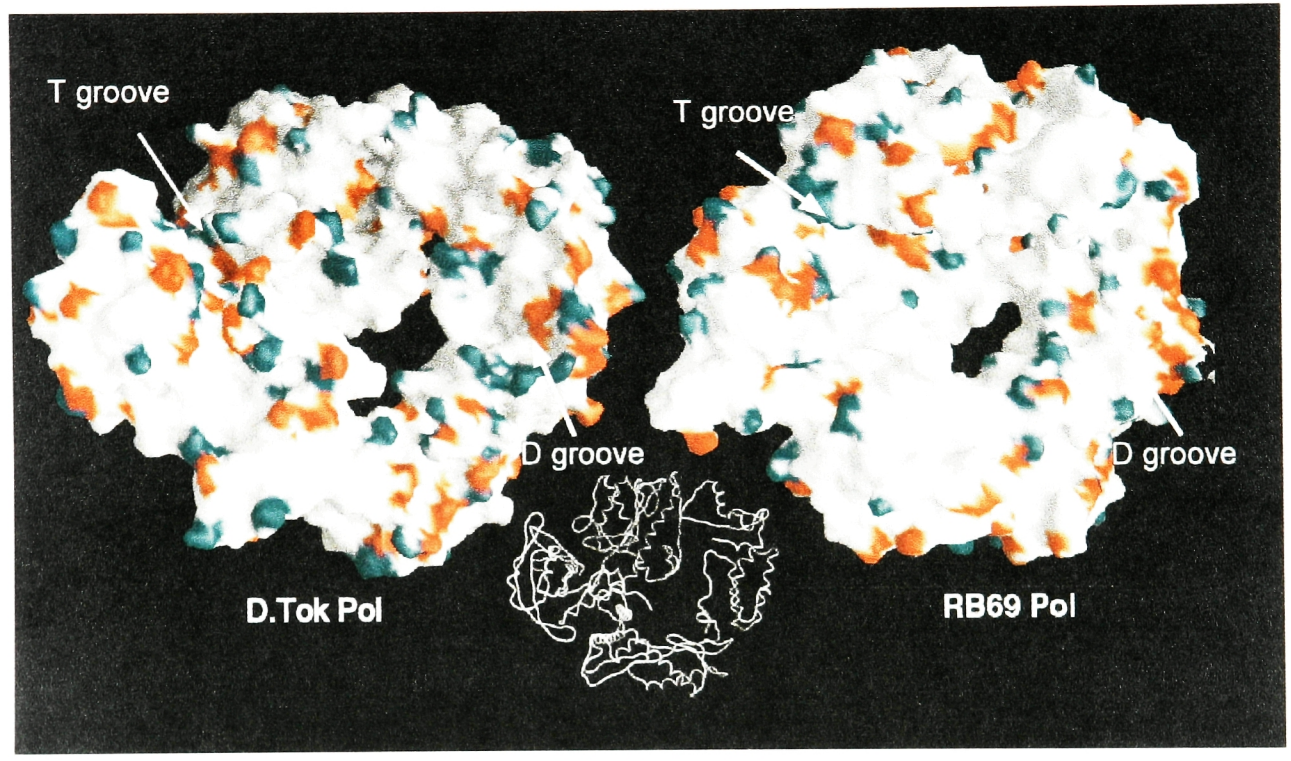


a part of the functional cycle of the protein, particularly since the two different forms of D. Tok Pol differ in the orientation of the exonuclease domain (not shown).

One interesting difference between D. Tok Pol and RB69 Pol is that the former is a thermostable DNA polymerase while the latter is not. Unfortunately, attempts to identify features in the D. Tok Pol structure that might be correlated with thermostability is complicated by the very low sequence similarity between the two enzymes. One feature that does stand out, however, is the increased formation of arrays of ionic interactions on the surface of D. Tok Pol when compared to that of RB69 Pol (Figure 1.12). The formation of networks of ionic interactions has been noted to correlate with thermostability in other proteins (Hennig et al., 1995; Korolev et al., 1995; Yip et al., 1995).

Generally, D. Tok Pol sub-domains tend to be more compact, with smaller helices and shorter loops than found in RB69 Pol, a feature that may be another important source of thermostability. For example, the palm sub-domain displays close structural conservation of elements near the catalytic aspartate residues. However, helix αR in D. Tok Pol is much shorter that its counterpart in RB69 Pol, and a small substructure in front of the palm sub-domain is entirely missing in D. Tok Pol (Figures 1.9, 1.11). Deletion of these elements is also seen in a representative set of archaebacterial DNA polymerases (Kong et al., 1993; Uemori et al., 1993). Likewise, the fingers sub-domain is missing a large mass from its tip in D. Tok Pol (Figures 1.9, 1.11). However, the RB69 fingers extension most likely plays a T4 phage-specific role since it is also missing from

Figure 1.12 Comparison of surface charges in D Tok Pol and RB69 Pol. Accessible surface representation of D. Tok pol and RB69 pol in the same orientation after superposition of their palm sub-domains. Surface regions corresponding to the terminal oxygen atoms of aspartate and glutamate are colored red. Likewise, surface regions contributed by the sidechain nitrogen of lysines and arginines are colored blue. D. Tok Pol has a striking pairing of oppositely charged residues not seen in RB69 pol. A representation of D. Tok Pol as a worm is included to orient.


T groove
D groove
D.Tok Pol
T groove
D groove
RB69 Pol

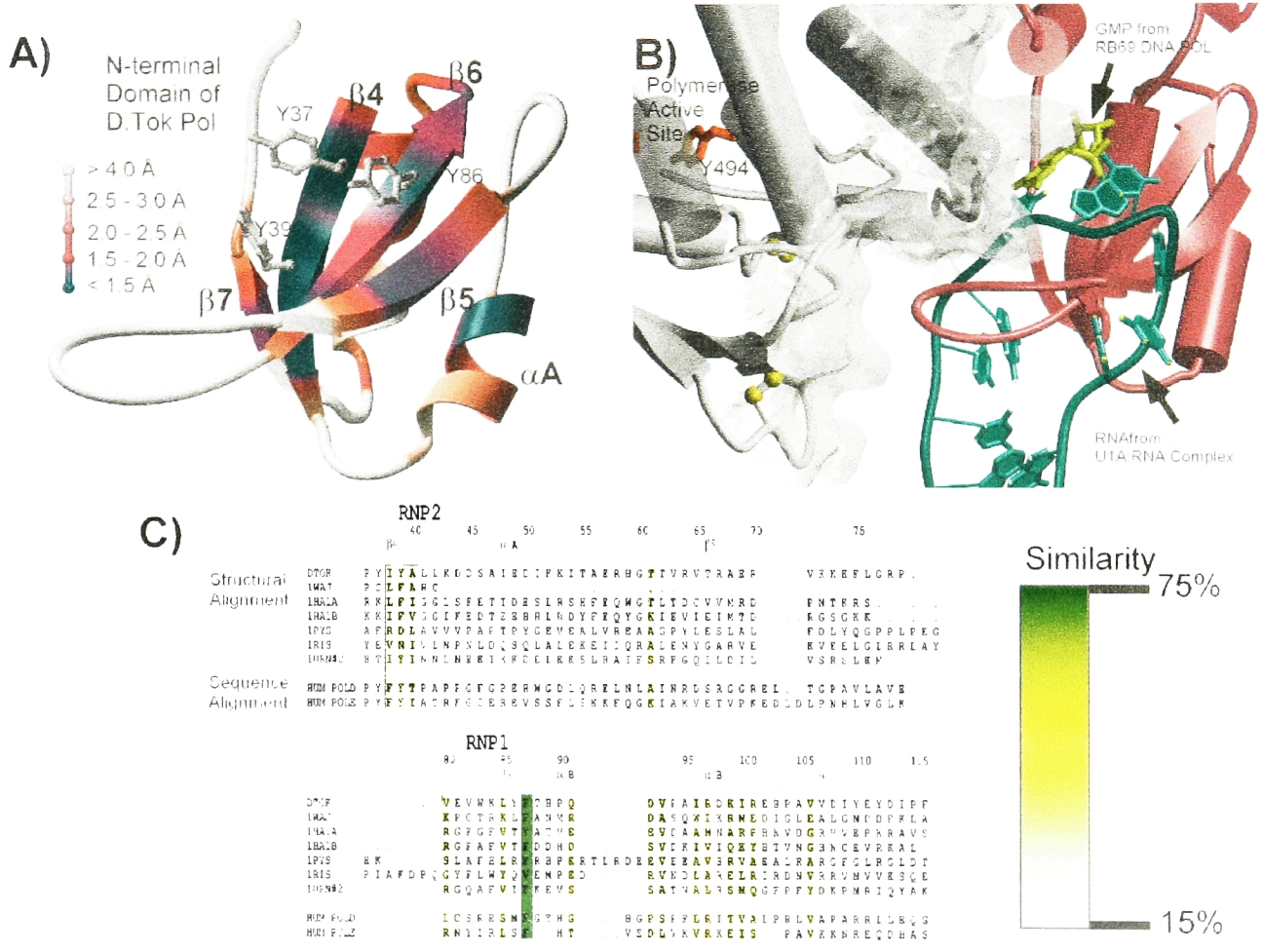
our alignments of archaebacterial DNAPs and eukaryotic polymerases δ(Figure 1.11).

1.3.3 The N-terminal domain resembles RNA Binding Domains

The N-terminal domain of D. Tok DNA polymerase has no corresponding element in Pol I-type polymerases. Analysis of the structure of this domain using DALI (Holm and Sander, 1993) (http://www.embl- ebi.ac.uk/dali/) revealed a previously unsuspected similarity to RNA binding domains (RBD, also known as RNA Recognition Motif, RRM). RBDs are small modules (80-90 residues) found in RNA binding proteins of prokaryotes, archaea, and eukaryotes (reviewed in (Varani and Nagai, 1998)). These modules adopt a conserved *βαββαβ* architecture and bind to single stranded RNA. Two conserved sequence motifs, referred to as RNP1 and RNP2, provide aromatic and charged residues that are important for RNA recognition (Oubridge et al., 1994) (Figure 1.13).

The N-terminal domain of D. Tok Pol can be superimposed closely onto the core secondary structural elements of RBDs from the U1A spliceosomal protein (Oubridge et al., 1994), ribosomal protein S6 (Lindahl et al., 1994), the hnRNP proteins (2 RBD domains) (Shamoo et al., 1997; Xu et al., 1997) and the anticodon binding domain from *Thermus thermophilis* phenylalanyl-tRNA synthetase (Goldgur et al., 1997). The RMS deviations in C*α* positions for these superpositions are in the range of 0.5 – 2.0 Å (Figure 1.13A). Differences between the structures of the loops in the N-terminal domain of D. Tok Pol and those of the RNA-binding domains are within the range of structural variation seen in the various RNA binding domains.

Figure 1.13 The N-terminal Domain of D. Tok Pol. (A) Structural conservation of RNA binding domains: Structures of RNA binding domains from 1HA1 (domains A & B), 1PYS, 1RIS, and 1URN (molecule 2) have been aligned by superimposing (LSQMAN, SUPERPOSE D. Tok Pol residues 40:110 which represent four strands (β4, β5, β6, β7) and two helices (αA and αB). The N-terminal domain of D. Tok Pol is shown. A color gradient is used to depict the average RMS deviation in Cα positions for the family of superimposed structures, ranging from blue (1.0 to 1.5Å) to white (>4.0Å). Certain aromatic residues in D. Tok Pol (white) are shown; these represent a potential RNA binding surface. This view is rotated by approximately 180° from that in Figure 1.9. (B) An RNA stem-loop from the U1A-RNA complex (PDB entry=1URN, (Nagai *et al.*, 1990)) modeled onto the N-terminal domain of D. Tok Pol. The model was generated by superimposing the U1A RBD onto the N-terminal domain of D. Tok Pol using the conserved structural elements. The RNA is drawn in blue with the sugar-phosphate backbone represented as a worm and the bases as a ball and stick representation. A partial surface that represents the interface between the N-terminal domain of D. Tok Pol and the exonuclease domain is shown in gray. The location of the modeled RNA relative to the polymerase active site is depicted by marking the position of residue Y494. The location (derived after superposition) of the guanosine monophosphate (GMP) molecule bound to the 'incomplete' RBD of RB69 Pol, drawn in light green, nearly overlaps with the positions of the bases of the modeled RNA stem-loop. (C) Structural and primary sequence alignment of RNA binding domains: Sequence alignment of the N-terminal domains from D. Tok Pol and RB69 Pol (incomplete domain) and the RBDs from 1HA1 (domain A and B), 1PYS, 1RIS, 1URN (molecule #2) superimposed as in (A). Alignments of the N-terminal domain of D. Tok Pol against DNA polymerase δ and ϵ were obtained using CLUSTALX (Thompson *et al.*, 1994) using its default parameters. The conserved primary sequence motifs RNP1 and RNP2 are boxed. The alignment is colored by sequence similarity (15%: white to 75%: green) calculated by averaging the similarity scores at each position of all possible pairs of sequences (D. J., unpublished software). Equivalence of non-identical residues was established by use of the BLOSUM62 amino acid substitution matrix (Henikoff *et al.*, 1992). Secondary structural elements corresponding to the N-terminal domain of D. Tok Pol are represented (magenta) with helices as cylinders, strands as arrows, and other as thin lines. Numbering of residues and naming of secondary structural elements is that of D. Tok Pol.


A)
N-terminal
Domain of
D.Tok Pol
> 4.0 Å
2.5 - 3.0 Å
2.0 - 2.5 Å
1.5 - 2.0 Å
< 1.5 Å
β4
β6
Y37
Y86
Y39
β7
β5
αA
B)
Polymerase
Active
Site
Y494
GMP from
RB69 DNA POL
RNAfrom
U1A RNA Complex
C)
RNP2
RNP1
Structural
Alignment
Sequence
Alignment
Similarity
75%
15%

There is no evidence at present to suggest that the N-terminal domain of D. Tok Pol binds RNA. However, comparison with the structures of RNA complexes of RNA binding domains shows that the N-terminal domain might in fact be a functional RNA binding domain (Figure 1.13). In particular, three aromatic residues in the N-terminal domain (Tyr 37, Tyr 39 and Tyr 86) could interact with RNA bases in a manner similar to that seen in crystal structures of RNA bound to RNA binding domains (Oubridge et al., 1994) (Figure 1.13). Interestingly, these residues are located near the position of a guanosine triphosphate molecule that is found bound to the N-terminal domain of RB69 Pol (Wang et al., 1997b) (Figure 1.13B). The DNA polymerases from bacteriophage T4 and its distant relative bacteriophage RB69 bind specifically to the ribosome binding site of their own mRNA, repressing its translation (Pavlov and Karam, 1994; Tuerk et al., 1990; Wang et al., 1997a). The N-terminal domains of T4 Pol and RB69 Pol are smaller than that of D. Tok Pol. In the RB69 Pol structure, the N-terminal domain appears to form an "incomplete" RNA binding domain (Figure 1.13C).

There is no significant overall sequence similarity between the N-terminal domain of D. Tok Pol and RNA binding domains, which is why the presence of this fold was not recognized previously (Figure 1.13C). Comparison of the sequences of other archaebacterial DNA polymerases and human polymerases δ and ϵ suggests that a corresponding structural element is likely to be found in these polymerases as well (Figure 1.13C). The sequence alignment in this region is unambiguous for the archaebacterial DNA polymerases. For eukaryotic

polymerases the alignment is less certain, but it appears to conserve the essential aromatic character of the RNP motifs (Figure 1.13C). Confirmation of the presence of these domains along with their ability to bind RNA, and their precise role in eukaryotic DNA synthesis awaits future structural and functional studies.

1.3.4 Conclusion

The 2.4 Å structure of D. Tok Pol revealed an overall architecture that retains the common right-hand analogy in its polymerase domain as seen in other DNA polymerase structures (Steitz et al., 1994). The similarity between D. Tok Pol and RB69 Pol suggests that these two structures are representative of a common Pol II polymerase fold. Members of this family carry out chromosomal DNA replication in eukaryotes, including humans, and yet there is no structural information available for any eukaryotic member of this family. The D. Tok Pol structure reported here, along with the RB69 Pol structure, should now make it possible to generate reliable structural models for eukaryotic DNA polymerases.

Pol I and Pol II families share very limited sequence identity and carry out different cellular functions. The structure of D. Tok Pol revealed that in the immediate vicinity of the central catalytic region of the polymerase domain there is close resemblance between these two families, indicating a common origin in evolution. But the overall architecture and the location of 3'-5' exonuclease domain is strikingly different. Modeling of DNA substrate onto D. Tok Pol using known structure of the Pol I type T7 DNA Polymerase ternary complex with primer-template DNA and incoming nucleotide (Doublie et al., 1998) confirmed

that the modes of binding of the template and extrusion of newly synthesized duplex DNA are likely to be similar in both Pol II and Pol I type DNA polymerases. However, the mechanism by which the newly synthesized product transits in and out of the proofreading exonuclease domain has to be quite different. The discovery of a domain that seems to be an RNA-binding module raises the possibility that Pol II family members interact with RNA.

A recently published structure of RB69 DNA Pol complexed with primer-template DNA and deoxynucleoside triphosphate (dNTP) at its polymerase active site provided a detailed picture of the state of the enzyme just before primer extension (Franklin et al., 2001). This ternary structure showed a similar conformational change in the fingers and thumb domains of RB69 Pol to form a more closed structure upon DNA and nucleotide binding as seen in that of Pol I type DNA polymerases complexed with DNA substrate (Doublie et al., 1998; Kiefer et al., 1998; Li et al., 1998). This behavior of alternating between open and close conformation during different catalytic steps as observed both for Pol I (Doublie et al., 1998; Li et al., 1998) and Pol II family suggests a preserved common functional mechanism that is possibly shared by all DNA polymerases.

Chapter 2. Molecular Dynamics Studies of Sliding Clamp Opening

2.1 Introduction

2.1.1 Sliding Clamp Proteins

Most DNA polymerases are distributive enzymes that stop and dissociate from the template after a few rounds of synthesis. The DNA polymerases that are specialized for chromosomal replication overcome this inefficiency by utilizing processivity factors. These are distinct subunits of the polymerase assembly (Kornberg and Baker, 1991). By forming a complex with their associated processivity factors and moving as one entity along the replication fork, DNA polymerases can synthesize thousands of bases without dissociation from the template. In prokaryotes, eukaryotes, archaebacteria and T4-related bacteriophages the processivity factors include a sliding clamp protein and a multiprotein clamp loader complex. The bacterial sliding clamp protein is referred to as the β subunit and it exists as a dimer. The eukaryotic sliding clamp is a trimer, and is called proliferating cell nuclear antigen (PCNA) because of its initial discovery as a cell cycle dependent antigen (Miyachi et al., 1978). Archaebacterial sliding clamp proteins are named PCNA too because of their close similarity to the eukaryotic proteins. Bacteriophage T4 also utilizes a trimeric sliding clamp, which is named gene protein 45 (gp45) and was the first processivity factor for which the term "sliding clamp" was used (Huang et al., 1981). It is striking that all these proteins have similar structures although

sequence similarity between these proteins from different organisms is insignificant.

Processivity factors confer upon the DNA polymerase the ability to move rapidly along DNA without dissociation. Biochemical studies in *E. coli* showed that the β subunit is able to move in both directions on duplex DNA in an ATP-independent fashion (Stukenberg et al., 1991). The half life of β dimer on circular DNA is over 1 hour (Yao et al., 1996). This picture of a sliding clamp suggests that processivity factors probably function as a stable and mobile platform encircling DNA to which the polymerases are attached. The C-terminal segments of the polymerases are crucial for interaction with the processivity factors (Berdis, 1996; Goodrich, 1997; Shamoo and Steitz, 1999). The experimental data suggest that the polymerases interact with the processivity factor in a fashion that does not hinder its DNA-encircling characteristics, so that the complex of these proteins is tightly associated with DNA, yet free to move along.

Crystallographic studies carried out in our lab and in others on different sliding clamps and their complexes with other proteins or peptides have elucidated fascinating details regarding the functional mechanism of these sliding clamps, in good agreement with previous biochemical work. The first clamp structure, the *E. coli* β subunit (Kong et al., 1992) revealed a circular protein with a central hole large enough (~35 Å internal diameter) to accommodate double-stranded DNA without steric hindrance (Figure 2.1). Although the overall charge of the protein is negative, the distribution of charged residues within the protein is

Figure 2.1 Ring-Shaped Sliding Clamps. (A) Ribbons representations of the polypeptide backbones of PCNA and β subunit, with hypothetical duplex DNA. In this schematic representation individual monomers within each ring are distinguished by different colors. (B) A hypothetical model of duplex B-form DNA is placed in the geometrical center of each structure to illustrate the hypothesis that the rings encircle duplex DNA (Reproduction from Krishna et al. 1994) (Permission from John Kuriyan).

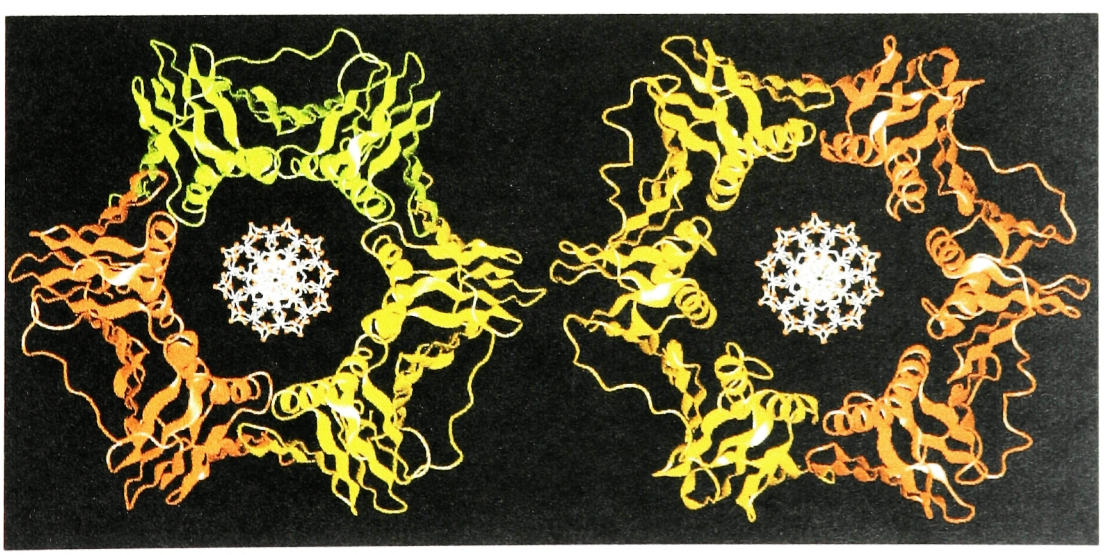
A)

E. coli Beta Subunit

Yeast PCNA

T4 gp45

B)



Yeast PCNA

E. coli Beta Subunit

such that the region inside the ring is positively charged and probably contacts the DNA phosphate backbone through nonspecific, water-mediated interactions. This allows the ring to slide in a sequence-independent manner on DNA, providing a DNA-bound platform for the polymerase to attach to. Two semi-circular β monomers form a tightly closed ring that can remain stably situated on circular DNA (Figure 2.1). The β dimer is formed in a head-to-tail fashion, and its two inter-subunit interfaces are stabilized by multiple interactions. Hydrophobic packing between two helices, one on each monomer, together with dimer formation buries about 1400 $Å^2$ surface at each interface (Kong et al., 1992). There are six potential ion pairs at each interface, which are expected to further stabilize the interface, together with hydrogen bonding between two β strands, one on each monomer.

Each monomer in the β dimer is composed of three domains of identical topology, each containing 1/3 of the total 366 residues (domain 1: residues 1-122, domain 2: residues 123-244, domain 3: residues 245-366) (Figure 2.1). Two inter-domain loops in each monomer connect the C-terminal residue of the preceding domain to the N-terminal residue of subsequent domain, i.e. domains 1 to 2, and domains 2 to 3, on the outside surface of the ring structure (Figure 2.1). The β dimer thus displays a pseudo 6-fold symmetry arising from its 6 similar domains in two monomers (Figure 2.1). Each individual domain in a monomer is again composed of two similar β-α-β-β-β structural units. The β dimer is thus made up by 12 such basic structural units situated adjacent to each other to form a closed ring. The arrangement of these twelve units is such that

the 12 α-helices line the inner surface of the ring and the 12 β sheets, each composed of four neighboring β strands, line the outer surface of the ring. This high symmetry in β dimer structure is not indicated by its protein sequence.

The structures of several other sliding clamps from different organisms, including the β subunit from *Streptococcus pyogenes* (Maria Argiriadi, unpublished), gp45 from bacteriophage *RB69* (Shamoo and Steitz, 1999), PCNA from yeast (Krishna et al., 1994), and PCNA from humans (Gulbis et al., 1996), and archaebacteria (Matsumiya et al., 2001), all form a toroidal shape with a central cavity large enough to accommodate DNA in a pattern highly similar to *E. coli* β despite no significant sequence similarity between these proteins (Figure 2.1). For the trimeric clamps, three monomers (each 2/3rds the size of the β monomer) are required to form a closed ring of about same dimension as that of *E. coli* β (Figure 2.1). Each monomer in the trimeric clamps is composed of two topologically identical domains with the same two-fold repeat of the β-α-β-β-β motif as in *E. coli* β dimer and one inter-domain connector loop resulting in the same pseudo 6-fold symmetry.

Although all processivity factors are highly negatively charged, the charge distribution is such that the center of the ring is positively charged, so as to complement the negative charges on DNA. Despite the overall striking similarities among structures of different processivity factors, there are abundant structural differences upon close examination. For example the gp45 structure was found to have much less structural regularity compared to β and PCNA,

giving it a triangular appearance rather than the hexagonal shapes common to *E. coli* β and PCNA (Moarefi et al., 2000) (Figure 2.1). Interfacial interactions were also found to be most extensive in *E. coli* β, less in trimeric PCNA, which lacks an extensive ion pairing network at the interface, and even less in gp45, which has fewer hydrogen bonds at the interface, with about half of the surface area buried as compared to PCNA (Moarefi et al., 2000). This may correlate to the fact that these oligomeric clamps vary in their ability to maintain their closed form in solution or on DNA, with *E. coli* being the strongest (Kd < 60pM, $t_{1/2}$ on DNA ~ 1 hr), PCNA weaker (Kd ~ 21 nM, $t_{1/2}$ on DNA ~ 24 min) and gp45 the weakest (Kd ~ 250 nM, requires DNA polymerase to form a stable complex on DNA) (Yao et al., 1996).

Besides their primary role in conferring processivity to DNA polymerases, sliding clamps are also involved in several other cellular mechanisms, such as DNA replication postprocessing, DNA repair and cell cycle regulation. Sliding clamps interact with many proteins involved in these diverse processes. In prokaryotes, the β subunit interacts with DNA ligase, DNA Pol I and the repair protein MutS that is involved in DNA mismatch repair (Lopez de Saro and O'Donnell, 2001). In T4 bacteriophage, the sliding clamp acts as an enhancer of late gene transcription (Herendeen et al., 1992). In eukaryotes, the so-called 9-1-1 DNA damage check point response involves the homologs of PCNA, known as Rad9, Rad1 and Hus1, which form a heterotrimeric ring (Griffith et al., 2002). Human PCNA also interacts with cell cycle proteins such as $p21^{WAF1/CIP1}$, which inhibits it (Luo et al., 1995).

The sliding clamps serve as communication points between the DNA replication machinery and other cellular mechanisms. How sliding clamps interact with these other proteins was not immediately obvious because the unique shape of these clamps does not immediately suggest a binding site for other proteins. Two crystal structures, one of the RB69 clamp in complex with a C-terminal peptide from its cognate polymerase (Shamoo and Steitz, 1999), the other of human PCNA complexed to a C-terminal peptide from $p21^{WAF1/CIP1}$(Gulbis et al., 1996) revealed a common binding site. These two peptides both bind in an extended fashion onto an interdomain connector on one face of their cognate sliding clamps (Figure 2.3C, 2.4D). The C-termini of the subunits of the ring are on this face. The peptides are stabilized by hydrogen bonding, ionic interactions, and hydrophobic packing. The internal channel of the clamp is totally uninterrupted, leaving the clamp free to slide on DNA. The overlap between the polymerase binding site and that of the replication inhibitor $p21^{WAF1/CIP1}$ suggests that the latter functions by blocking the access of polymerase to the clamp, consistent with biochemical results (Warbrick et al., 1995). This discovery potentially mapped the region on processivity factors that might be shared among all of their interacting partners.

2.1.2 The *E. coli* Sliding Clamp is opened by the δ subunit and loaded onto DNA by the Clamp Loader Complex

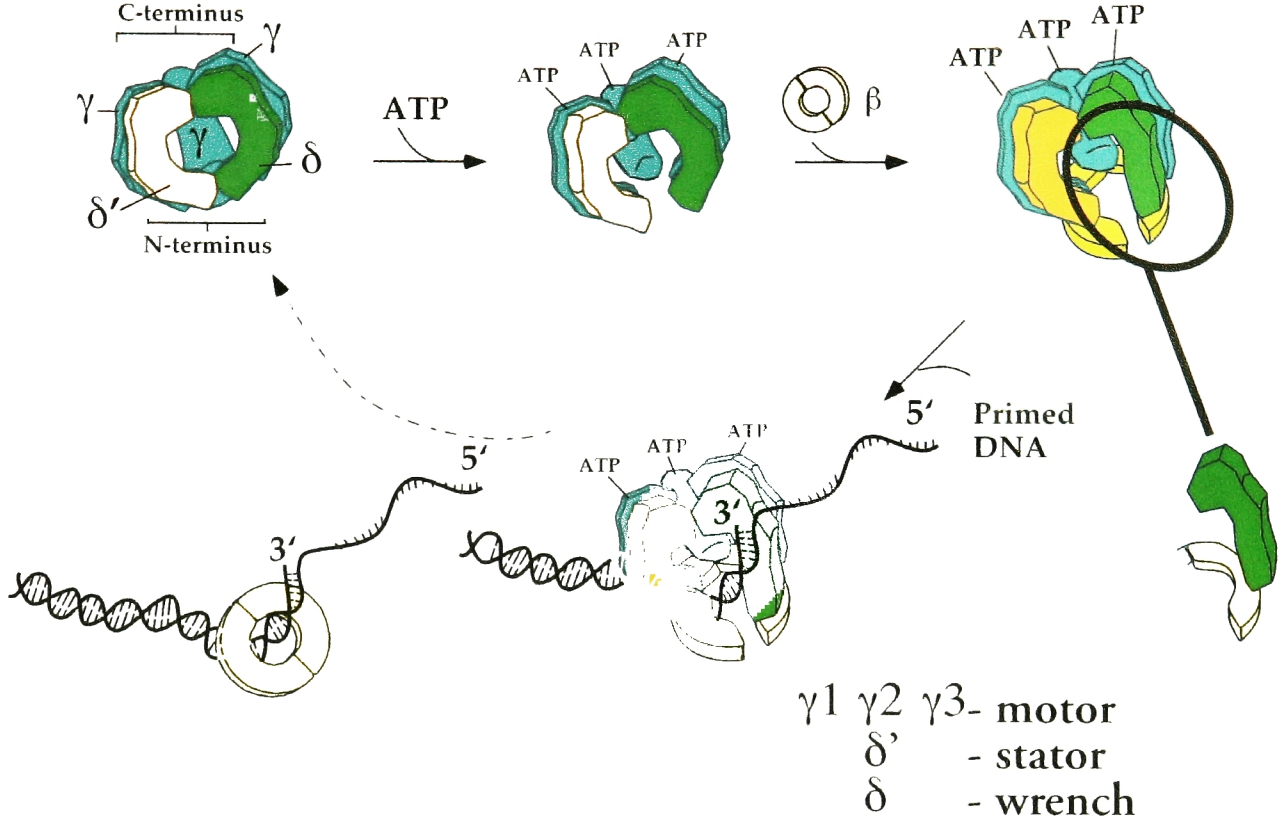
Because sliding clamps do not self-assemble onto DNA other proteins are required to load and unload them appropriately. One interface in the ring needs

to be perturbed momentarily to create a gap large enough to allow DNA to pass through to enter or exit the inner region of the ring. Biochemical studies using the *E. coli* system demonstrated that only one interface is disrupted, while the other interface remains intact, so that the ring does not totally fall apart (Turner et al., 1999). It is also surmised that once the loading is finished, the closed-ring structure will reform itself on DNA.

This task of opening the ring is performed by the clamp loader complex, a sub-assembly of the DNA polymerase complex. Utilizing energy from ATP binding and hydrolysis, the clamp loader cycles through conformational changes that results in the loading of sliding clamps on DNA. The subunits of clamp loader complexes in *E. coli*, bacteriophage and eukaryotes are related evolutionarily (O'Donnell et al., 2001). In bacteriophage the clamp loader consists of 4 copies of gp44 and one copy of gp62. In eubacteria such as *E. coli*, an assembly of γ_3, δ, δ' is sufficient for effective clamp loading. In eukaryotes this complex is referred to as replication factor C (RFC), a complex of five distinct subunits (RFC 1-5) with significant sequence homology to *E. coli* counterparts. Archaebacteria use a clamp loader similar to that of eukaryotes, also named RFC complex because of sequence similarity to eukaryotes.

Using *E. coli* DNA Polymerase III as a model system, extensive biochemical and structural studies of this clamp loading cycle yielded the current picture, summarized by Jeruzalmi et al. in their recent paper (Jeruzalmi et al., 2001a) (Figure 2.2). A minimal complex of γ, δ, δ' in the stoichiometry of $\gamma_3\delta\delta$' is

Figure 2.2 Opening of sliding clamps by the clamp loader complex (Reproduction from Jeruzalmi et al. 2001a) (Permission from David Jeruzalmi)


C-terminus
γ
γ
γ
δ
δ′
N-terminus
ATP
ATP
ATP
ATP
β
ATP
ATP
ATP
5′
Primed
DNA
5′
3′
ATP
ATP
ATP
3′
γ1 γ2 γ3- motor
δ' - stator
δ - wrench

needed to perform ring loading. The γ subunits are the ATPases, δ is the subunit responsible for binding to and opening the β clamp and δ' is thought to act as a stator (Jeruzalmi et al., 2001a). This five-subunit complex may exist as a closed pentameric ring in the nucleotide-free form, with δ sequestered between the neighboring γ and δ' subunits. Upon ATP binding, the γ subunits go through conformational changes that are thought to pull δ away from δ' so as to expose its β-interacting elements. δ will then recruit β and open one interface of it. This intermediary assembly of δ:β is surmised to be further stabilized through interactions of β with other subunits in the complex.

The affinity of DNA for the clamp loader complex facilitates the passage of primer-template DNA to the inside of the opened ring. The addition of DNA stimulates hydrolysis of ATP by the clamp loader, which is thought to cause conformational changes in the γ subunits, pulling δ away from β, resulting in the original closed conformation. Departure of the clamp loader finishes this cycle, with the net result of a closed β clamp on DNA, which is now free to interact with DNA polymerase and initiate DNA synthesis.

Structural analysis of one of the two subunits of archaebacterial RFC (which forms a trimer) shows striking similarities in structure and inter-subunit interfaces to the bacterial clamp loader (Oyama et al., 2001). Sequence homology between the archaebacterial RFC subunit and the eukaryotic proteins is extensive enough to suggest that the RFC complex will be similar in general term to the bacterial complex. However, detailed biochemical analysis of the

clamp loading mechanism of yeast RFC also suggests that there will be significant differences (Gomes and Burgers, 2001; Gomes et al., 2001; Schmidt et al., 2001a; Schmidt et al., 2001b).

Although the process of loading the ring onto DNA appears complicated, the actual opening of the ring requires only δ in an ATP-independent fashion (Turner et al., 1999). However, weak interactions between the clamp and the other subunits of the clamp loader complex probably further stabilize the opened ring and coordinate the other loading steps (Mike O'Donnell, unpublished data). The question of how the δ subunit pries open a stable dimer interface (Kd in the picomolar range) with no additional energy input provokes fascination. The crystal structure of a complex between δ and a monomeric form of the β clamp provided valuable information regarding the mechanism of ring opening. Wild-type β molecules exist predominantly in a dimeric form. The monomeric form of β used in crystallization, referred to as βm1, was engineered by mutating two hydrophobic residues at the inter-subunit interface in β into alanines (I272A, L273A) so as to weaken the dimer interface (Stewart et al., 2001). This modification facilitated complex formation with δ and δ binds to this monomeric β ~ 50 times more strongly than to dimeric β (Stewart et al., 2001), providing a more suitable complex for crystallographic studies.

In the structure of the δ:βm1 complex, one molecule of δ binds to βm1 with two major interactions. A loop region in δ containing two key hydrophobic residues interacts with a hydrophobic pocket on β that is located between domain

2 and 3, which is also composed of conserved residues (Jeruzalmi et al., 2001b) (Figure 2.3A, 2.3B). This binding site overlaps with that seen in the structure of RB69 sliding clamp complexed with a C-terminal peptide from its cognate DNA polymerase (Shamoo and Steitz, 1999) (Figure 2.3C) and the structure of human PCNA in complex with a C-terminal peptide of cell cycle inhibitor protein $p21^{WAF1/CIP1}$ (Gulbis et al., 1996) (Figure 2.3D), suggesting once again that the various sliding clamps have a conserved binding site on their surfaces that may be utilized by many associated factors with distinct functions. The second binding site involves the helix on δ that immediately precedes the loop region that makes contact with the conserved binding site on β. The interaction between this δ helix and a loop in β results in conformational changes in β that weaken the interface and favor a more opened state of the molecule (Jeruzalmi et al., 2001b).

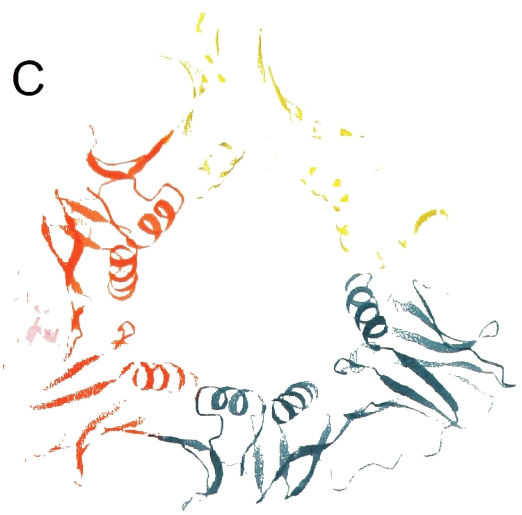
In dimeric β, the semi-circular monomers are joined in a head-to-tail manner so as to form a closed ring. Comparing the curvature of mutated monomeric β (βm1 as in δ:βm1 complex) to that of a wild-type β monomer in the dimer revealed unexpected differences. When the two β monomer structures are superposed on domain 2 of the structure, βm1 has a less-curved crescent shape while the β1 is more bent toward the center of the ring (Figure 2.4) (Jeruzalmi et al., 2001b). Rough calculations indicated that domain 1 is rotated 12° away from the inner side of the ring and the C-terminal section, domain 3, moved 5° outward, resulting in a more relaxed crescent (Jeruzalmi et al., 2001b) (Figure 2.4).

Figure 2.3 Interactions with Sliding Clamp By Multiple Proteins (A) and (B) The structure of *E. coli* mutant monomeric β in complex with δ subunit (Jeruzalmi et al. 2001). (C) The structure of sliding clamp in RB69 in complex with a 11-residue peptide derived from its cognate DNA polymerase (Shamoo et al. 1999). (D) The structure of human PCNA with 22-residue peptide from p21 bound on its surface (Gulbis, et al 1996). The $p21^{WAF1/CIP1}$ peptide is colored magenta. All proteins are shown in Ribbons representation. The β-interaction elements in all three structures are colored magenta. The hydrophobic residues that insert into the surface of sliding clamps are shown in bond-and-stick representation.

A

B

C

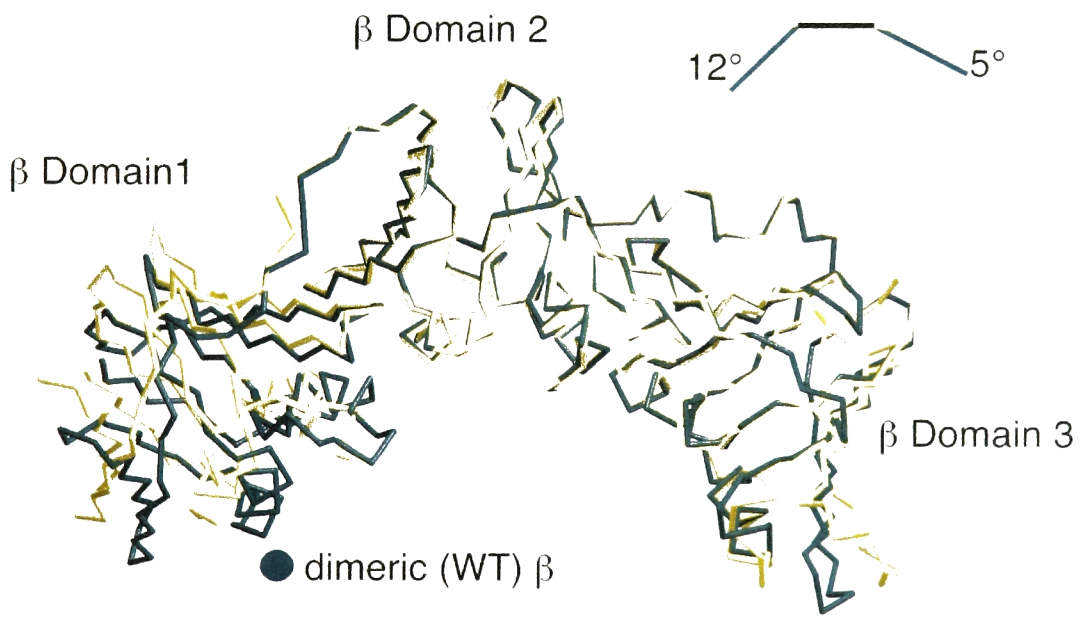


RB69 gp45

D

human PCNA

Figure 2.4 Conformational Change in β upon Binding of δ. The backbone structures of one subunit from the dimer (blue) and the β monomer in the β:δ complex (yellow) are shown, superimposed on Domain 2 (Reproduction from Jeruzalmi et al. 2001) (Permission by David Jeruzalmi).

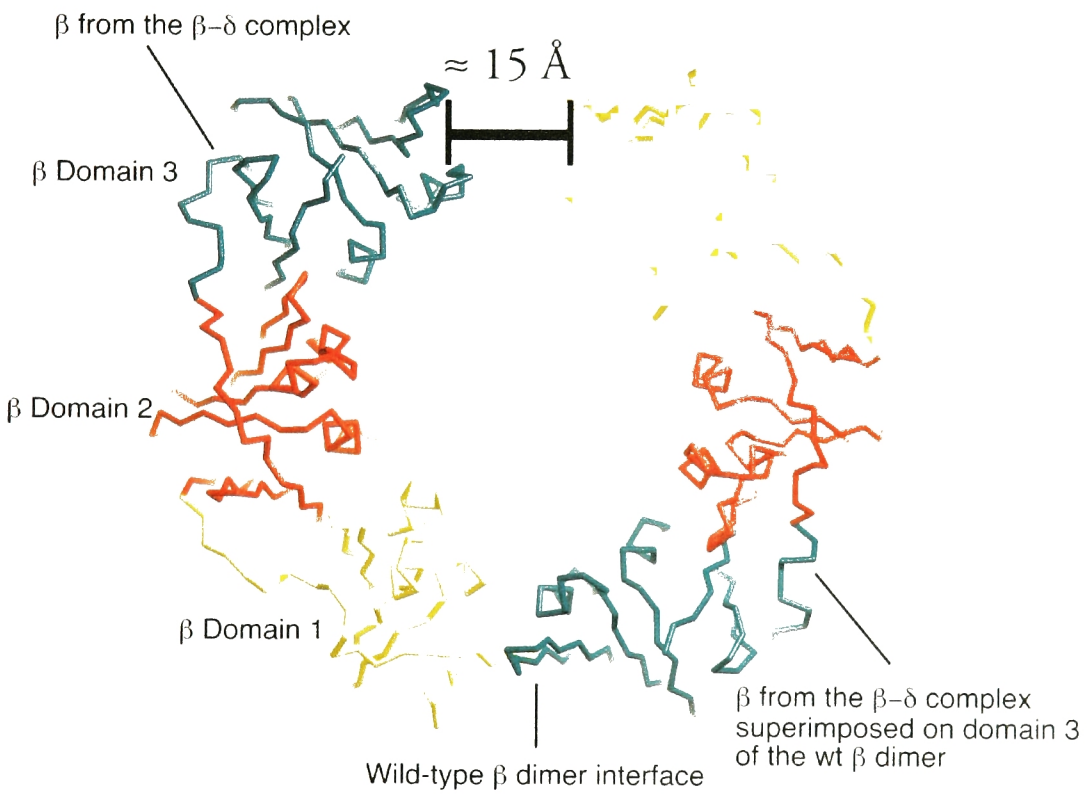


β from the β:δ complex

The significance of this relaxation in bending can be better appreciated if one generates a model of a β dimer with one interface as that in the dimeric form but in which the two monomers adopt the less-bent conformation as seen in βm1:δ complex. This superposition generates a gap at the other interface between the two monomers, with a width of about 15 Å (Figure 2.5), enough for single stranded but not double stranded DNA to pass through (Jeruzalmi et al., 2001b). β clamps are loaded onto DNA at junctions of single stranded/double stranded DNA, and so the passage of single stranded DNA into the ring might suffice for clamp loading. This surprising result, together with the fact that there is a significant rotation between domain 1 and 2 in βm1 structure as compared to wild-type β monomer in the dimer structure, with no direct influence from δ at this site, enticed us to speculate that in a closed dimeric β there is internal strain in each protomer molecule, which is compensated by favorable interactions at the dimer interface. Once δ binds to β and weakens one interface, each monomer could spring into the open conformation, as seen in δ:βm1 complex, so as to create a gap at that interface to allow DNA pass through.

One conceptual roadblock needs to be removed before this hypothesis is fully acceptable. We have to establish that the relaxing of βm1 into an open conformation as seen in the crystal structure is the same as what a protomer of β in the dimer would experience once the interfacial restraint is perturbed in the ring opening process, rather than a response to the mutations we introduced or due to crystal packing forces. To see whether there is internal strain in the

Figure 2.5 A Model for opened β dimeric. The structure of β from the β:δ complex is superimposed twice on one intermolecular interface in the crystal structure of dimeric β (Kong et al., 1992), once using Domain 1 as a reference, and once using Domain 3. A hybrid molecule was then created, which has one open interface and one closed one. Domains 1 and 3 at the closed interface were retained from the crystal structure of dimeric β, whereas the rest of the domains are from the structure of β in the β:δ complex (Reproduction from Jeruzalmi et al. 2001) (Permission from David Jeruzalmi).


β from the β–δ complex
≈ 15 Å
β Domain 3
β Domain 2
β Domain 1
β from the β–δ complex superimposed on domain 3 of the wt β dimer
Wild-type β dimer interface

dimeric β which would drive the ring to open upon interface disruption we turned to molecular dynamics (MD) simulations.

2.1.3 Molecular Dynamics Simulations

Solving the three-dimensional structures of proteins and their complexes with other biomolecules using either X-ray crystallography or Nuclear Magnetic Resonance (NMR) is a powerful route for biologists to gain detailed atomic resolution information on functional mechanisms. However, further studies are usually necessary to fully answer the questions being addressed. This is because proteins are dynamic macromolecular systems, and a static picture of the three-dimensional spatial arrangement of the atoms is usually not sufficient to understand fully how they function. This is especially true in the case of the mechanism by which β is opened by δ, for which we have structures for two states of the same protein (β) but miss the information about the path linking one to the other and the nature of the driving forces. Little information is available experimentally about the transient conformations of the clamp in the course of opening. In contrast to direct experimental procedures, molecular dynamics (MD) simulations can, in principle, provide a complete theoretical description of the structure and motions of the biomolecules in the presence of solvent molecules and counter ions. With the sharp increase in computational power MD is becoming a valuable means of developing models and interpreting experimental data on molecular structures.

In molecular dynamics simulations, each atom in the system studied is associated at time point *t* by its spatial position (x,y,z) and velocity (Vx,Vy,Vz). Its location and velocity at the next time point $t+\Delta t$ is inferred from all the acting forces on the structure according to Newton's law (for reviews of molecular dynamics simulations, see Brooks, (1995); Karplus and Petsko, (1991)). This calculation generates a trajectory that specifies how the atoms in the systems move as time progresses. The time scale of such simulations for proteins is limited to the range of nanoseconds due to the intensive calculations involved.

The description of all the forces in a structure is implemented as a force field in MD simulations. This force field is an empirical parameterization that takes into account the bonded and non-bonded interactions in the system and is optimized through iterative modification to match computational results with experimental data on various systems, such as water and other small molecules (Cornell et al., 1995). When conducted properly, MD simulations can provide information about both the structural and dynamic properties of a system in detail which is often inaccessible to direct experimental observations.

Although MD is a well-defined theoretical methodology, it is not unlimited in terms of accurately or completely describing and predicting the behavior of a system. The biggest and almost insurmountable problem right now is the limited sampling of conformational space. Biological systems can be quite large, with independent variables in the number of tens of thousands or even more. Usually one stable state of the system corresponds to an ensemble of different

conformations of enormous number. Due to limited computational power in such calculations, it is impossible to sample all the possible conformations the system, which in some cases seriously damages the credibility of these calculations. Secondly, approximations inherent in the force field bring associated error, whose accumulation may sometimes divert the system to a physically and chemically irrelevant state. Thus, care must be taken in analyzing a simulation and a tight connection with experiments is essential for the validity of the simulation. If the experimental and the simulated results agree, there is a greater probability that the results of the simulation can be used for further prediction and also give detailed explanations of experimentally observed phenomena.

Molecular dynamics simulations made their entry into the field of proteins in 1977, with the simulation of the protein bovine pancreatic trypsin inhibitor (BPTI) *in vacuo* (McCammon et al., 1977). With improved computational performance, the inclusion of explicit solvent became feasible and the first MD simulation of a hydrated protein was performed in 1984, also on BPTI (van Gunsteren and Berendsen, 1984). Recent advances in computational power and simulation methodology have enabled the use of MD methods to simulate large biological systems, including solvent water molecules and salt ions to represent physiological solvent conditions for trajectories extended well into the nanosecond region of time (reviewed by Brooks et al. 1995). New versions of force fields have been implemented to represent the interaction terms more accurately, especially for cases like nucleic acids (Cornell et al., 1995). In addition, a treatment of long-range electrostatic interactions, the particle mesh

Ewald (PME) method (Darden et al., 1995), has been implemented to deal with the long-standing issue of the treatment of elestrostatics without truncation. Molecular dynamics has been used recently to study macromolecular assemblies in detail. Systems that MD simulations have been applied to include DNA, RNA, protein, DNA/protein complex, RNA/protein complex, protein-ligand binding etc, with good degree of success (reviewed by (Kollman et al., 2000)). All these efforts have provided an independent theoretical means of investigating further the problems identified above, which also served as tests for the validity of this method.

The goal of our study in utilizing molecular dynamics was to begin to explore whether there is an intrinsic tendency in monomeric β to spring from a closed to open conformation once the restraint from the dimer interface is removed. Our approach was to simulate one β monomer starting from its closed conformation as seen in the original dimeric β structure (Kong et al., 1992), but without the presence of the other molecule. By doing this we hoped to see whether this "closed" β monomer relaxes into a conformation similar to that seen in the δ:βm1 structure. A simulation of the dimeric β clamp was also conducted, to serve as a control calculation. Once the MD trajectories were generated we analyzed the energetics of the molecules, which was used to examine the nature of the internal strain, if any, in the dimeric form of the ring. Our aim with these studies was to complement the information that was available from experiments and, ultimately, to refine our understanding of the clamp opening.

2.1.4 Free Energy Calculations from MD Simulations

It is enticing, yet challenging, to compute the free energy (G) of a biological system at certain state, or the difference in free energy (ΔG) between two states. Unfortunately, the free energy is a difficult quantity to obtain from molecular dynamics simulation of large macromolecule systems using explicit solvent models. There are two major sources of error associated with such a calculation. Errors may rise from inaccuracies in the force field chosen to model the system. With improvements in parameters and utilization of the Particle Mesh Ewald method (Darden et al., 1995) to avoid truncation in the calculation of electrostatic terms, such errors may be reduced to reasonable levels. The second source of error arises from an insufficient sampling of phase space due to limited computational resources. This problem is particularly severe for large-sized systems, which take a long time to equilibrate. One strategy commonly utilized by researchers is to use a free energy perturbation method, but this method works only for very limited conformational or topological changes (Kollman, 1993).

Despite the apparent difficulties in estimating G from MD simulations, Kollman and coworkers have proposed an empirical approximation method that produces reasonable results (Kollman et al., 2000). In this approach, the free energy of a system is partitioned into several components, each of which can be computed readily and the free energy is treated as a sum of all these components. The protocol for this method is as follows. A trajectory from a

simulation is post-processed to remove all the solvent water molecules and ions, which are later represented by a featureless continuum with constant dielectric to simplify calculations. For each snap-shot conformation of the system the free energy, G, is calculated according to the following equations:

$$G = E_{MM} + G_{PBSA} - T\Delta S_{MM} \quad (2.1)$$

$$E_{MM} = E_{bond} + E_{angle} + E_{dihedral} + E_{vdw} + E_{elec} \quad (2.2)$$

$$G_{PBSA} = G_{PB} + G_{SA} \quad (2.3)$$

$$G_{SA} = \gamma^{*}SASA + \beta \quad (2.4)$$

where E_{MM} is the molecular mechanical energy as a sum of the bonded (bond, angle, dihedral) and nonbonded (van der Waals, and electrostatic) energy terms in the molecular mechanical force field, evaluated with no nonbonded cutoff. G_{PBSA} is the solvation free energy, which is a sum of the G_{PB} term (calculated by a numerical solution of the Poisson-Bolzmann equation) and the G_{SA} term (an estimate of the nonpolar free energy contributed by the solvent accessible surface area (SASA)). γ and β are two constants with γ = 0.02268 kJ/mol$Å^2$ and β = 3.85 kJ/mol (Wesson and Eisenberg, 1992).

$-T\Delta S_{MM}$ is the solute entropy, which can be estimated by quasi harmonic analysis of the trajectory (Srinivasan et al., 1998) or inferred from a static structure through normal-mode analysis (Hinsen, 1998) or some other method. Except in cases where large conformational changes are involved, the change in

the internal entropy term is expected to be small. Normal mode analysis for a large system can only give qualitative estimations because of the intrinsic limitation in the method. The quasi-harmonic method infers the entropy of a system from its RMS displacement over a time period, thus simulations over long time scale (exceeding nanoseconds) are required to reduce computational error. Because of the imperfect tools available for estimating the entropy of a system, it is often a practice to omit this term from calculations except for systems undergoing substantial conformational changes. For the system we study qualitatively we can safely state that the entropy of the β molecule in a monomeric form is higher than in the dimeric form because interfacial restraints reduced the degrees of freedom in the dimer. We do not, however, estimate the magnitude of this effect.

The value of G calculated using this method for instantaneous structures is usually averaged over a reasonable time period to include inherent thermal fluctuations in MD simulations. For sake of brevity this method will be referred to Molecular Mechanics – Poison Boltzman Surface Area (MM-PBSA) method, as first denoted by Kollman (Kollman et al., 2000) and will be utilized in our energetics analysis.

This decomposition and approximation procedure enables us to calculate a crude average of the absolute value of G over a certain time period for a system. It is obvious that this method has intrinsically much larger errors than free energy perturbation/thermodynamic intergration calculations. But one

advantage we gain from the MM-PBSA method is that we can calculate the absolute free energy difference (ΔG) between two end states directly instead of calculating the relative free energy along a reaction coordinate, as required by the other two methods. We thus skip computations on intermediate states to save computer time and avoid calculation errors in those steps. The MM-PBSA method has been applied to several situations and often ΔG can be calculated in respectable agreement with experiments (Kollman et al., 2000). For example, the MM-PBSA method was used to calculate the binding free energy for the oncoprotein MDM2 and the tumor suppressor p53. Computational alanine scanning mutagenesis of 12 residues in the N-terminal helical portion of p53, important for binding activity, was performed by either generating a new trajectory for each mutation or modifying the side chain of the substituted residue in the wild type trajectory (Bottger et al., 1997). $\Delta\Delta G_{bind}$, the difference in binding free energy between wild type and mutants, was obtained as $[(\Delta G_{bind})_{mut} (\Delta G_{bind})_{wt}]$. It was found that the four residues that have the largest $\Delta\Delta G_{bind}$ were the same residues most sensitive to substitution experimentally. As a second evaluation of the validity of this method, three charged residues, which were found to cause no significant loss of binding when mutated to alanine in experiments, did have small $\Delta\Delta G_{bind}$, which has an effective cancellation between $\Delta\Delta E_{elec}$ and $\Delta\Delta G_{PB}$ in the magnitude of ~50 kcal/mol (Bottger et al., 1997).

As will be reported later in the Results and Discussion section, we did observe spontaneous opening of monomeric β from its starting close conformation, as seen in the dimeric structure, to the more opened conformation

in δ:βm1 complex. In contrast, the dimeric β remained stable and closed during an MD simulation. This natural tendency to adopt a more open conformation for monomer β once interfacial restraints are removed is very intriguing in the sense that there must be some internal stress stored in the closed structure for it to take place. The MM-PBSA method was applied to our simulation studies to investigate if there was any kind of free energy change coupled with what we observed structurally in our simulations. The changes in decomposed free energy terms were then compared to check if the structural change we saw in monomer β simulation is a direct result of one or two energy terms relaxing or a distributive effect from all the factors. By doing this we hoped to discover the spring-loaded mechanism that governs the ring opening once one of its interfaces are weakened by the clamp loader complex.

2.2. Methods and Materials

We carried out multiple MD simulations to confirm the conformational stability of the β dimer and to study the intrinsic ability of the β monomer to open up in the absence of dimer interfacial restraints. For brevity, the individual simulation runs will be referred to by their assigned codes (Table 2.1). In all simulations, Na^+ (or K^+) and Cl^- counter-ions were added to balance the charges on the protein so as to create a neutral environment. Explicit water molecules and additional ions were added so as to represent aqueous solution with a salt concentration of 150 mM NaCl (or KCl). In MD run β2, a trajectory was obtained for the *E. coli* β dimer, starting from the crystal structure (Protein Data Bank 2POL, (Kong et al., 1992)). Simulations β1A and β1C are of isolated monomers taken from the structures of *E. coli* and *S. Pyogenes* β dimers, set up by using the atomic coordinates of only one molecule from their original dimer structures ((Kong et al., 1992) and Maria Argiriadi, unpublished structure). Simulation β1B is essentially the same as simulation β1A except that it was generated using a different software package (CHARMM instead of AMBER). The MD runs β2, β1A, β1C were all done using AMBER, with the setup, equilibration and production procedures being the same. The β1B run was calculated using CHARMM, and followed essentially the same route as the above three, with minor implementation differences that were dictated by CHARMM specifics.

2.2.1 Construction of the Microscopic Model

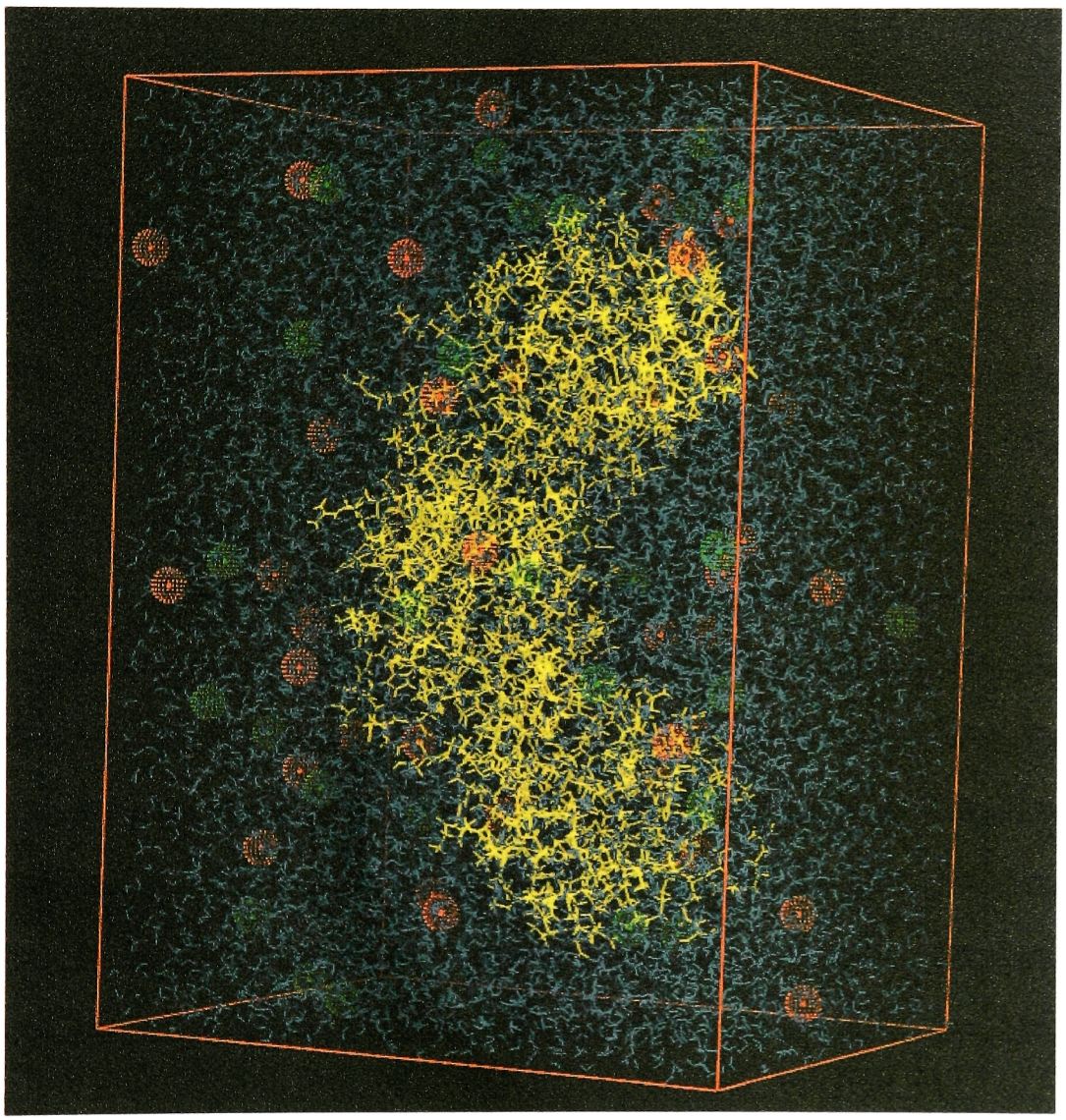
Table 2.1 System Setup for Multiple MD Simulations

Table 2.1 System Setup for Multiple MD Simulations

Trajectory	Protein	Software	Box Size (Å) (X*Y*Z)	Number of Protein Atoms	Number of Ions(+/-)	Number of Water Molecules	Total Number of Atoms
β2	*E. coli* β dimer	AMBER	107.4*64.1*113.1	11,410	67/45	16,486	60,977
β1A	*E. coli* β monomer	AMBER	75.5*63.5*104.7	5,705	41/30	11,046	38,914
β1B	*E. coli* β monomer	CHARMM	62*70*120	5,705	49/38	14,966	40,407
β1C	*S. pyogenes* β monomer	AMBER	77.6*61.5*107.8	5,684	42/32	11,518	50,690

The procedure for the AMBER simulations was as follows. The crystal structures of *E. coli* and *S. pyogenes* β were prepared for dynamics using the Leap module of AMBER (Pearlman et al., 1995). This involved first adding protein hydrogen atoms to the structure, aligning the principle axes of the protein with the Cartesian axes of the simulation box, and neutralizing the net charge of the protein by adding Na^+ ions. The PARM98 version of the AMBER force field was used (Cornell et al., 1995). All ionizable side-chains were configured in their characteristic ionized states at pH 7.0. All crystallographically determined water positions were retained during the simulation setup. The simulation unit cell was a rectilinear box, which extended 6.0 Å beyond the protein in each dimension. The protein, crystallographic waters, and ions in the simulation box were then surrounded with water molecules by overlaying a periodic box of waters on top of the crystallographic coordinates and removing any overlapping water molecules (Figure 2.6). The periodic box of waters (TIP3P water model) (Jorgensen, 1981) had been pre-equilibrated at 298 K via Monte Carlo simulation (Matthew Young, unpublished). Finally, Na^+/Cl^- ion pairs were added at random positions (replacing water molecules in the event of an overlap), such that the molar salt concentration was ~150 mM. The number of ion pairs needed was determined by a crude estimation of the solvent volume as the difference between the volume of the simulation unit cell and the volume of the protein. The protein volume is approximated as the product of $(X_{min}+ X_{max})^*(Y_{min} + Y_{max})^*(Z_{min} + Z_{max})$, because the protein lies within the spatial boundary defined by $(-X_{min}, X_{max})$, $(-Y_{min}, Y_{max})$, and $(-Z_{min}, Z_{max})$. For the details of the final simulation system for each trajectory,

Figure 2.6 Setup system for β1 MD simulation. Molecular representation of the atomic model of the monomeric β molecule (shown in yellow) bathed in an aqueous salt solution of 150 mM NaCl. Explicit water molecules are shown in blue, Na^+ ions are shown as red spheres and Cl^- ions are shown as green spheres.



including the size of the simulation unit cell and composition of protein, ions and water molecules in it, please see Table 2.1.

For the monomeric β run using CHARMM (MD run β1B), the setup procedure was essentially the same except for using CHARMM specifics, which included hydrogen rebuilding, axes realigning, charge neutralization, ion pair addition to satisfy salt concentration and water solvation in the end. The simulation unit cell was constructed to extend 10.0 Å beyond the protein, larger than the 6.0 Å extension used for the AMBER trajectory. One notable difference is that K^+ ions were used instead of Na^+ ions, due to their improved potential function parameters in the CHARMM suite (Benoit Roux, personal communication). The number of cations and anions needed to balance the charge in the system and maintain ~150mM salt concentration was determined by a CHARMM subroutine by solving the Poisson-Boltzmann equation (Im et al., 1998), a more accurate procedure than that implemented in AMBER. The final size of the system is slightly different from the AMBER trajectory (Table 2.1). The calculations were performed using the academic version c27a1 of CHARMM (Brooks et al., 1983). The all-atom potential energy function PARAM-22 for proteins (MacKerell et al., 1998) was used. The TIP3 potential was used for the water molecules (Jorgensen, 1981). The Lennard-Jones (LJ) parameters for K^+ and Cl^- were adjusted to yield the experimental solvation free energy in bulk water (Roux, 1996). Periodic rectangular boundary conditions were applied in all directions.

2.2.2 Equilibration and Dynamics Procedures

For the β2, β1A, and β1C runs using AMBER, the conformation of the solvated protein system was relaxed via a series of energy minimizations before initiating dynamics, using the SANDER module of AMBER. Periodic imaging of the central cell was included to model an infinite solvated protein environment and to eliminate solvent/vacuum boundaries, and the particle mesh Ewald summation technique (Darden et al., 1995) was employed to calculate electrostatic interactions between atoms at distances greater that 9 Å, effectively eliminating the truncation of long-range electrostatic interactions. Following minimization, the system was gradually heated to 298 K by slowly increasing atomic velocities over a period of 10 ps. Harmonic positional restraints were imposed on the protein atoms for the first 50 ps of the simulation to allow the solvent to equilibrate around the protein. The force constants of the positional restraints were gradually reduced in magnitude over this 50 ps time segment (from 25 kcal/mol/Å to 2.5 kcal/mol/Å). The remainder of the dynamics was carried out in the absence of any restraints, under constant temperature and constant pressure conditions, and using a 2 fs time step. Production dynamics generated roughly 100 ps of trajectory per day using six processors in parallel on a 195 MHz R10000 SGI computer.

In the β1B CHARMM run, a similar procedure was implemented using CHARMM specifics. The electrostatic interactions were computed with no truncation, using the particle mesh Ewald algorithm with a B-spline order of 4 and an FFT grid of one point per Å (Essmann et al., 1995). The list of nonbonded

interactions was truncated at 13 Å, using an atom-based cutoff. The nonbonded van der Waals terms were smoothly switched off at 8-10 Å. The trajectory was generated under constant pressure and temperature conditions with extra degrees of freedom for a Langevin thermostat and piston (Feller et al., 1995). The SHAKE algorithm (Ryckaert et al., 1977) was used to fix the length of all bonds involving hydrogen atoms, and the equations of motion were integrated with a time step of 2 fs. The coordinates were saved every 0.5 ps. The model structure was first minimized with 50 steps of the steepest-descent (SD) algorithm, followed by 50 steps using the adopted basis Newton-Raphson (ABNR) method. The system was first coupled to a heat bath at 300 K and maintained at constant volume with Langevin dynamics for 500 steps. A friction coefficient of 5.0 ps^{-1} was used for the nonhydrogen atoms. The equilibration of solvent molecules lasted 20 ps at constant temperature and pressure, with decreasing harmonic restraints on the protein atoms (from 25 kcal/mol/Å to 5 kcal/mol/Å). This was followed by a 3.4 ns unconstrained simulation at 300 K at constant temperature and pressure. A time-step of 2 fs was used and coordinates were saved every 500 steps (1 ps). A notable disadvantage of CHARMM as compared to AMBER is its slower computation speed. It takes about 35-45% longer time than AMBER to perform the same length of simulations.

2.2.3 Free Energy Calculations

Although trajectories β2, β1A and β1C were generated using AMBER, all free energy calculations were done using the CHARMM suite. The AMBER force

field was used in CHARMM so as to ensure unity between the AMBER MD simulations and free energy calculations within CHARMM. All trajectories were first post-processed to remove all water molecules and ions and realigned so that each instantaneous structure of the molecule had the same center of mass (set to the origin). Using the MM-PBSA method as described in the introduction, the free energy of each instantaneous structure was calculated as a total of several energy terms, including E_{bond}, E_{angle}, $E_{dihedral}$, E_{vdw}, E_{elec}, G_{PB} and G_{SA}. E_{bond}, E_{angle}, $E_{dihedral}$, E_{vdw}, E_{elec} were obtained directly by issuing the "ENERGY" command in CHARMM with no non-bonded cutoff, no periodic boundary condition and the dielectric constant $\epsilon = 1$, effectively calculating the molecular mechanics energy of one molecule in vacuum. The solvent accessible surface area (SASA) was obtained using the "SURFACE" function in CHARMM with a default probe radius of 1.6 Å. G_{SA} is then deduced from SASA according to eq. (4) (Introduction) but the constant term β is not added in. The solvation free energy (G_{PB}) was estimated as the sum of electrostatic solvation, calculated by the finite-difference solution to the Poisson–Boltzmann equation (PBEQ) as implemented in the CHARMM program (Im et al., 1998) using $\varepsilon = 1$ for the solute and $\varepsilon = 80$ for the solvent. A probe radius of 1.6 Å and the atomic radii parameter set from Nina et al. (Nina et al., 1997) were used to determine the molecular surface. The boundary of the grid for the PBEQ module was determined by assuming that the protein occupies 70% of the volume of the cubic lattice. A grid resolution of 0.5 Å/grid point were used. In our calculations the free energy contribution from

entropy changes in the trajectories was not included as explained before in the Introduction section.

In order to plot the free energy as a function of time for each trajectory, the free energy of each instantaneous structure was calculated every 10 ps for the entire course of the simulation. To calculate the average free energy for the initial and final stages of each trajectory, the free energy was averaged over a 250 ps window by computing it every 1 ps within such windows at the beginning and the end of the simulation. Exact ranges of this window for block averaging in different trajectories are shown in Table 2.2.

Free energy calculations on the two "static" *E. coli* X-ray crystal structures, the closed dimer β structure (Kong et al., 1992) and the open monomeric β structure (Jeruzalmi et al., 2001b), were done using the same MM-PBSA method as used for the MD trajectories. Both protein structures were prepared for energy calculations by generating the hydrogen atoms using CHARMM, and minimizing the energy of both structures using the steepest descend (SD) algorithm in CHARMM, with the PBEQ module turned on to mimic a solvated environment. The PBEQ term was updated every 10 steps during minimization. 300 steps of SD were found to be sufficient for the minimization to converge in both cases. Free energy calculations were then performed on the two minimized structures, but only using one molecule of β from the closed dimer structure to match the opened monomer β structure.

2.3 Results and Discussion

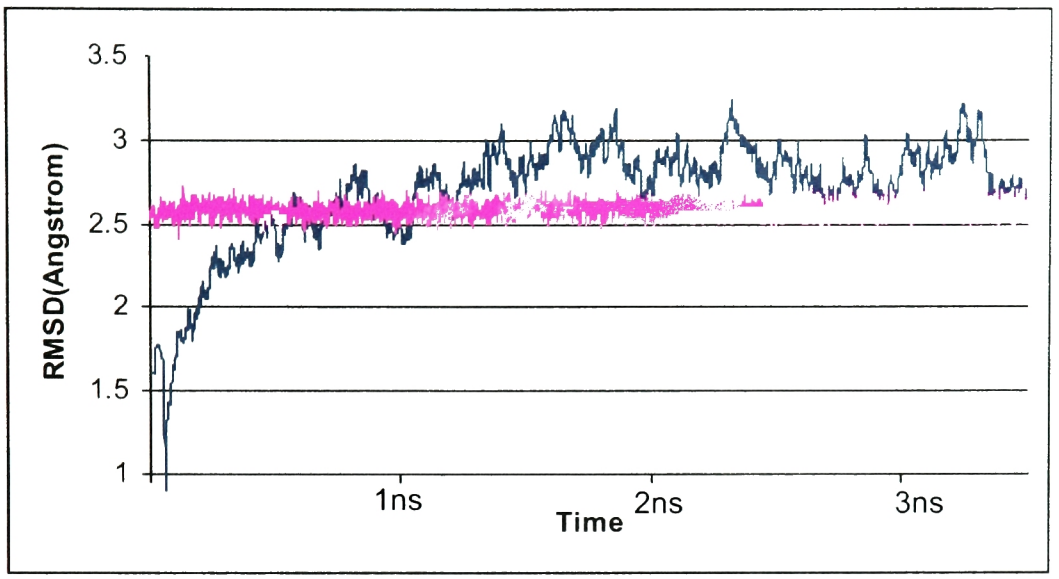
2.3.1 Dimeric Structure is Stable in Simulation

To testify the stability of the simulation methodology, we carried out a 3.4 ns simulation of the *E. coli* β homodimer. The crystal structure of this clamp was solved at 2.6 Å resolution (Kong et al., 1992) and used as the starting conformation. Dimeric β is stable relative to the crystal structure over the course of the simulation, with both interfaces remaining intact and without a net change in the overall shape of the ring. When all the Cα atoms in the ring are superimposed, the RMS deviation in Cα positions with respect to the crystal structure oscillates around ~2.7 Å during the trajectory. The distance between the closest pair of atoms at the interface, one on each monomer, oscillates around ~ 2.5 Å too, indicating no spatial separation of the two monomers at the dimer interface (Figure 2.7A).

When individual structures from the simulation are superimposed on one interface of the ring (domain 1 in one molecule and domain 3 in the other) the RMS deviation of one component of the other interface region (domain 3 of the first molecule) rarely deviated beyond 6 Å from the X-ray structure, with an average of about 5 Å. When compared to the domain 3 of the monomeric βm1 structure as seen in βm1:δ complex with domain 1 superimposed, the RMS deviation is always larger than 6 Å and mostly oscillates around 10 Å (Figure 2.7B). More importantly there is no change in the trend of the two RMS deviation plots throughout the simulation, indicating that the structure stays at the dimeric closed structure and never relaxes to the open conformation as seen in the

Figure 2.7 RMS Deviation in Cα atoms for β2 Run. (A) (blue) RMS deviation in Cα atoms for entire β dimer as compared to original crystal structure. (magenta) the distance of closed pair of atoms at the dimer interface (d_{min}), one from each monomer. (B) (blue) RMS deviation in Cα atoms for domain3 of molecule 1 in β dimer as compared to original crystal structure. (magenta) RMS deviation in Cα atoms for domain3 of molecule 1 in β dimer as compared to the crystal structure of βm1 in βm1:δ complex crystal structure. Instantaneous structures from the trajectory are superimposed on Domain 1 of the reference structure, which is the crystal structure of dimeric β. RMS deviations in A are calculated and plotted for the entire trajectory (3.5 ns) but only up to 2.3 ns in B.

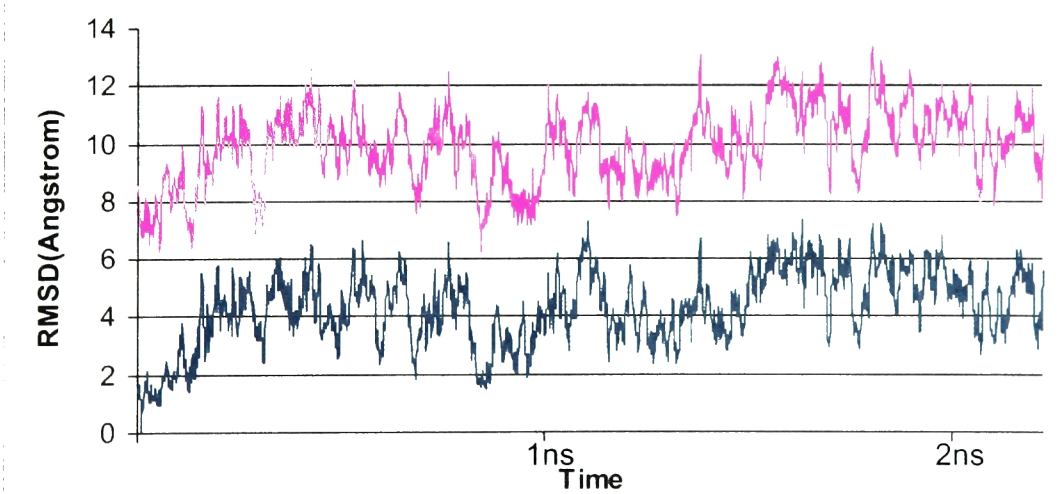
A



overall structure vs. beta from wild type beta dimer

d_{min} at dimer interface

B



domain 3 vs. beta from wild type beta dimer

domain 3 vs. beta from beta-delta complex

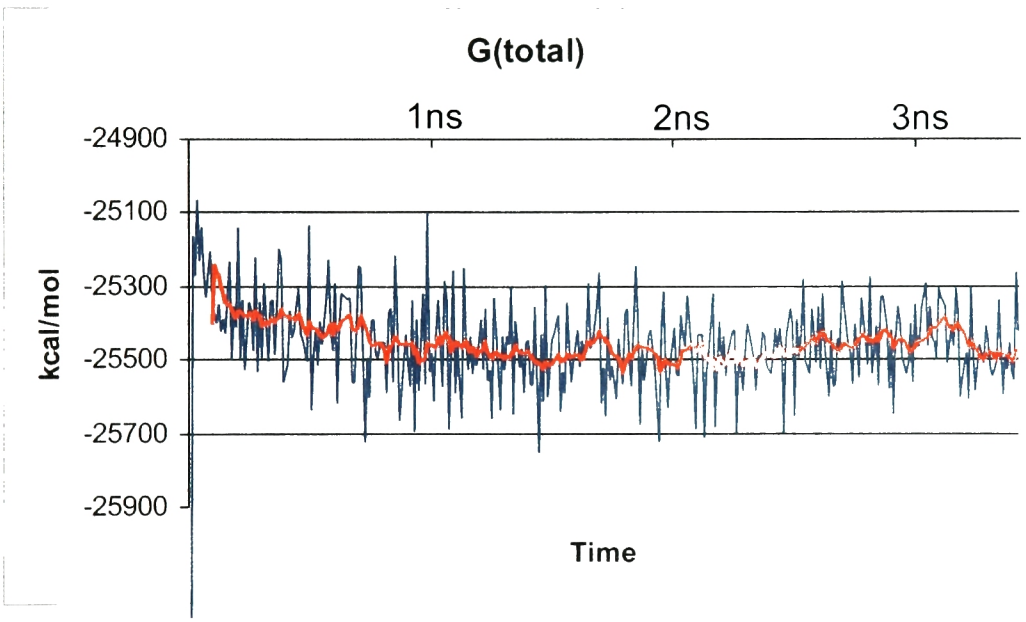
βm1:δ complex (Figure 2.7B). This is consistent with experimental data indicating a very stable dimer in solution (Yao et al., 1996).

We performed MM-PBSA free energy calculations on post-processed protein-only coordinates from this simulation, and the time dependence of the free energy is plotted in Figure 2.8. Thermal fluctuations in the values of G of ~0.5% are present throughout the whole trajectory, which is characteristic of MD simulations. An initial steep drop in the value of G in the first 50 ps or so reflects the quick relaxation of the system from the X-ray structure. This decrease is then followed by a general downward trend until it becomes stabilized after about 1.0 ns. The second stage of steady decrease can be regarded as a result of the molecule relaxing further into a different parameter space from that used in the crystal structure refinement. In the final stage, the free energy oscillates around an average value, indicating that the system has reached equilibrium. This phenomenon of decrease in free energy is common in all simulations, and should not be taken as a meaningful indication of the molecule going from a higher to a lower free energy conformation.

2.3.2 MD Simulations of Monomeric *E. coli* β Suggest a Spring-Loaded Component to the Ring Opening Mechanism

We performed two simulations (β1A, β1B) of the nanosecond timescale dynamics of the *E. coli* monomeric β subunit in a solvated environment using standard procedures. For these monomeric β runs, we started from the same

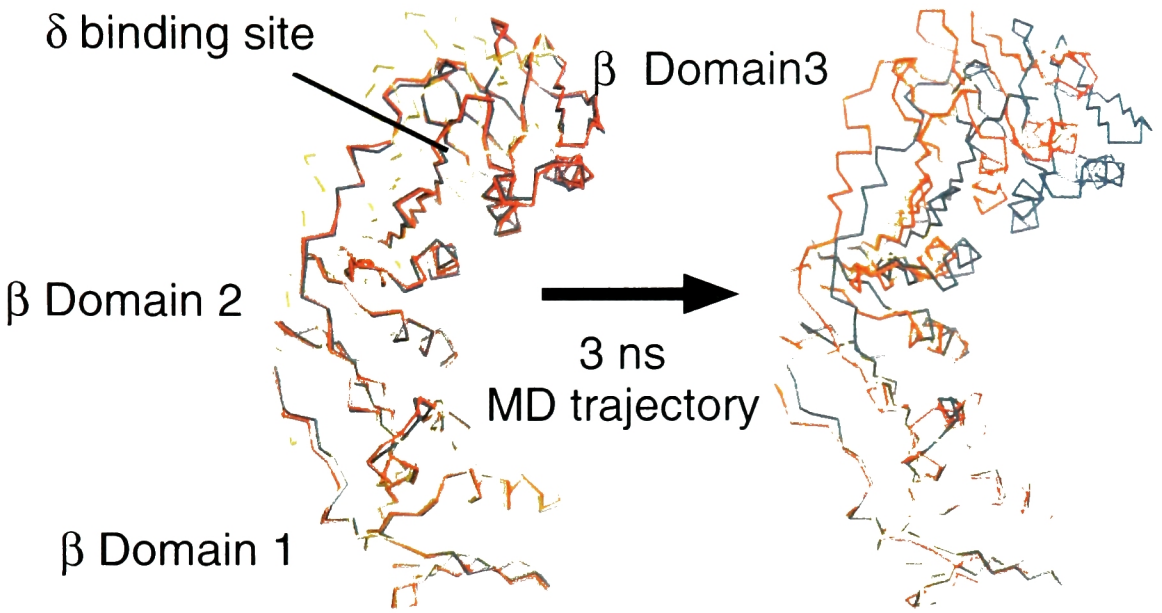
Figure 2.8 Free Energy vs. time in β2 MD Run (Trajectory β2) Free Energy values obtained from MM-PBSA method are plotted vs. time for β dimer simulation. Simulation is sampled at a frequency of 10 ps. The red trend line is the moving average over 100 ps.


G(total)
1ns
2ns
3ns
-24900
-25100
-25300
-25500
-25700
-25900
kcal/mol
Time

crystal structure of wild-type dimeric β as before, but we removed one of the two molecules at the start of the trajectory. This allowed us to follow the relaxation of the structure upon removal of the dimer restraints. No artificial driving forces were applied in these trajectories. Both simulations of monomeric β were extended to the nanosecond (ns) range, with the β1A run calculated beyond 3.0 ns and β1B to 2.0 ns.

What we obtained from these MD simulations was much more satisfying than we first imagined. In contrast to the stable conformation seen in the β dimer (trajectory β2), the simulation of monomeric β produces a remarkable result. Starting from a structure that is identical to that of one monomer in dimeric β, the monomer relaxes within ~1.5 ns in trajectory β1A to a structure that resembles the more open form of β seen crystallographically in the βm1:δ complex (Figure 2.9). This is illustrated most clearly by superimposing domain 1 from the instantaneous structures in the trajectory onto domain 1 of the dimeric crystal structure, and then monitoring the RMS deviations of Cα atoms in domain 3 with respect to either domain 3 in the crystal structure of dimeric β (blue trace in Figure 2.9) or βm1 in the crystal structure of the βm1:δ complex (red trace in Figure 2.9). Starting from an RMS deviation of 0 Å with respect to domain 3 of dimeric β, the structure moves away rapidly, and eventually oscillates around an RMS deviation of ~10 Å away from domain 3 in the initial structure. In contrast, domain 3 in the trajectory is initially ~10 Å away from domain 3 in the βm1:δ complex, but relaxes toward it, eventually oscillating around an RMS deviation of ~3 Å from the structure in the βm1:δ complex. The close overlap in the position

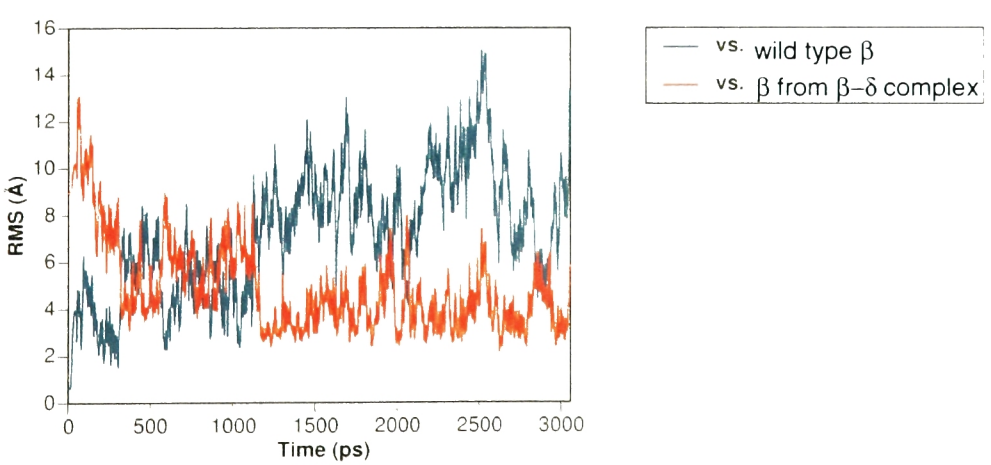
Figure 2.9 Opening of β1 observed in β1 simulation (Trajectory β1A). (top) The backbone structure of one subunit from the dimeric form of β is shown in blue. This structure was used to initiate the molecular dynamics trajectory. The crystal structure of the β monomer in the structure of the βm1:δ complex is shown in yellow. An instantaneous structure from the start of the trajectory (left) and from the end (right) are shown in red. The structures are superimposed on Cα atoms of Domain 1. (bottom) RMS deviation on Cα atoms in domain 3 of β are plotted as a function of time. Instantaneous structures from the trajectory are superimposed on domain 1 of the reference structure, which is the crystal structure of dimeric β for the blue trace, and the crystal structure of βm1 in the βm1:δ complex in the red trace (Reproduction from Jeruzalmi et al. 2001) (Permission from David Jeruzalmi).



dimeric (WT) β crystal structure

β from β–δ complex crystal structure

instantaneous β from MD trajectory



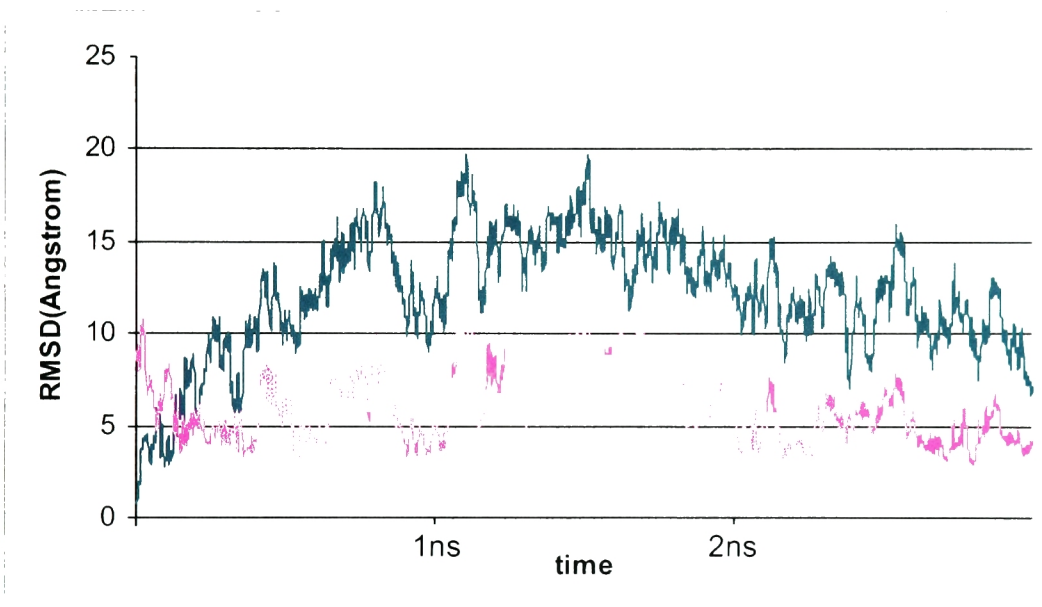
of domain 3 in the trajectory with that of the βm1:δ structure can be readily visualized in Figure 2.9. It is also noted from the RMS deviation plot that once the β monomer opened, it stayed open in the absence of the other monomer, never falling back to its initial closed conformation (Figure 2.9).

Large scale conformational changes in proteins are not usually observed on a nanosecond timescale. The rapid structural relaxation in β monomer indicates that the monomer readily adopts a stable and more open structure when not constrained by dimer interactions, without having to surmount large energy barriers. The convergence to the open conformation of the β monomer as seen in the crystal structure of the βm1:δ complex indicates that the this conformation is a preferred state of lower free energy that the β monomer quickly occupies. The forces underlying this "spring" are difficult to disentangle, especially with a system of this large size. There is no obvious steric effect that caused the domains of β to open up when the dimer restraint is released. However, we did carry out energy calculations that shed light on possible sources of internal strain that might have contributed to this spontaneous opening.

In order to rule out possible artifacts in our computations that might have caused an accidental result, we performed a second MD simulation of *E. coli* monomeric β using the CHARMM program (run β1B) as an independent check of our AMBER run β1A. The major differences between the AMBER and CHARMM programs are in the programming implementation of computational algorithms. In our particular setup, there are other minor specific differences too. For example,

different cations are used in these two MD runs with Na^+ in AMBER and K^+ in CHARMM. The number of total ions is slightly different and the size of the system box is also slightly different (Table 2.1). To draw a not-so-perfect analogy, it is equivalent of solving a crystal structure in two different crystal lattice environments to confirm the central features of a structure. Although trajectory β1B was ended at 2.0 ns, and is shorter than β1A, we did observe the same type of spontaneous opening in the structure. Similar to our analysis for trajectory β1A, the opening process was monitored by measuring the RMS deviation of domain 3 in β monomer when we superimpose domain 1 onto either the crystal structure in the closed conformation (dimer crystal structure) or the crystal structure in the open conformation (βm1:δ complex structure). The RMS deviation as a function of time versus these two reference structures indicated that the monomeric β quickly deviated from its starting closed conformation and merged onto the open conformation in a matter of about 250 ps (Figure 2.10). This transition seemed to happen in a much shorter time scale as compared to β1A run using AMBER (~1.0 ns) and the thermal fluctuations in RMS deviation seemed to be a little larger (Figure 2.9, 2.10). The fact that we did observe the same behavior of spontaneous opening in two different MD simulations using two different programs suggests that the result is not a coincidence. Both trajectories suggest that an internal strain in the monomeric β structure facilitates δ-mediated ring opening in the clamp-loader process.

Figure 2.10 RMS deviation of Cα atoms in CHARMM β1 Run (Trajectory β1B). Rms deviations in Cα atoms in Domain 3 of β monomer are plotted as a function of time. Instantaneous structures from the trajectory are superimposed on Domain 1 of the reference structure, which is the crystal structure of dimeric β for the red trace, and the crystal structure of β in the βm1:δ complex in the blue trace.


25
20
15
10
5
0
RMSD(Angstrom)
1ns
2ns
time
vs. beta from wild type beta dimer
vs. beta from beta-delta complex

2.3.3 MD Simulation of Monomeric *S. pyogenes* β Suggests Similar Structural Relaxation as in *E. coli*

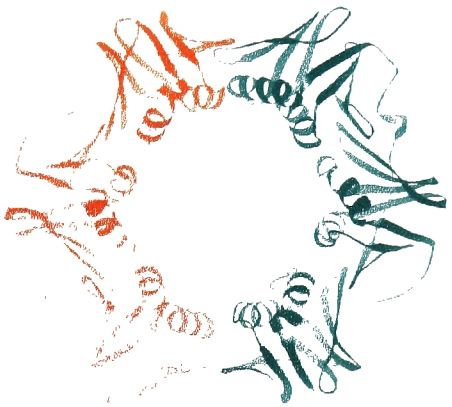
To explore whether this spontaneous opening in a spring-loaded clamp is a universal mechanism employed by different processivity factors, we carried out a simulation of *S. pyogenes* β monomer using AMBER. Although the sequence identity between *E. coli* and *S. pyogenes* β is less than 15%, the overall structure of *S. pyogenes* β dimer resembles that of *E. coli*, with a closed ring formed by two head-to-tail monomers, each consisting of three topologically identical domains (Maria Argiriadi, unpublished results) (Figure 2.11A). Biochemical data also confirmed its function as a sliding clamp that confers processivity on its cognate DNA polymerase, in the same fashion as *E. coli* β does (Bruck and O'Donnell, 2000).

One distinct difference between the structures of these two β subunits is the way that the three domains are aligned relative to each other within a monomer. In *E. coli* these domains are arranged as to produce a pseudo 6-fold symmetry in the ring, and the central channel of the ring has a circular shape. In *S. pyogenes*, within each monomer the angle between domains 1 and 2 is larger than that between domains 2 and 3, so that although the ring is still composed of 6 topologically similar subdomains the hexagonal symmetry is distorted. This irregularity in inter-domain packing produces an oval-shaped central channel, in distinct difference to *E. coli* (Figure 2.11B). We believe the structural differences and low sequence identity between these two β clamps means that *S. pyogenes* β can serve as an independent model system.

Figure 2.11 Sliding clamps from *E. coli* and *S. pyogenes* Ribbons diagram of two different sliding clamps, from *E. coli* and *S. pyogenes* respectively, are shown to indicate the difference in overall curvature of their ring-shaped topology. (A) (left) Structure of dimeric β subunit of *E. coli* with two monomers colored red and blue respectively (Kong et al. 1992). (right) Structure of dimeric β subunit of *S. pyogenes* with two monomers colored red and blue respectively (Maria Argiriadi, unpublished). (B) Schematic representation of the origin of different shapes for *E. coli* and *S. pyogenes* sliding clamps. (top) Structures of one β molecule from *E. coli* and *S. pyogenes* dimer structures, superimposed on domain 2. (bottom) Diagram illustrating the origin of hexagonal shape for *E. coli* and oval shape for *S. pyogenes*.

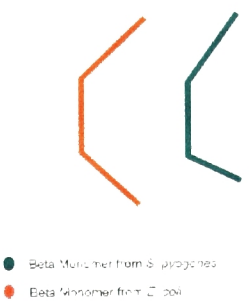
A

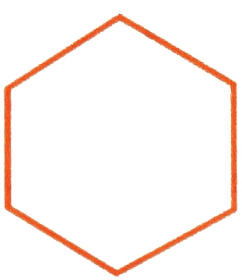
E. coli Beta Dimer



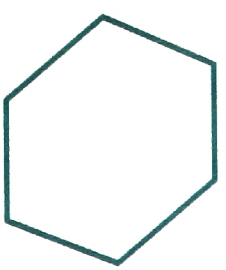
S. pyogenes Beta Dimer

B





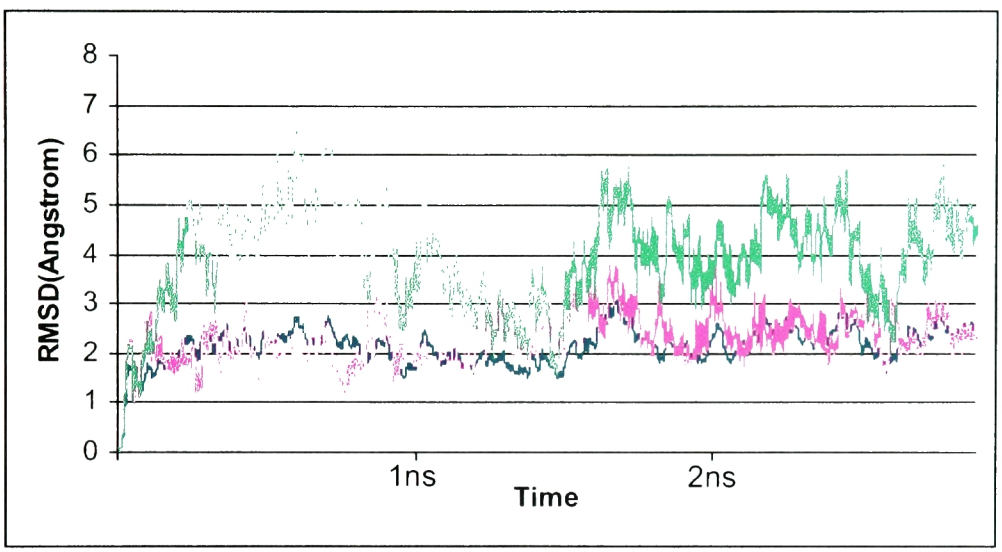
E. coli Beta Dimer's Hexagonal Shape



S. pyogenes Beta Dimer's Oval Shape

An MD trajectory (run β1C) was generated employing the same procedure as used in β1A reported above. We simulated one molecule of *S. pyogenes* β in explicit solvent environment, starting from the conformation as seen in the crystal structure of the dimer, but without the other molecule in the dimer being present. The trajectory generated in this way again showed a spontaneous conformational change in the monomer structure which would disrupt dimer formation, similar to that seen for *E. coli* β. When individual structures in the trajectory are superimposed on domain 2, we saw that domain 3 of *S. pyogenes* β monomer quickly deviated from the original conformation in less than 300 ps, and started to oscillate around a different conformation (Figure 2.12). Averaging instantaneous structures of the molecule in the trajectory after 1.5 ns produced a molecule with reduced curvature as compared to the starting conformation in the trajectory (Figure 2.13). If we superimpose domain 2 of the averaged *S. pyogenes* β structure onto its starting conformation we see that the relaxation happened mostly between domains 2 and 3, which enlarged the angle between them to match that seen in the structure of *E. coli* β, while the angle between domains 1 and 2 remained mostly unchanged (Figure 2.13). When the averaged structure of *S. pyogenes* β is compared to the crystal structure of *E. coli* β monomer in the βm1:δ complex, also superimposed on domain 2, it was satisfying to find that the averaged *S. pyogenes* β monomer structure from the MD trajectory closely resembles the conformation of *E. coli* β monomer in βm1:δ complex, with the same reduced curvature as compared to their closed conformation in the dimer (Figure 2.13).

Figure 2.12 RMS Deviation of Cα atoms in *S. Pyogenes* Run (Trajectory β1C). RMS deviations in Cα atoms in overall structure and domains 1 and 3 are plotted as a function of time. Instantaneous structures from the trajectory are superimposed on Domain 2 of the starting structure, which is the crystal structure of one monomer from dimeric β. RMS deviations plots are colored differently for each comparison as indicated under the plot.



overall structure vs. initial conformation

domain 1 vs. initial conformation

domain 3 vs. initial conformation

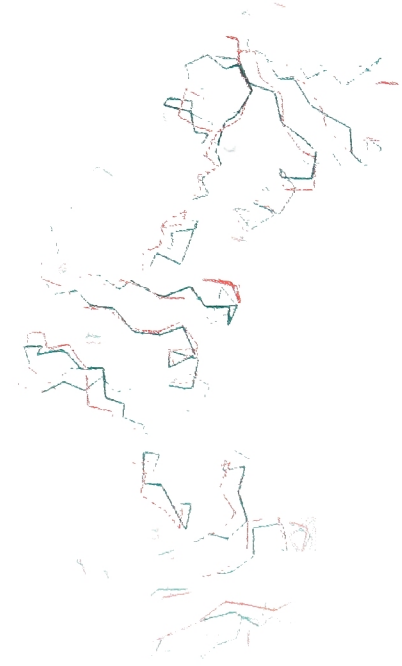
Figure 2.13 Opening of *S. pyogenes* β1 observed I MD (Trajectory β1C). (A) Superposition of the averaged structure of β monomer from MD trajectory (after 1.5 ns) onto the structure of β monomer from dimeric crystal structure, aligned on domain 2. (B) Superposition of averaged structure of *S. pyogenes* β monomer from MD trajectory β1C onto the *E. coli* βm1 structure as seen in the βm1:δ complex structure, aligned on domain 2.

A

Beta Monomer in Dimeric form

Averaged Structure (>1.5ns) from MD

B



E. coli Beta from Beta:Delta Crystal Structure

Averaged Structure (>1.5ns) from MD

A model of a relaxed *S. pyogenes* β dimer generated by superimposing two averaged monomer structures from the simulation onto the original dimer structure with one interface preserved, however, did not produce a gap at the other interface as seen in *E. coli.* Instead the β strands from each molecule at that interface collide, and the domains at this interface twisted out of plane of the ring in opposite directions. Although the structural components utilized to achieve a dimer interface are the same between β of *E. coli* and *S. pyogenes*, domain 1 of one molecule and domain 3 of the other at the interface of *S. pyogenes* β dimer do not interact symmetrically. In *E. coli* these two domains are symmetrical. A small kink or rotation at one interface would be propagated to the other interface as a gross error in our modeling exercise for the *S. pyogenes* clamp. Because we lack an X-ray structure of either *S. pyogenes* β monomer or its opened dimer it is not possible to relate our results obtained from the MD trajectory to experimental data. However, the simulation does indicate that *S. pyogenes* β goes through spontaneous relaxation upon dimer interface disruption. This finding suggests that internal strain may be a common feature in different clamps.

2.3.4 Energy Calculations Point to Decrease in Molecular Mechanical Energy during Structural Relaxation

We performed free energy calculations for the two β monomer trajectories we generated (β1A, β1C), using the MM-PBSA approach. In contrast to the

Figure 2.14 Free energy vs. time in *E. coli* β1 Run (Trajectory β1A). Free Energy values obtained from MM-PBSA method are plotted vs. time for β1A run. Simulation is sampled at a frequency of 10 ps. The red trend line is the moving over 100 ps.

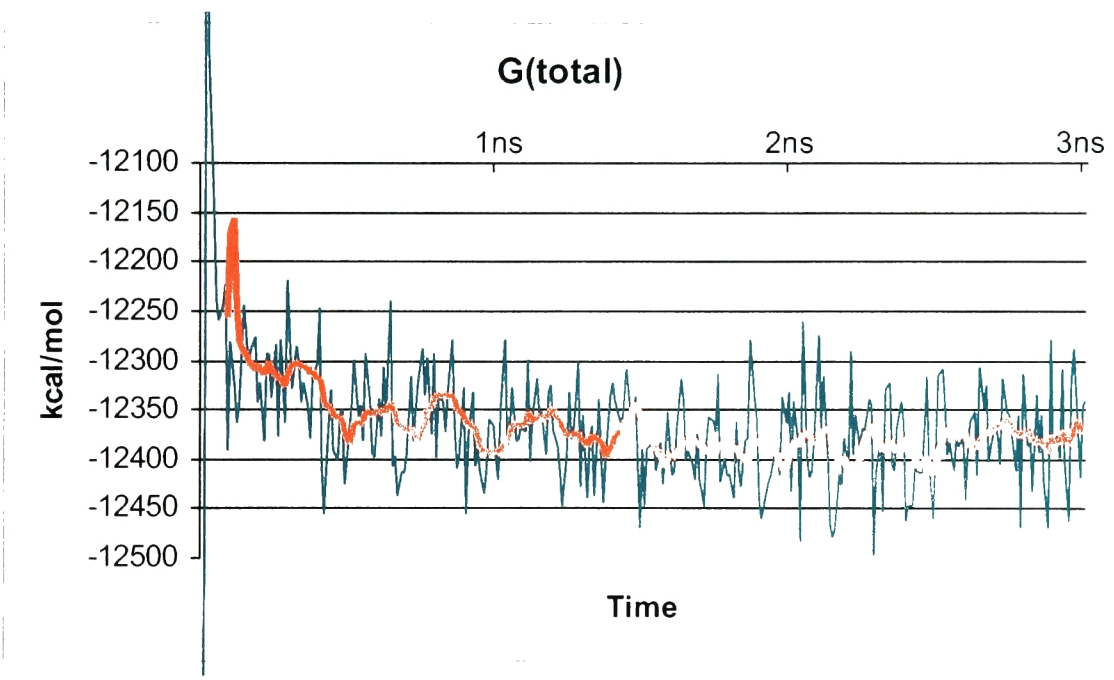
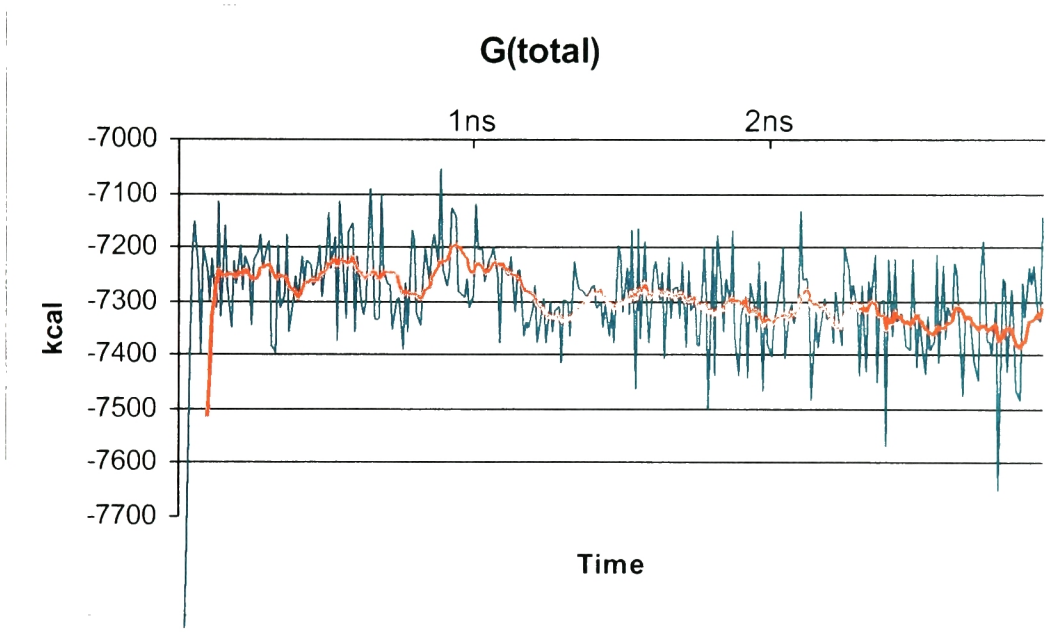

G(total)
1ns
2ns
3ns
-12100
-12150
-12200
-12250
-12300
-12350
-12400
-12450
-12500
kcal/mol
Time

Figure 2.15 Free energy vs. time in *S. pyogenes* Run (Trajectory β1C). Free Energy values obtained from MM-PBSA method are plotted vs time for β1C simulation. Simulation is sampled at a frequency of 10 ps. The red trend line is the moving average over 100 ps.


G(total)
1ns
2ns
-7000
-7100
-7200
-7300
-7400
-7500
-7600
-7700
kcal
Time

gradual relaxation in G until about 1.0 ns as seen in the dimeric β simulation the free energy in β1A and β1B did not stabilize until after 1.5 ns (Figure 2.8, 2.14, 2.15). Although the time scale over which the free energy relaxes is somewhat longer for the monomer simulations it is not clear if this is significant.

In order to identify the nature of the internal strain in the system, we calculated the change in free energy G at the end of simulation as compared to the beginning in all three trajectories β2, β1A and β1C for all decomposd energy terms as dictated by the MM-PBSA method ($\Delta G_{end\text{-}beginning}$). If we identify a change that only exists in the β1A and β1B calculations but not in β2, this is likely to be responsible for the conformational relaxations we observed in our MD simulations. While it is only truly meaningful to look at the sum of all energy terms as a representation of the free energy of a system, we think in our case it is helpful to look at the changes in individual components to locate the possible driving force for the spontaneous opening as observed in our simulations. We first calculated the averaged value for all free energy components at both the initial and final time periods (250 ps) and deduced the difference from those data as listed in Table 2.2 and plotted in Figure 2.16. In order to compare β1A and β1B to β2, all numbers obtained from β2 were divided by a factor of 2 to reflect the fact there are two molecules in the β2 simulation while β1A and β1B only have one.

Only the angle and dihedral energy terms displayed a prominent decrease for both the *E. coli* and *S. pyogenes* β monomer trajectories (MD run β1A and β1C) during the simulations, which was not the case for the *E. coli* β dimer (MD

Table 2.2 Free Energy Calculations from MD Trajectories and X-ray Crystal Structures

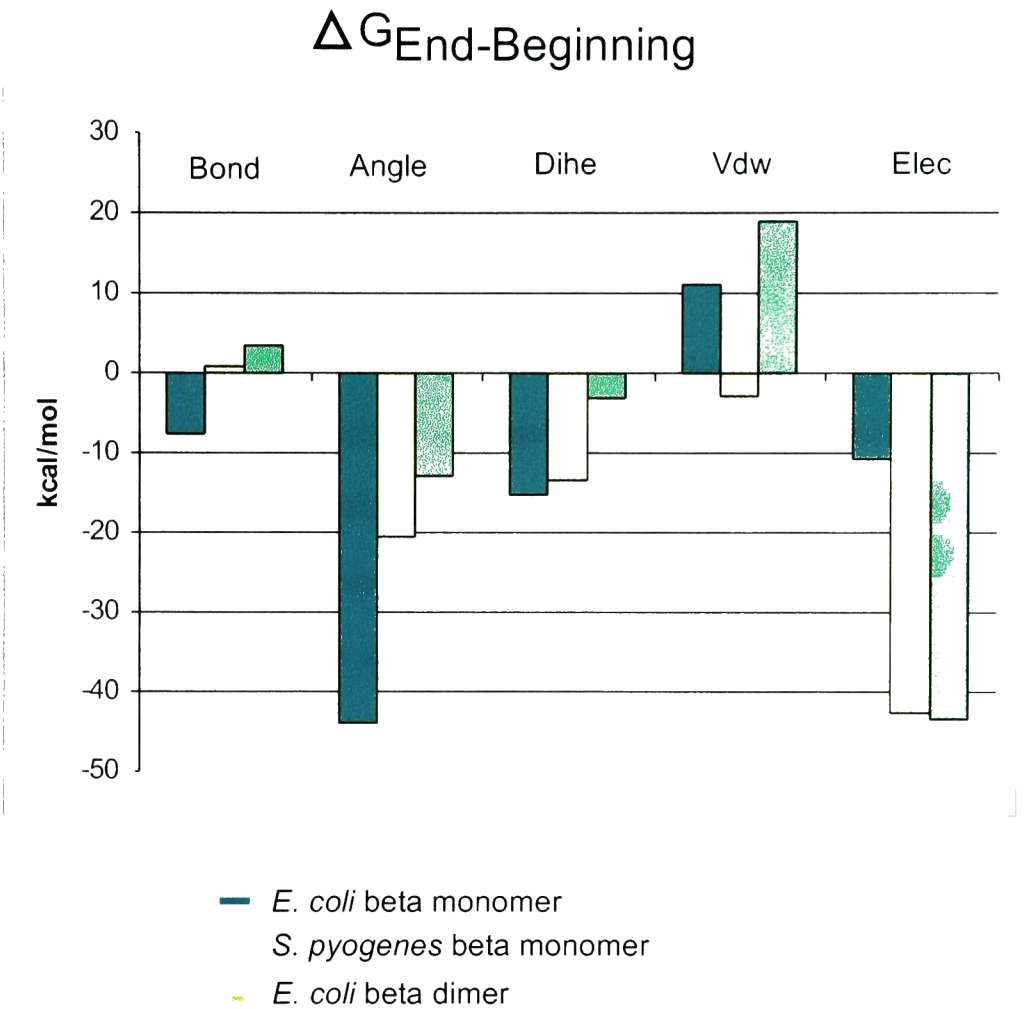
(A) Energy Analysis from Crystal Structures

	Ebond	Eangle	Edihedral	Evdw	Eelec+G_{PB}	G_{SA}	Gtotal
β2	380	1592	3976	-3748	-34294	223	-31871
β2*mol1	186	790	1970	-1824	-16995	123	-15750
βm1	186	729	1917	-1715	-17054	127	-15810
βm1-β2*mol1	0	-61	-53	109	-59	4	-60

(B) Energy Analysis from MD Simulations

Trajectory		Ebond	Eangle	Edihedral	Evdw	Eelec+G_{PB}	G_{SA}	Gtotal
β2(*E.coli* dimer)	Begin(76-298ps)	2089	5616	4382	-3235	-34209	237	-25120
	End(3.0-3.2ns)	2096	5590	4376	-3197	-34296	237	-25194
	$Diff_{End-Begin}$	7	-26	-6	38	-87	0	-74
β1A(*E.coli* monomer)	Begin(60-250ps)	1047	2814	2208	-1541	-16862	127	-12207
	End(2.8-3.0ns)	1039	2770	2192	-1530	-16872	133	-12268
	$Diff_{End-Begin}$	-8	-44	-15	11	-10	5	-61
β1C(*S. pyogenes* monomer)	Begin(61-248ps)	1022	2799	2188	-1478	-14284	128	-9753
	End(2.75-3.0ns)	1023	2778	2174	-1488	-14327	128	-9840
	$Diff_{End-Begin}$	1	-21	-14	-10	-43	0	-87

Figure 2.16 $\Delta G_{[(End)-(Beginning)]}$ Deduced from MD simulations. For three simulation runs β2, β1A, β1C, free energy calculations are performed for 250 ps periods at a frequency of 1ps both at the beginning and the end of the simulations (specific window of time is indicated in table 2.1). Free energy is decomposed as described by MM-PBSA method and the average value for each component for the beginning and end period is obtained. Difference between average value at the final and initial period for each free energy component is plotted. Values from dimeric β simulations were divided by 2 to compensate for the factor of two β molecules in system as compared to one molecule in monomer simulations.


ΔGEnd-Beginning
30
20
10
0
-10
-20
-30
-40
-50
kcal/mol
Bond
Angle
Dihe
Vdw
Elec
E. coli beta monomer
S. pyogenes beta monomer
E. coli beta dimer

run β2). The $(\Delta G_{end\text{-}beginning})_{angle}$ and $(\Delta G_{end\text{-}beginning})_{dihedral}$ difference is ~-44 kcal/mol and ~-15 kcal/mol respectively in trajectory β1A (*E. coli* β monomer) and ~-20 kcal/mol and ~-13 kcal/mol respectively in trajectory β1C (*S. pyogenes* β monomer) but only -12.8 kcal/mol and ~-3 kcal/mol respectively in trajectory β2 (*E. coli* β dimer) (Table 2.2, Figure 2.16). This finding suggests that the higher values of the angle and dihedral energy terms in the initial structure of the MD trajectory, which is of the closed conformation, probably played a role for the spontaneous structural relaxation as seen in MD trajectories for both *E. coli* and *S. pyogenes*.

The change in electrostatic energy, $(\Delta G_{end\text{-}beginning})_{elec}$, which included the molecular mechanical Coulombic electrostatic energy (E_{elec}) and the Poisson-Boltzmann electrostatic solvation free energy (G_{PB}), had little decrease in β1A (~-10 kcal/mol) but a large change in both β1C (~-42 kcal/mol) and β2 (~-43 kcal/mol) (Table 2.2, Figure 2.16). The lack of a large decrease for $(\Delta G_{end\text{-}beginning})_{elec}$ in trajectory β1A is in contradiction to an initial hypothesis that the highly charged *E. coli* β monomer (net change -11) might spring open due to electrostatic repulsion within itself. However, given the lack of a consistent change in the electrostatic energy during the monomer and dimer simulations we are cautious about ascribing a major role to electrostatics in driving the process.

The change in the van der waals energy term, $(\Delta G_{end\text{-}beginning})_{VDW}$, is small for all three trajectories with an increase in β1A (~10 kcal/mol) and β2 (~19 kcal/mol), but a decrease in β1B (~-3 kcal/mol) (Table 2.2, Figure 2.16). This behavior indicates that it probably does not play a role in driving the ring to open

up. The change in bond energy term, $(\Delta G_{end\text{-}beginning})_{bond}$, is also very small for all three trajectories (less than 10 kcal/mol) (Table 2.2, Figure 2.16), is thus also unlikely to play an important role during structural relaxation.

G_{SASA}, the energy term accounting for the surface area contribution to the free energy, does increase in the MD simulation β1A (~5 kcal/mol) and remains unchanged in the β1C trajectory (Table 2.2), reflecting an increase in surface area for relaxed *E. coli* β monomer as compared to its closed state, while *S. pyogenes* β monomer maintains an almost constant surface area whether at open or closed states. In the dimer simulation, this term stayed unchanged, reflecting the stability of the molecule. Increase in G_{SASA} in the *E. coli* monomer trajectories suggests that this term opposes the ring opening process.

In conclusion, we believe that an increase in the angle and dihedral energy terms as the system moves from the open to the closed state is the most likely candidate for the driving force. This aspect is shared by *E. coli* and *S. pyogenes* β monomers for their structural relaxation to their final "open" state. The force field used for the simulations couples the angle and dihedral terms so that they account for the energy of distorting torsion angles in the system. This result therefore suggests that torsional strain in the system is an important component of the driving force.

2.3.5 Free Energy Calculation on Crystal structures of *E. coli* β Monomer Confirmed Findings from MD

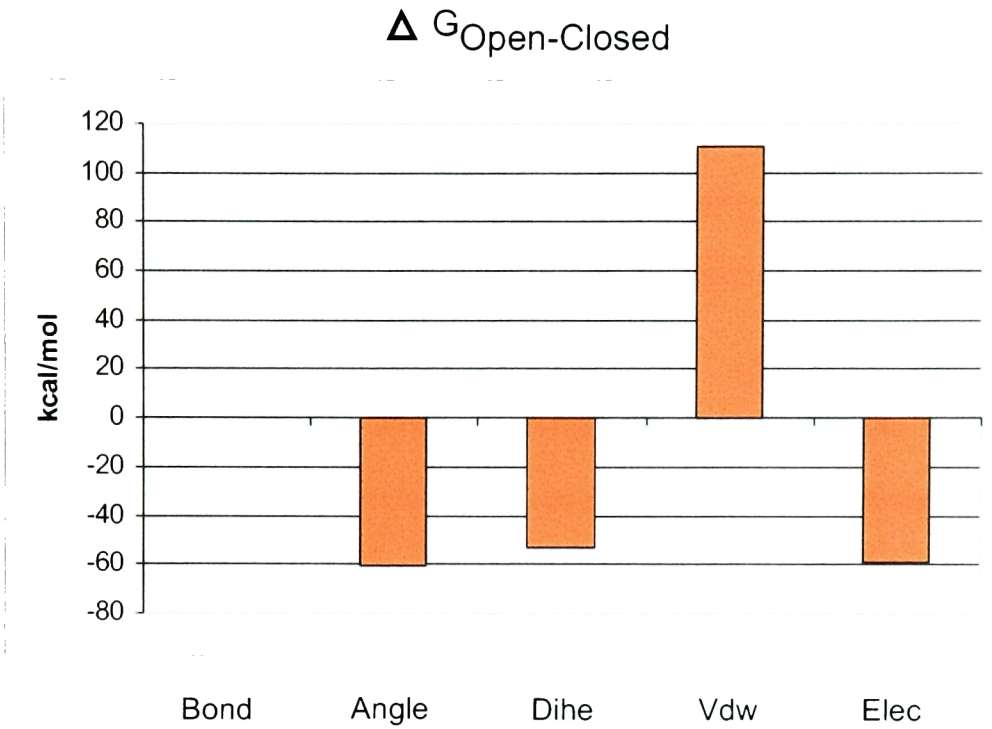
Given that the energetic decomposition of the MD trajectories suggests an important role for torsional strain, we set out to confirm it by comparing the angle and dihedral terms in the energy of the crystal structures of the "open" monomeric βm1 and the "closed" β monomer in the dimer conformation. The idea was that if the angle and dihedral energy is higher in the closed state as compared to the relaxed "open" state then this difference should be manifested in the crystal structures.

First we subjected the crystal structures of monomeric βm1 (as in βm1:δ complex) and closed dimer β2 to steepest-descent energy minimization in CHARMM with the PBEQ module switched on to mimic a solvent environment. This minimization is necessary because it smoothes the transition of the structure from the force field used in CNS or XPLOR (the software packages used for structural refinement for β dimer structure and the δ:βm1 structure) to that of CHARMM. The minimization moved the structures only slightly, with a RMS deviation of 0.25 Å for all heavy atoms between the initial and final structures. MM-PBSA free energy calculation was applied to the energy minimized βm1 structure directly, while the same calculation was done on just one molecule from the energy minimized structure of β dimer. All the energy terms obtained from the crystal structure are listed as in Table 2.2.

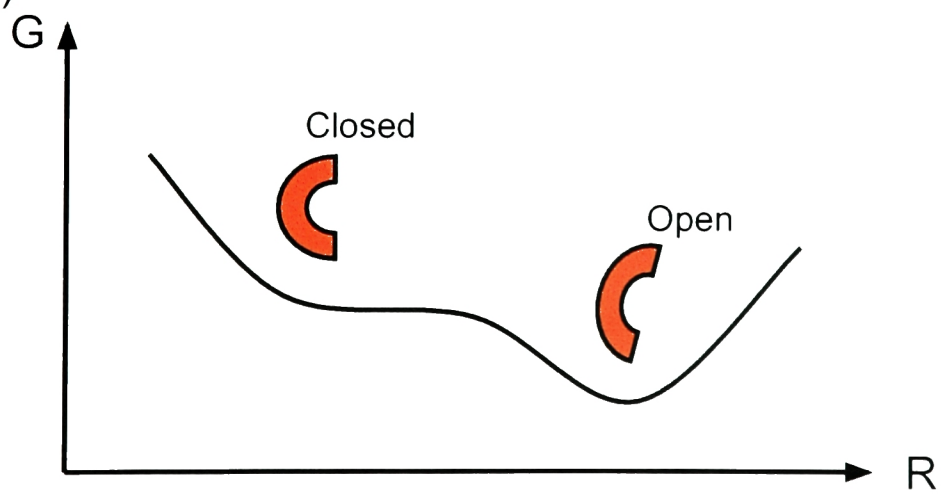
What we see is satisfactory because again the angle and dihedral terms have significantly large differences between the "open" and "closed" states (~-61 kcal/mol for angle term, ~-53 kcal/mol for dihedral term), in agreement with our observation from MD studies (Figure 2.16, 2.17A).

Figure 2.17 $\Delta G_{[(Open)-(Closed)]}$ Deduced from Crystal Structures. (A) Differences in each individual free energy term between two states (open and close) are plotted. (B) Schematic Presentation of Free Energy Landscape Governing β Opening. Plot of free energy vs. reaction coordinates is constructed to illustrate spontaneous opening of monomeric β when dimer restrain is removed.

A)


Δ G Open-Closed
120
100
80
60
40
20
0
-20
-40
-60
-80
kcal/mol
Bond
Angle
Dihe
Vdw
Elec

B)


G
Closed
Open
R

The van der Waals energy term, $(\Delta G_{open\text{-}closed})_{VDW}$, displayed a large increase (~111 kcal/mol). The electrostatic term, $(\Delta G_{open\text{-}closed})_{ELEC}$, including both the E_{ELEC} and G_{PB}, produced a large decrease of ~59 kcal/mol from the "open" state to the "closed" state (Figure 2.17A). The unusual large changes for the two non-bonded energy terms are likely a result of the fact that both the closed and open β monomer structures we utilized for energetic calculations had hydrophobic and ionic interactions with their dimer partner or ring opener (δ) which were omitted in our calculations.

The change in angle and dihedral energy for each individual residue between the open and closed state of *E. coli* β monomer, obtained from energetic calculations on crystal structures, was projected onto the closed monomer structure to identify regions that may have increased contribution to the driving force for structural relaxation. It turns out that the strain in both the angle and dihedral terms is distributed throughout the structure (Figure 2.18).

Ramachandran plots of the "open" and "closed" crystal structures after energy minimization didn't display any notable differences in distribution of ϕ/ψ angles (Figure 2.19A) but the "open" structure had much less deviation in the ω torsion angle as compared to the "closed structure" at the two termini in the structure where the dimer interface would form (Figure 2.19B) and the Chi1/Chi2 distribution plot for residue Leucine is quite different too (Figure 2.19C). The "closed" structure had poor Chi1/Chi2 distribution but this situation was greatly improved in the "open" structure (Figure 2.19C). This suggests that the internal distortion in bond and dihedral angles is unlikely to reside in the secondary

Figure 2.18 Distribution of $\Delta G_{[(Open)-(Closed)]}$ by Residues for Angle and Dihedral energy term. Difference in angle and dihedral term between the "open" conformation (crystal structure of βm1 in βm1:δ complex structure) (Jeruzalmi et al. 2001) and the "closed" conformation (crystal structure of one β molecule in the dimeric structure) (Kong et al. 1992) for each residue is plotted onto the structure using Ribbons presentation. A smooth color gradient is used to color each residue from red to blue when the value of $\Delta G_{[(Open)-(Closed)]dngle}$ or $\Delta G_{[(Open)-(Closed)]dihe}$ varies from -3 to +3. So blue-colored residues would experience an energey decrease from the closed state to the open state while the red-colored residues experience the opposite.

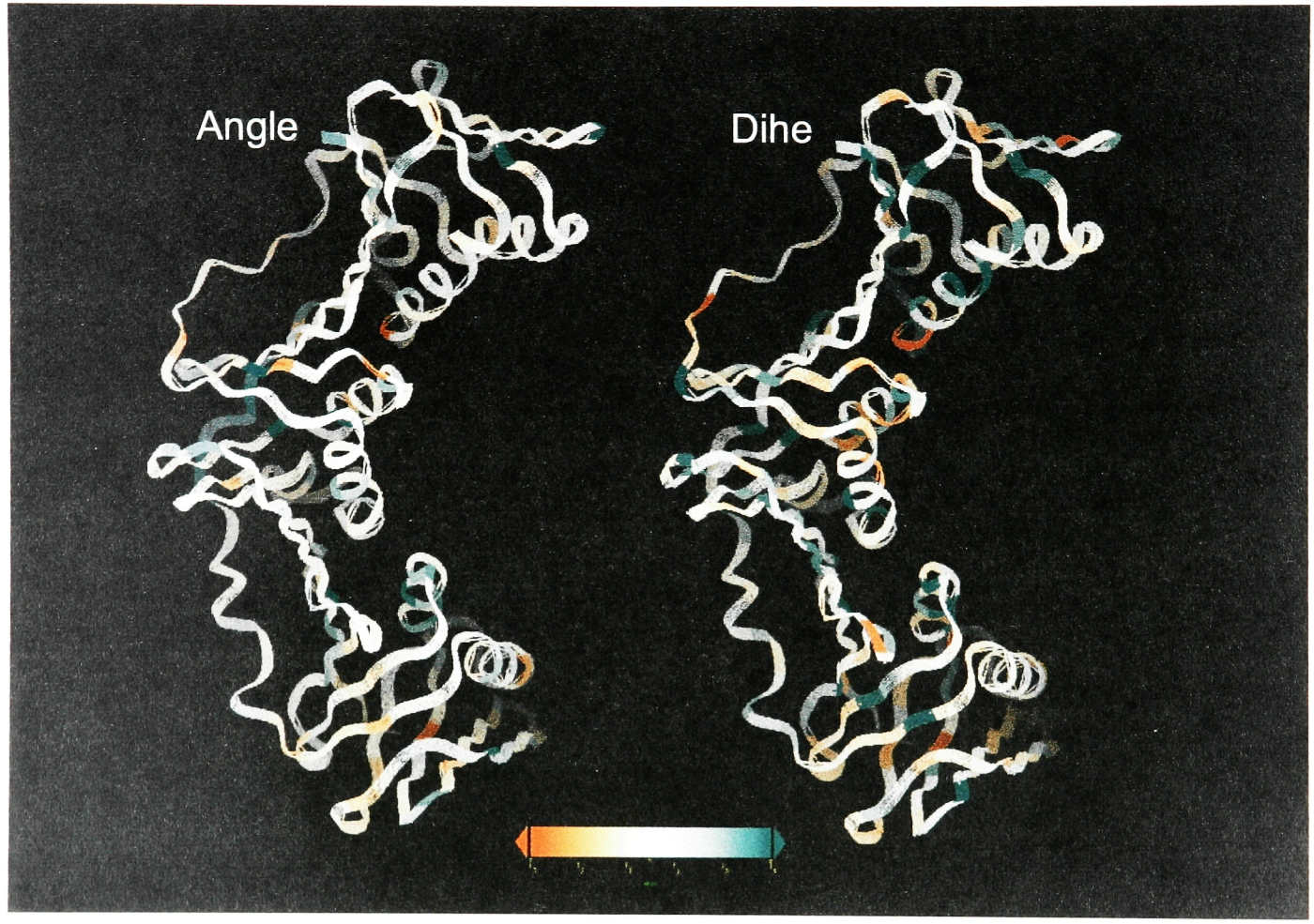
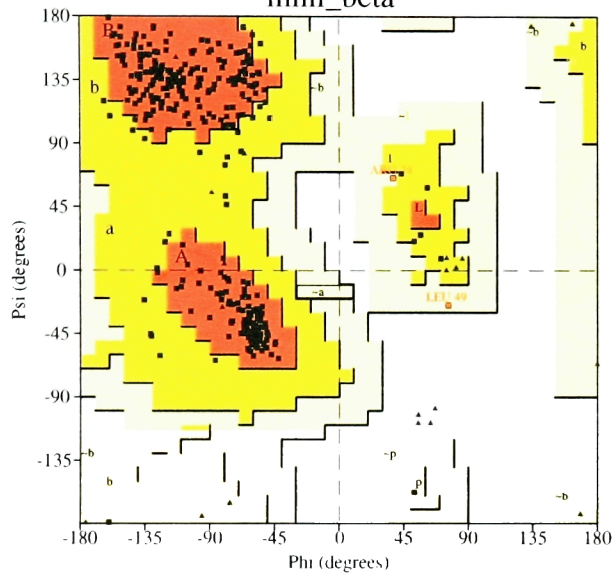

Angle
Dihe

Figure 2.19 Ramachandran Plots of Crystal Structures of "Open" and "Closed" State after Energy Minimization. All these plots are generated using PROCHECK (Laskowski et al. 1993). Plots labeled as "mini_beta" refer to the "closed" crystal structure after energy minimization, "mini_delta" refers to the "open" crystal structure after energy minimization. (A) Plots of ϕ/ψ distribution. (B) Plots of the ω distribution and other relevant terms. (C) Plots of the Chi1/Chi2 distribution.

PROCHECK

Ramachandran Plot

mini_beta



Plot statistics

Residues in most favoured regions [A,B,L]	286	89.7%
Residues in additional allowed regions [a,b,l,p]	31	9.7%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.6%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	319	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	25	
Number of proline residues	20	
Total number of residues	366	

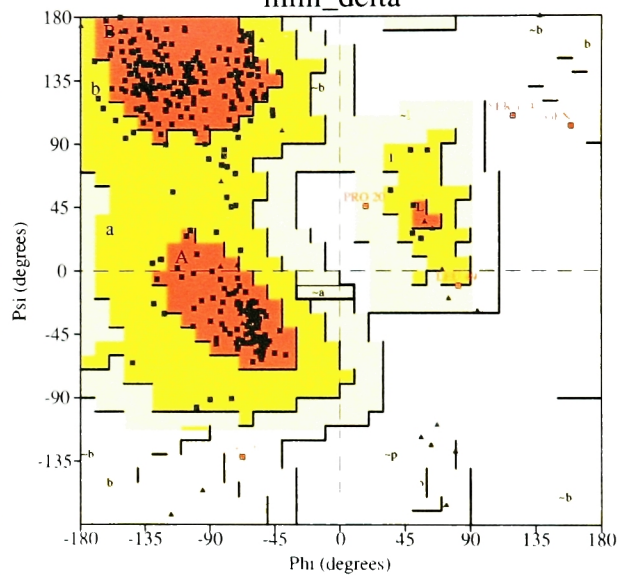
Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

mini_beta_01.ps

PROCHECK

Ramachandran Plot

mini_delta



Plot statistics

Residues in most favoured regions [A,B,L]	278	87.1%
Residues in additional allowed regions [a,b,l,p]	37	11.6%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.9%
Residues in disallowed regions	1	0.3%
Number of non-glycine and non-proline residues	319	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	25	
Number of proline residues	20	
Total number of residues	366	

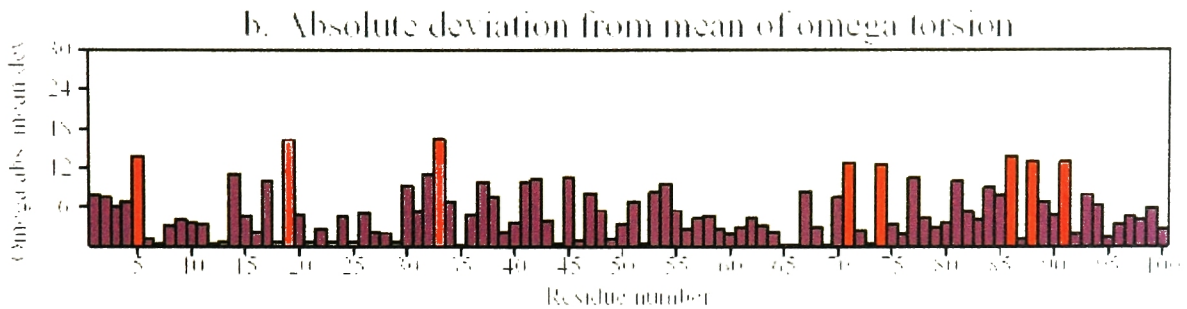
Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

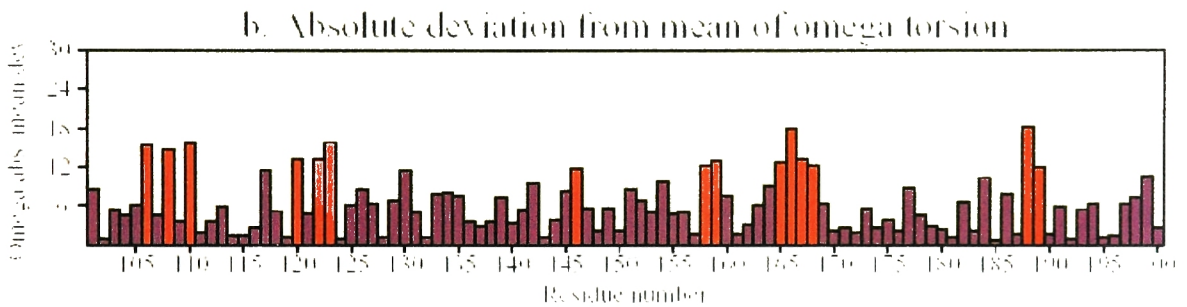
mini_delta_01.ps

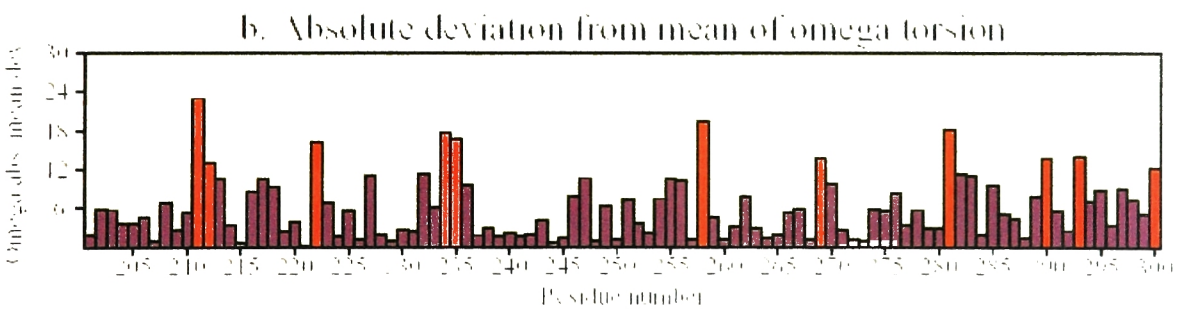
(b)

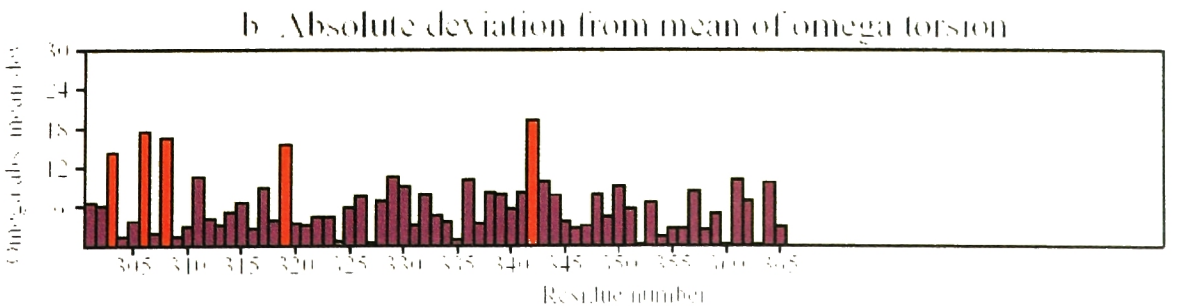
Residue properties

mini_delta

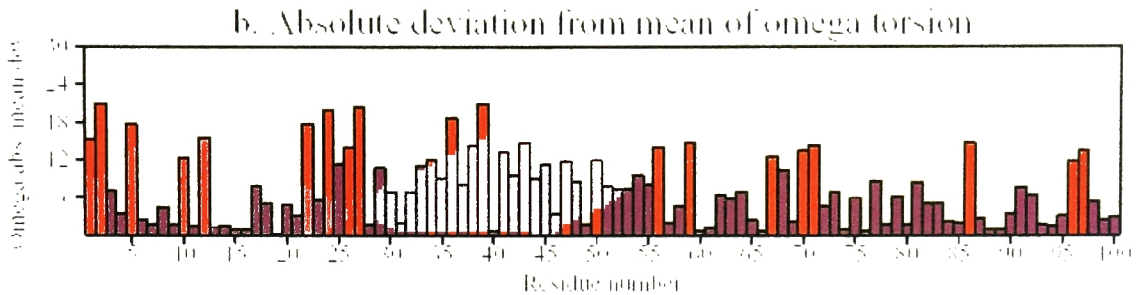


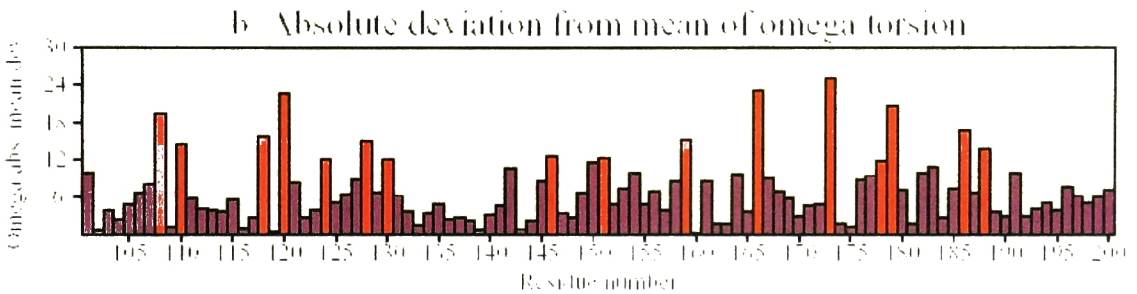


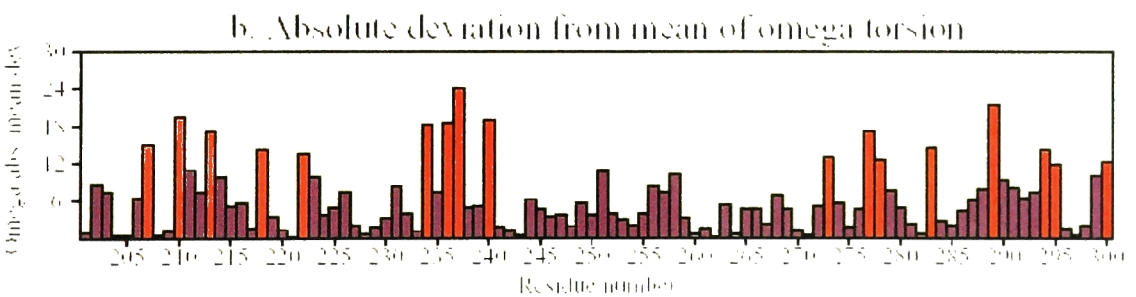


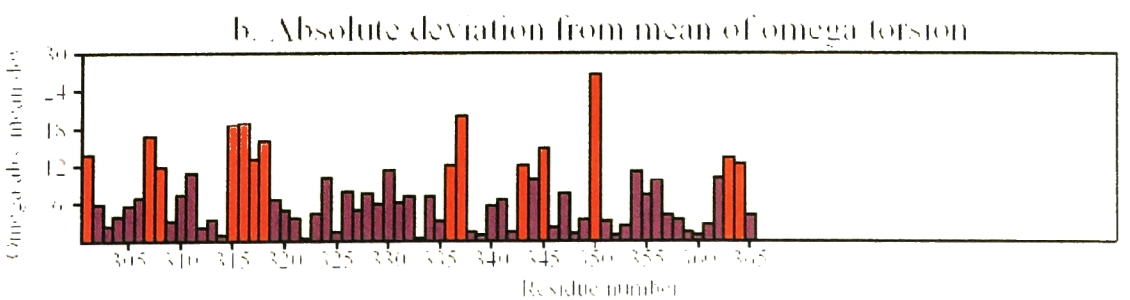


Residue properties
mini_beta



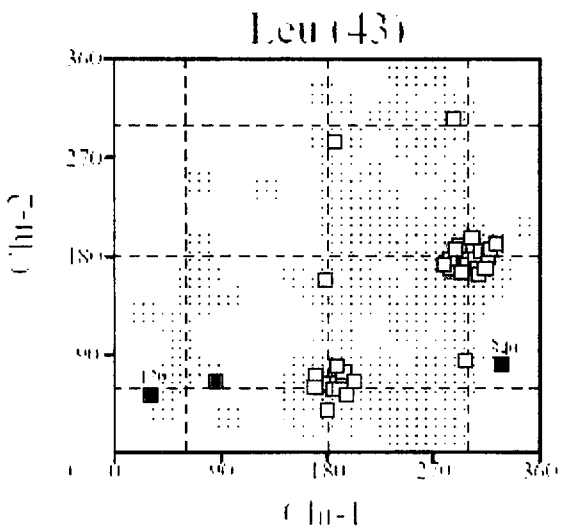




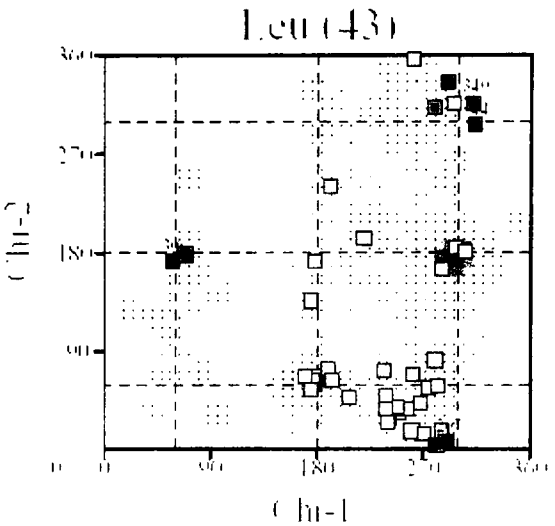


(C) Chi1-Chi2 plots

mini_delta



mini_beta



structural elements (ϕ/ψ) but rather in the imperfect peptide bond angle (ω) at the dimer interface and dihedrals of the side chains for certain residues, such as Leucine.

A free energy profile was constructed based on the above analysis to illustrate the mechanism of spontaneous ring opening process which does not have to overcome an energy barrier (Figure 2.17B). Free energy calculations from both MD simulations and crystal structures pointed to a large decrease in total free energy from the closed to the opened state for monomeric β.

$$\Delta G_{monomer} = G_{open\_monomer} - G_{closed\_monomer}$$

$$\Delta G_{monomer\_MD} = -61 \text{ kcal/mol}$$

$$\Delta G_{monomer\_crystal} = -60 \text{ kcal/mol}$$

By constructing a simple model of opened dimeric ring using two "opened" monomer with one intact dimer interface (Figure 2.20A) and using the various energy terms available from our MM-PBSA calculations performed on the two energy-minimized "open" and "closed" structures (Table 2.2A), we were able to obtain a crude estimate of the ΔG ($\Delta G_{[(open)-(closed)]dimer}$) between the two states as shown in the following steps of calculation.

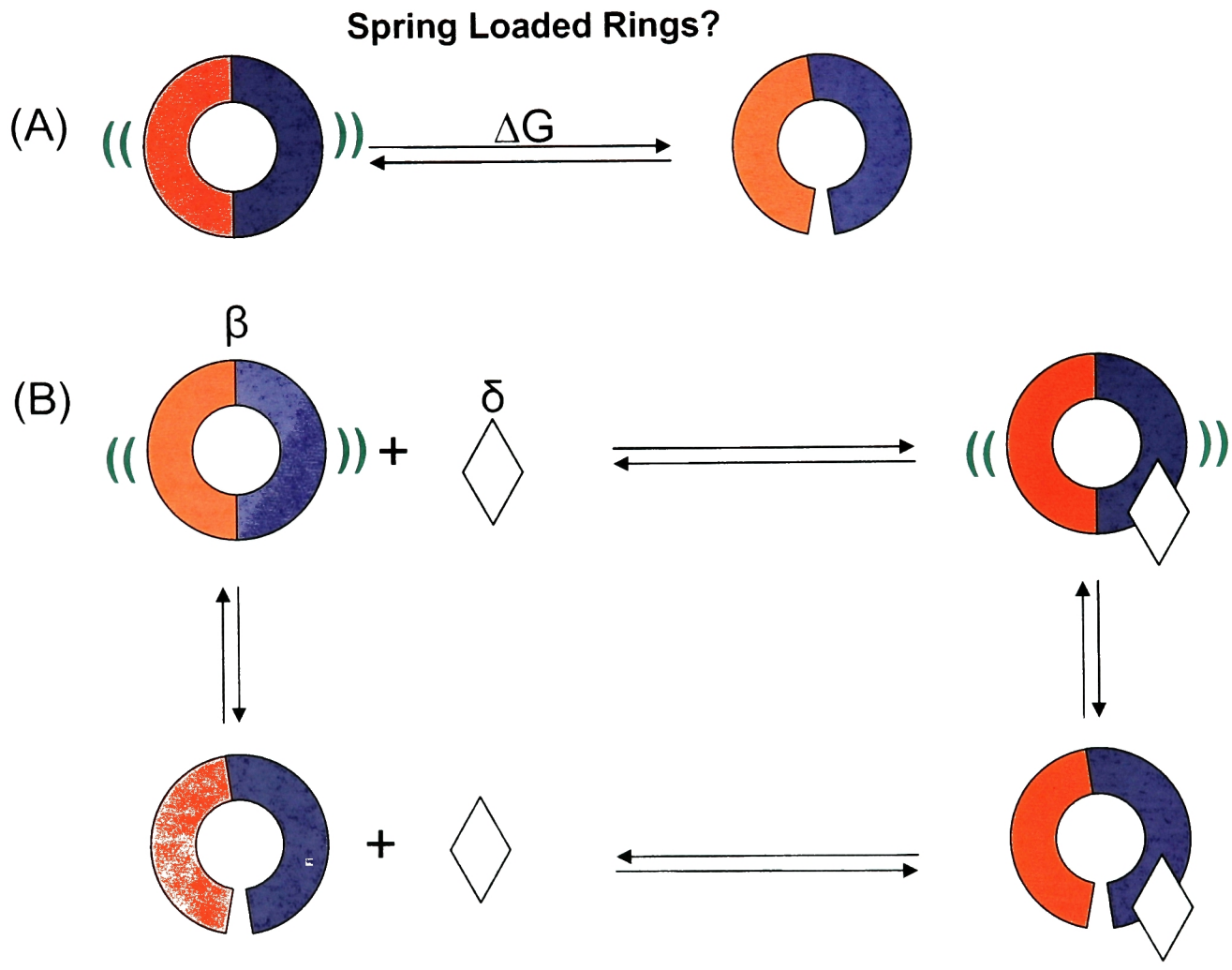
$$\Delta G = 2^{*}\Delta G_{monomer\_crystal} - G_{one\_dimer\_interface}$$

$$= 2^{*}\Delta G_{monomer\_crystal} - (G_{vdw(\beta 2)}/2 - G_{vdw(\beta 2/mol1))}) - (G_{elec(\beta 2)}/2 - G_{elec(\beta 2/mol1))})$$

$$= \sim 82 \text{ kcal/mol}$$

The absolute value of ΔG obtained from our modeling exercise should only be acknowledged at the qualitative level, indicating a likely large energy cost for the ring to open up. Considering that the contribution from the entropy term, which

Figure 2.20 Free Energy Thermodynamic Cycle of Ring Opening and Interaction with δ Subunit. (A) The model of an opened dimeric ring (B) Thermodynamic cycle of the ring opening by δ subunit.


Spring Loaded Rings?
(A)
ΔG
β
(B)
δ
+
+

would favor the opened conformation, has not been included, $\Delta G_{[(open)-(closed)]dimer}$ should be reduced even lower. Energy contributions from binding of δ to β and also interactions with β from other components in the clamp loader complex will likely overcome this energetic barrier to render the ring spring open. Once the clamp loader complex separates from the ring after loading it onto DNA, it is hypothesized that the opened clamp will reform a closed ring driven by this same energetic difference between the open and closed states.

A simplified thermodynamic cycle to reflect the actual ring opening process by δ subunit is shown in Figure 2.20B. However the complete process of ring opening by γ complex as facilitated by ATP binding and hydrolysis is far more complicated and attempts to measure or even theoretically estimate the ΔGs along the path is going to be challenging.

2.3.4 Conclusion

Molecular dynamics simulations of both *E. coli* and *S. pyogenes* β subunits suggest that a spring-loaded mechanism governs the spontaneous opening of β rings once a dimer interface is perturbed. The internal strain for this spring-loaded clamp is likely due to an increase in the angle and dihedral energy terms in the dimeric closed structure as compared to the opened conformation. This strain is likely distributed through the whole molecule. Future studies to further address the question of spring-loaded ring opening are abundant. First a control run of MD simulation should be conducted starting from a closed dimer structure taken from the simulated dimer run (trajectory β2) after equilibration.

The idea behind this is to make sure that if there were some initial internal strain in the structure, it remains in the closed ring even after extensive equilibration in MD simulation. Secondly, to further investigate the contribution of entropy to the process of ring opening, we can borrow a readily available computer program from Willy Wriggers' group (Tama et al., 2002) to compute the normal modes of both the "open" and "closed" structures and deduce the entrophy difference between these two states from the normal mode analysis. Finally we could conduct MD simulations of a carefully constructed model of an open dimer to explore certain features of the opened dimeric ring.

Chapter 3. Detwining of a 2.0 Å Data Set With Perfect Merohedral Twinning for PCNA/p21$^{WAF1/CIP1}$ Complex

3.1. Introduction

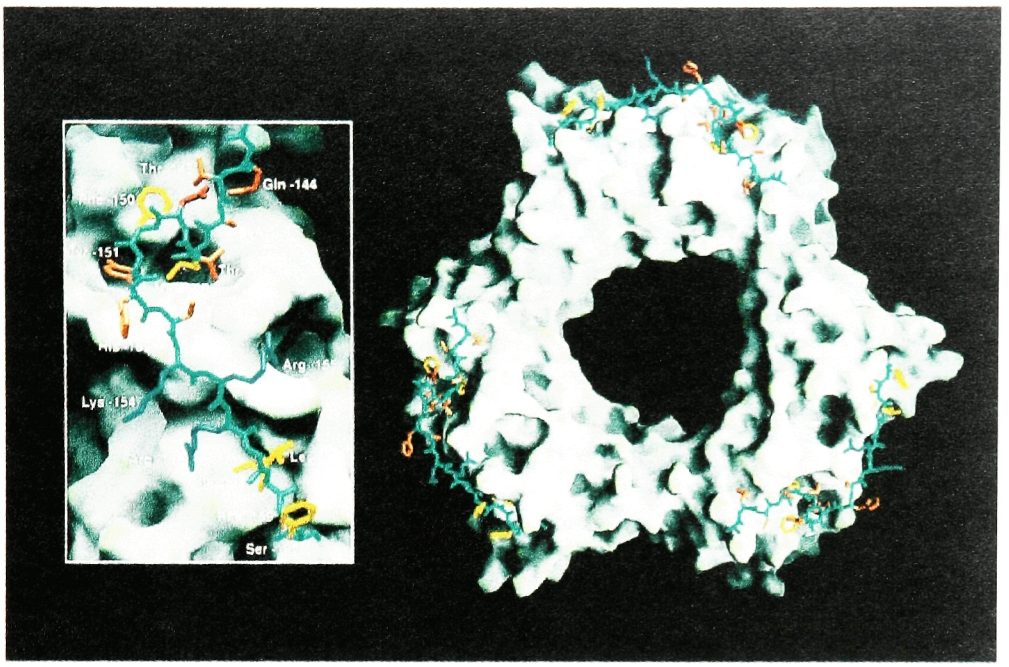
3.1.1 PCNA/p21 $^{WAF1/CIP1}$ Complex

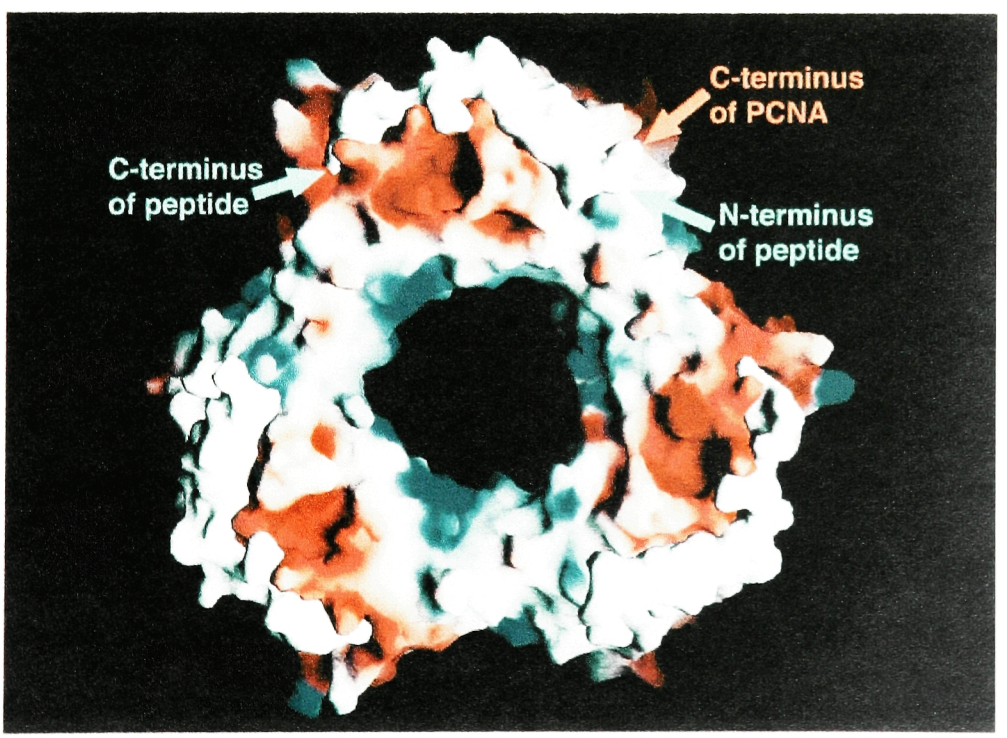
The protein p21$^{WAF1/CIP1}$ (referred to as p21 in this chapter) belongs to a family of inhibitors of cyclin-dependent protein kinases (cdks), which are part of the p53 signaling pathway to stall DNA replication, thus cell mitosis, in response to DNA damage in eukaryotic cells (reviewed by (Peter and Herskowitz, 1994)). p21 is also able to interact directly with human PCNA (hPCNA), the processivity factor for human DNA polymerase δ and ϵ, arresting DNA replication directly (Luo et al., 1995). This PCNA-interacting element in p21 resides in a C-terminal peptide consisting of the last 22 residues of the protein (Chen et al., 1995; Luo et al., 1995; Warbrick et al., 1995). This region is functionally distinct from the N-terminal region, which is responsible for cdk interaction and has sequence similarity with other cdk-interacting proteins, like p27 ^{Kip1} (Polyak et al., 1994), p57 ^{Kip2} (Lee et al., 1995) and p27 ^{XIC1} (Su et al., 1995).

The 2.6 Å crystal structure of human PCNA in complex with the C-terminal 22-residue peptide of p21 (Gulbis et al., 1996) revealed the structure of PCNA as a circular ring formed by three identical monomers with a central hole large enough to accommodate duplex DNA (Figure 2.1, 3.1), a common architectural theme shared by the sliding clamps for *E. coli* and yeast, whose structures were known at that time (Kong et al., 1992; Krishna et al., 1994). Each PCNA molecule is composed of two identical domains linked by a long inter-domain

Figure 3.1 Solvent-Accessible Surface of the PCNA Trimer. (a) The peptide is shown as a stick representation, with side chains colored according to their physical properties (negatively charged residues are drawn in red, positively charged in blue, polar in orange, and hydrophobic in yellow). An enlarged view of the peptide-binding surface is shown in the inset. This figure was generated using GRASP (Nicholls et al. 1991). (b) Molecular surface of the complex, with PCNA colored according to electrostatic potential and the peptide shown in white. In the calculation of electrostatic potential, we used dielectric constants of 2.0 for the protein interior and 80 for solvent at an ionic strength equivalent to 100 mM KCl. Side chains of lysine and arginine residues were given a net positive charge and aspartate and glutamate negative, with other residues considered neutral. Regions of intense positive charge appear blue, and electronegative regions of the surface are red (Reproduction from Gulbis et al. 1996) (Permission from John Kuriyan).

a


Gln -144
-151
Ser


C-terminus
of PCNA
C-terminus
of peptide
N-terminus
of peptide

b

connector loop running across the outer surface of the ring structure. Each monomer in the trimeric ring has one p21 peptide bound to it (Gulbis et al., 1996) (Figure 2.3C, 3.1), in agreement with the approximate 1:1 stoichiometry observed experimentally (Li et al., 1994). p21 peptide binds on the outer surface of the PCNA molecule, leaving the central channel and trimer interface intact, consistent with experimental data indicating that p21 binding to PCNA does not affect its sliding on DNA (Podust et al., 1995). Anchored by hydrophobic interactions between three residues in the central region of the peptide and a conserved hydrophobic pocket in PCNA underneath the connector loop, the C-terminal region of the peptide forms an anti-parallel β sheet with the interdomain loop through extensive hydrogen bonding. The N-terminal region of peptide extends to interact with C-terminal residues of PCNA, which are disordered in the structure.

The fact that p21 peptide does not interfere with the sliding clamp feature of PCNA, i.e. the PCNA trimer still forms a ring, suggests that p21 inhibits DNA replication by disrupting the association between PCNA and its DNA polymerase. This is consistent with experimental observation (Luo et al., 1995). Therefore, the interaction regions between p21 peptide and PCNA most likely identify potential PCNA surface area that is important for polymerase association. Subsequently, the crystal structure of the processivity factor from RB69 bacteriophage complexed with a C-terminal peptide of its associated DNA polymerase revealed a striking similarity between the nature of this interaction and that between p21 and PCNA (Shamoo and Steitz, 1999). This finding suggests that indeed the

processivity factor has one binding site shared between two molecules of opposing function, the DNA polymerase and the replication inhibitor p21. Increased levels of cellular concentration for p21 as part of the p53 pathway upon detection of DNA damage (Waldman et al., 1996) probably results in p21 winning over DNA polymerase in their competition for PCNA binding.

The interaction regions between p21 and PCNA are targets for the design of small molecules that might mimic the action of the p21 peptide. This has potential therapeutic application in cancer treatment because over-proliferating cancerous cells have markedly high level of DNA replication activity. For structure-based small molecule design it is usually helpful to have a structure with a sufficiently high resolution (usually ~ 2.0 Å) to offer detailed description and accurate geometry of the interactions of interest. This is especially true for the case of p21/PCNA complex because p21 peptide interacts with PCNA at numerous locations through hydrogen bonding, ion pairing and hydrophobic packing.

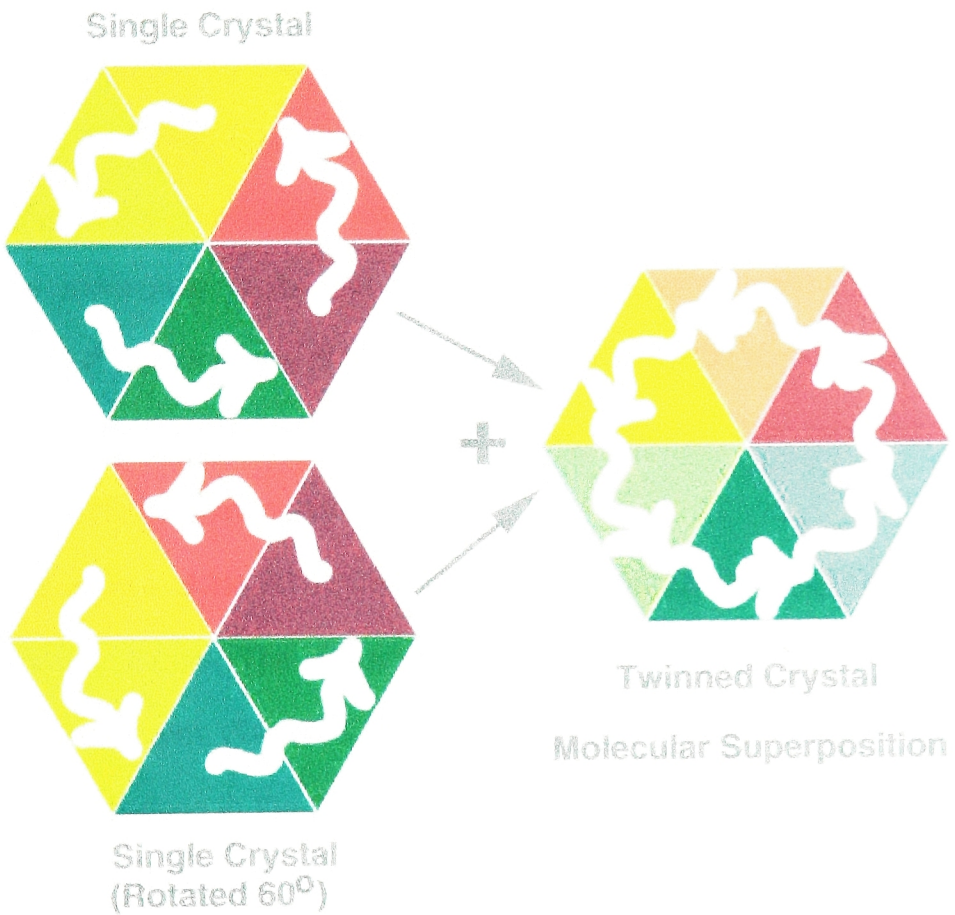
Jacqueline Gulbis, a former postdoc in our lab who solved the 2.6 Å structure as described above, did obtain a crystal form of the complex that diffracted to 2.0 Å during her crystallization trials (Gulbis et al., 1996). Unfortunately it suffered from perfect hemihedral twinning. How this twinning was identified is described in the following section. I set out to deconvolute the twinned data ("detwin") to obtain a higher-resolution structure of this complex (at 2.0 Å), hoping to discover more details of the interaction between p21 peptide and PCNA. This project also provided valuable training in understanding the

complexity of crystal symmetry and space groups, a common issue of perplexity to crystallographers.

3.1.2 Twinning of Crystals

Protein crystals are formed by orderly repetition of a basic unit according to certain symmetry. Twinning arises when one single crystal contains two or more crystal domains arranged in a special relationship, but with the same basic unit and internal symmetry. Epitaxial twinning referred to the kind where the twinning operation, the symmetry operation relating one crystal domain to the other, is not a member of the lattice symmetry of the crystal. Crystals containing this kind of twinning are easily identified because monochromatic X-ray diffraction patterns from these kinds of crystals will often (but not always) present distinct reciprocal lattices for each crystal domain. In merohedral twinning, the twinning operation is a member of the lattice symmetry but not a member of the crystal symmetry. This is possible when the crystal lattice is capable of supporting higher symmetry than the space group of the crystal (Figure 3.2). Crystals suffering this kind of twinning are hard to detect at first sight because the twinning operation exactly superimposes the reciprocal lattices of individual crystal domains, generating a diffraction pattern that is not visibly abnormal. Hemihedral twinning refers to the simplest of merohedral twinning in which there are only two crystal domains with a special orientation to each other. A perfect hemihedral twinning means the crystal contains an equal weight of the two orientations (Koch, 1992).

Figure 3.2 Twinning of P3 symmetry into P6 symmetry.


Single Crystal
+
Twinned Crystal
Molecular Superposition
Single Crystal
(Rotated 60°)

Perfect merohedral twinning in crystals usually escapes initial detection because the higher symmetry created by the twinning operation is mistaken as the true crystallographic symmetry. This kind of special twinning is eventually detected because it can prevent a successful structure determination, since the space group assignment is wrong. Initial suspicion of perfect merohedral twinning can usually be confirmed by analysis of the crystal-packing density and intensity statistics of the diffraction data set. Twinning usually fools a crystallographer into processing data using a higher symmetry space group, thus creating impossibly too many molecules in its unit cell. Intensity statistics dictate that the ratio $<I^2(h)>/<I(h)>^2$ in thin shells of resolution for acentric reflections is expected to be 2.0 for normal crystals and 1.5 for crystals with merohedral twinning (Redinbo and Yeates, 1993; Stanley, 1972).

In the case of the p21/PCNA complex, we discovered the twinning through the following steps. The Laue group of the data from twinned crystals corresponds to 6/m (Table 3.1). With screw symmetry ruled out by lack of systematic absences in the data, P6 is the only hexagonal space group suitable. Due to the trimeric nature of PCNA molecule it cannot center on the 6-fold axis at the origin. Given the size of the cell (a=143.2Å, c=41.4Å) and the dimensions of PCNA (~85Å diameter for the ring, with a thickness of ~40 Å), a location of the PCNA trimer not on the two 3-fold axes would generate symmetry-related PCNA molecules overlapping with each other. The only remaining possibility is for PCNA trimers to be centered around the two 3-fold axes in each P6 cell. However, this possibility was ruled out because the native Patterson map

Table 3.1 Statistics for Data Processing

Table 3.1. Statistics for Data Processing in P3 and P6 for Perfectly Merohedrally Twinned p21 PCNA Complex Crystal

	6 m Data Processing				3 m Data Processing			
Resolution Shell (Å)	<I>	σ(I)	R_{sym}	Completeness(%)	<I>	σ(I)	R_{sym}	Completeness(%)
20.0 – 4.56	21431.6	625.9	0.042	88.3	21392.5	816.5	0.086	85.2
4.56 – 3.63	21458.4	665.6	0.056	92.4	21526.1	889.1	0.056	87.9
3.63 – 3.17	12258.3	402.7	0.061	96.6	12074.3	532.3	0.057	92.5
3.17- 2.88	6051.1	233.3	0.069	98.6	6069.8	321.9	0.064	94.4
2.88 – 2.68	3597.3	174.5	0.089	99.2	3587.8	242.4	0.084	95.1
2.68 – 2.52	2606.0	153.4	0.109	98.7	2614.4	213.7	0.102	94.7
2.52 – 2.39	1820.4	138.2	0.128	98.2	1837.8	194.6	0.119	94.6
2.39 – 2.29	1502.3	137.0	0.152	98.8	1507.2	194.8	0.141	95.8
2.29 – 2.20	1331.4	150.2	0.273	99.3	1361.6	215.6	0.334	96.0
2.20 – 2.13	1070.8	143.3	0.223	99.6	1076.3	204.8	0.206	96.3
2.13 – 2.06	811.2	150.3	0.337	99.3	836.9	214.8	0.376	95.0
2.06 – 2.00	592.8	153.1	0.344	95.6	606.8	218.1	0.311	89.3
All Reflections	6082.1	257.3	0.075	97.0	5885.5	347.5	0.088	93.1

contains strong peaks (~1/2 to 1/3 the height of the origin peak, depending on the resolution) that correspond to translations in the horizontal x,y plane between the locations of rotation axes in the cell, as well as translations of ~1/3 along the vertical axis. The vertical translation is inconsistent with space group P6, which requires all the PCNA rings in a unit cell to lie in the same plane. Space group P6 is therefore ruled out. Given the 3-fold symmetry of PCNA, we suspected that the true space group is P3, with merohedral twinning resulting from equal mixtures of different lattices that are rotated by 60° with respect to each other (Figure 3.2). This is confirmed using intensity statistics. The ratio of $<I^2(h)>/<I(h)>^2$ is close to 1.6 for the twinned PCNA/p21 peptide crystal form, whereas a ratio of ~2.0 is obtained for the normal crystal form which was used to solve the structure at 2.6 Å.

3.1.2 Detwinning Algorithm for Molecular Replacement Method

Normal methods of structure determination cannot succeed in a case of merohedral twinning because the intensity for each diffraction spot collected is a sum of contributions from two different crystal orientations (Redinbo and Yeates, 1993).

$$I_{obs}(\mathbf{h}_1) = (1-\alpha)\, I(\mathbf{h}_1) + \alpha\, I(\mathbf{h}_2)$$

$$I_{obs}(\mathbf{h}_2) = \alpha\, I(\mathbf{h}_1) + (1-\alpha)\, I(\mathbf{h}_2) \qquad (3.1)$$

$\mathbf{h}_1$, $\mathbf{h}_2$ are two reflections (h_1,k_1,l_1), (h_2,k_2,l_2) related by the twinning operation, α is the twinning fraction. $0 \leq \alpha \leq \frac{1}{2}$. In perfect twinning, $\alpha = \frac{1}{2}$. This resulted in an apparent higher symmetry because

$$I_{obs}(\mathbf{h}_1) = I_{obs}(\mathbf{h}_2) = [I(\mathbf{h}_1) + I(\mathbf{h}_2)]/2 \qquad (3.2)$$

Rebindo and Yeates (1993) have shown that with a reasonably good starting model for molecular replacement, the two twinned components in the intensity can be deconvoluted in an iterative process (Redinbo and Yeates, 1993). A simplified description of this algorithm starts with the assumption that $I_{calc}(\mathbf{h}_1)$, obtained from the starting model, can be regarded as a close estimation of $I(\mathbf{h}_1)$, the real intensity for that diffraction if there was no twinning. Thus we can deduced from equation 3.2 that

$$I(\mathbf{h}_1) = 2^* I_{obs}(\mathbf{h}_1) - I_{calc}(\mathbf{h}_2)$$

$$I(\mathbf{h}_2) = 2^* I_{obs}(\mathbf{h}_2) - I_{calc}(\mathbf{h}_1) \qquad (3.3)$$

Least-square optimization reduces these equations to

$$I_{detwin}(\mathbf{h}_1) = [I_{obs}(\mathbf{h}_1) + I_{calc}(\mathbf{h}_1) - I_{calc}(\mathbf{h}_2)]/2 \qquad (3.4)$$

while ignoring the arbitrary factor of 2. Considering the fact that

$$F_{detwin}(\mathbf{h}_1) = [I_{detwin}(\mathbf{h}_1)]^{1/2}$$

$$F_{detwin}(\mathbf{h}_1) = [I_{detwin}(\mathbf{h}_1)]^{1/2} \qquad (3.5)$$

A standard refinement procedure can now proceed by refining $F_{calc}(\mathbf{h}_1)$ against $F_{detwin}(\mathbf{h}_1)$ as if they are a matching pair of $F_{calc}(\mathbf{h}_1)$ and $F_{obs}(\mathbf{h}_1)$. The following definition for R factors are utilized to evaluate the model structure.

$$R_{untwin} = \sum |F_{calc}(\mathbf{h}_1) - F_{detwin}(\mathbf{h}_1)| \; / \; \sum F_{detwin}(\mathbf{h}_1)|$$

$$R_{twin} = \sum |[F^2_{calc}(\mathbf{h}_1) + F^2_{calc}(\mathbf{h}_2)]^{1/2} - F_{obs}(\mathbf{h}_1)| \; / \; \sum F_{obs}(\mathbf{h}_1)| \qquad (3.6)$$

After initial round of detwinning and refinement using the above algorithm this deconvolution process is repeated to improve the estimation of $F_{detwin}(\mathbf{h}_1)$. This iterative process continues until R factors converge. The weakness of this

procedure is that the parameters of the model enter into the estimate of the untwined structure factor amplitudes; the procedure is therefore biased by the starting model. However, Redinbo and Yeates (1993) showed that when a good starting model is available, the procedure is able to yield interpretable electron density maps with new information not contained in the model. For the case of a plastocyanin crystal form that suffered from merohedral twinning (Laue group of data, 6/m; true space group, $P3_2$), refinement at 1.5 Å resolution using deconvoluted structure factors resulted in the incorporation of reliable solvent molecules into the model (Redinbo and Yeates, 1993).

We carried out the detwinning of our p21/PCNA data set using the above described detwinning algorithm. Based on our analysis on twinning, we processed our data set in P3 symmetry. Rotational and translational searches led to three PCNA molecules being put onto the three 3-fold axes in the unit cell. Detwinning and structure refinement followed using this initial model and achieved success in detwinning, bringing out new structural details at several regions previously disordered in the starting model.

3.2 Methods and Materials

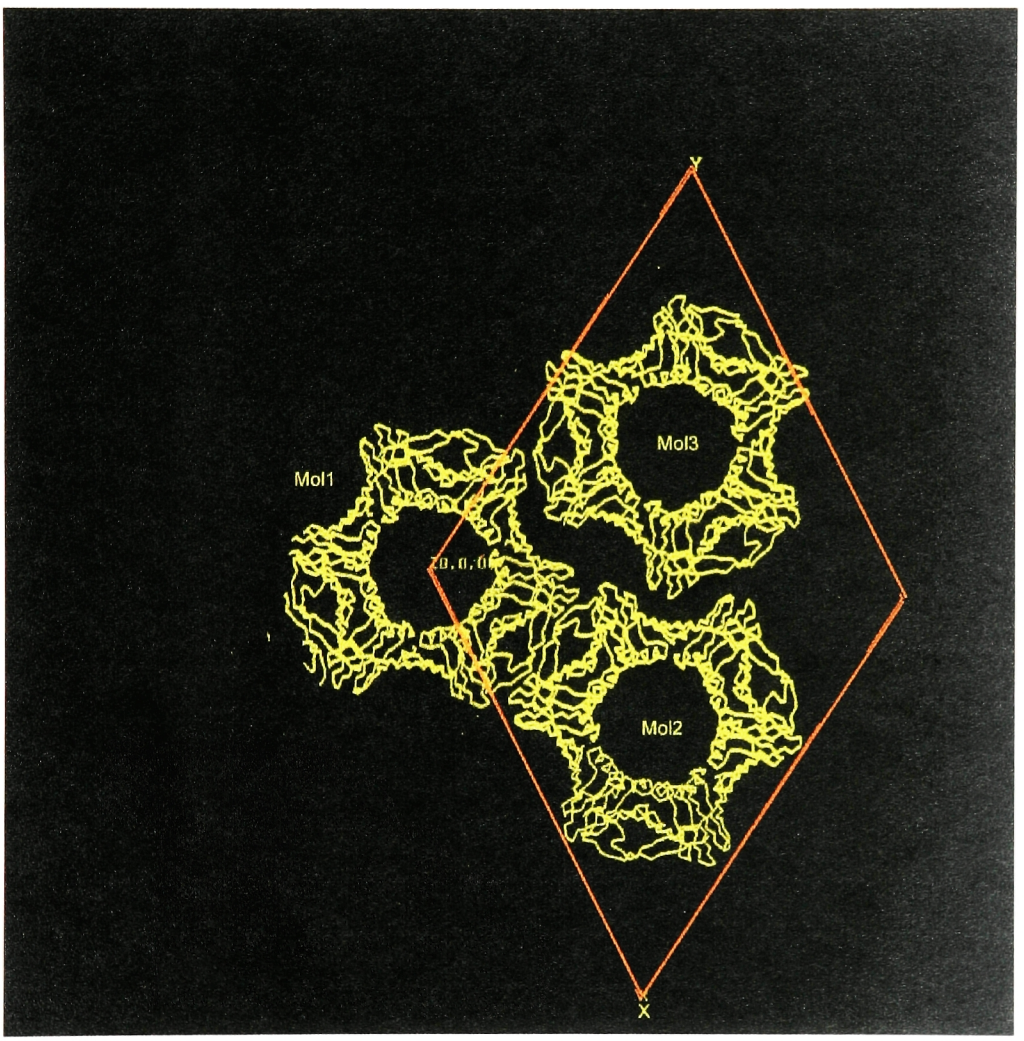
3.2.1 Molecule Placement using XPLOR

The search to place PCNA molecules in the P3 unit cell was done using XPLOR (Brünger, 1992). The twinned 2.0 Å data set initially processed in symmetry P6 was first expanded to P1 symmetry and subsequently reduced to P3 symmetry using XPLOR. The 2.6 Å structure of p21/PCNA complex (Gulbis et al., 1996) (RCSB deposition code 1AXC) without the p21 peptide was used as the starting model. The possibility of the PCNA trimer being centered at (0,0,0), (1/3,2/3,z) and (2/3,1/3,z') in the unit cell was explored sequentially by rotational and translational searches using data from 10.0 – 4.0 Å. This resulted in placing one PCNA molecule at each of the three sites (Figure 3.3). The z coordinate and orientation of each molecule was identified by generating a peak value above noise level in Patterson correlation coefficients calculated from each step of the search. This initial model of three PCNA molecules was then used in the detwinning and refinement steps.

3.2.2. Detwinning and Structure Refinement

Implementation of detwinning was done using X-PLOR (Brünger, 1992). $F_{calc}(\mathbf{h}_2)$ is obtained by calculating F_{calc} from a model rotated by 60° as compared to that of $F_{calc}(\mathbf{h}_1)$. The F_{obs} array in XPLOR was modified to be F_{detwin} according to equation 3.4 and 3.5. Subsequent structure refinement using $F_{calc}(\mathbf{h}_1)$ and F_{detwin} are done using a standard protocol, including positional refinement, simulated annealing and tightly constrained temperature factor refinement in XPLOR

Figure 3.3 Placement of Molecules in P3 Unit Cell. Three molecules of PCNA are placed based on self-correlation coefficient calculations. Molecule 1 is placed at the (0,0,0). Molecule 2 is placed at (1/3, 2/3, 0.08) and molecule 3 is placed at (2/3, 1/3, 0.365). The trimer molecules are generated by symmetry.


Y
Mol3
Mol1
Mol2
X

(Brünger, 1992). 2Fo-Fc electron density maps were calculated and then improved by 3-fold symmetry averaging using the program RAVE (Kleywegt and Jones, 1994). The initial resolution for detwinning and refinement included data from 50.0 – 2.6 Å. Electron density maps after three rounds of detwinning and refinement showed clear density for the p21 peptide, which was not included in the starting model (Figure 3.3). Subsequent iterative detwinning and refinement included data up to 2.3 Å. Density for residues 186-191, 106-109 which were disordered in starting model appeared in this higher resolution map. R_{twin} according to equation 3.6 is 21.4% and free R_{twin} is 23.7%.

3.3 Results and Discussions

3.3.1 Building a start model for detwinning

With the possibility of P6 ruled out and the perfect hemihedral twinning confirmed in our 2.0 Å data set of the p21/PCNA complex, we identified the real space group of the crystal to be P3 and reprocessed our data in P3 symmetry. Two P3 lattices with the same unit cell dimensions but a 60° rotation between them about c produced the higher 6-fold symmetry seen in the diffraction data (Table 3.1). The choice of P3 as the space group requires that the PCNA molecules be centered around one or more of the three 3-fold axes in the cell. Ambiguities about how to place these molecules, such as how many molecules exist in the unit cell, their rotation and translation about each axis, and whether the ring is to be placed "face up" or "face down" (i.e., flipped about an axis in the x,y plane) were resolved by carrying out systematic searches and the calculation of Patterson correlation coefficients for each step of the search using XPLOR (Brünger, 1992). These searches revealed that there are three molecules in the asymmetric unit of the unit cell (Figure 3.3). Their locations are (0,0,0), (1/3, 2/3, 0.08) and (2/3, 1/3, 0.365) respectively. The fact that one of the molecules is displaced by ~1/3 along Z with respect to the others explains the peak in the native Patterson function, which was the initial signal for us to suspect twinning. Rotation of this model by 60° about the Z axis resulted in an identical Patterson correlation coefficient, confirming the underlying merohedral twinning.

3.3.2 Successful Detwinning and Improved Structure

We have implemented the detwinning procedure using X-PLOR command scripts (Brünger, 1992). In order to be certain that the procedure was working properly, the p21 peptide was not included in the molecular replacement searches and the initial detwinning or refinement calculations. After three iterations of detwinning and refinement at 2.6Å resolution, averaged difference electron density maps revealed strong and tracable density for the p21 peptide (Figure 3.4), with sidechains and the 3_{10} helix clearly visible. The resolution of the refinement and detwinning was gradually increased to 2.3 Å. At this stage the map began to reveal features that were not present in the lower resolution starting structure. Several residues that were not modeled in different parts of the structure now became visible. For example, clear density for the previously disordered loop connecting βD2 and βE2 (residue 186-191) is present in the detwinned electron density map (Figure 3.5), but was apparently missing in the previous structure. At this stage of refinement at 2.3Å, R_{tw} is 0.21 and $R_{tw(free)}$ is 0.24 (the presence of 3-fold non-crystallographic symmetry artificially lowers the free R-value). This corresponds to conventional R-values of 0.29 and 0.34, respectively (Redinbo and Yeates, 1993). The refinement can be extended to 2.0Å resolution, and the model will be completed with the inclusion of solvent molecules. As shown in Figure 3.5, there is promising electron density for other parts in the structure which were apparently disordered at 2.6Å resolution, especially at the N-terminal end (four residues) of the p21 peptide interacting with the C-terminus of PCNA (six residues) which were poorly ordered in the 2.6 Å

Figure 3.4 Electron Density for p21 Peptide in p21-PCNA complex. Electron density from a 2.6 Å simulated annealing 2|Fo|-|Fc| OMIT map is shown for the region of p21 peptide. The simulated annealing is carried out against the detwinned fobs for only one orientation with three molecules in unit cell. Electron density is contoured at 1.0 σ. The PCNA atoms in vicinity to p21 peptide are shown in ball and stick representation. p21 peptide is illustrated as thin trace through the electron density.

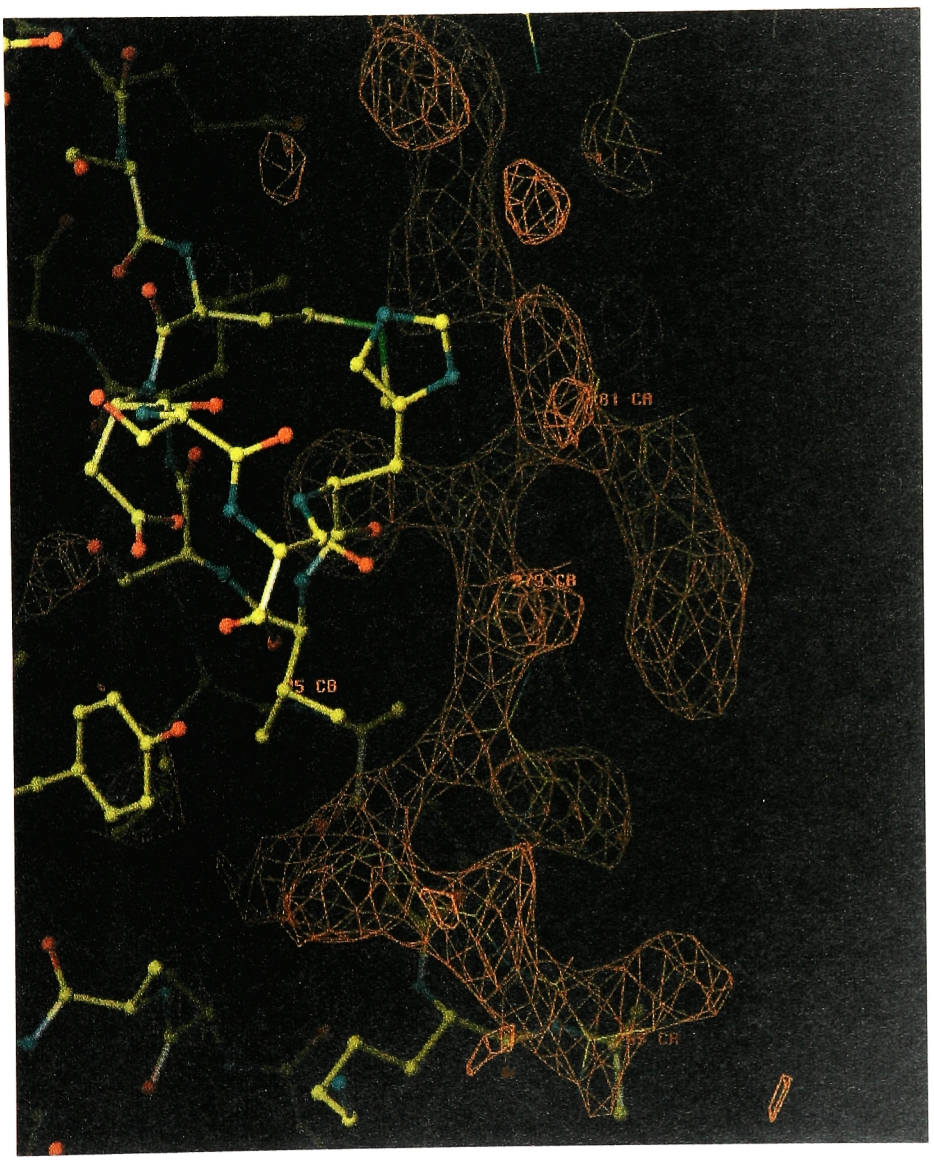
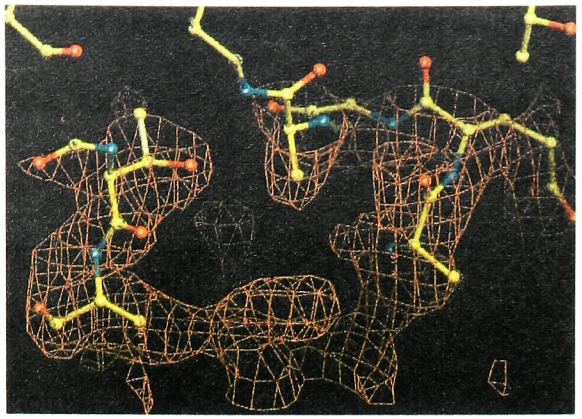
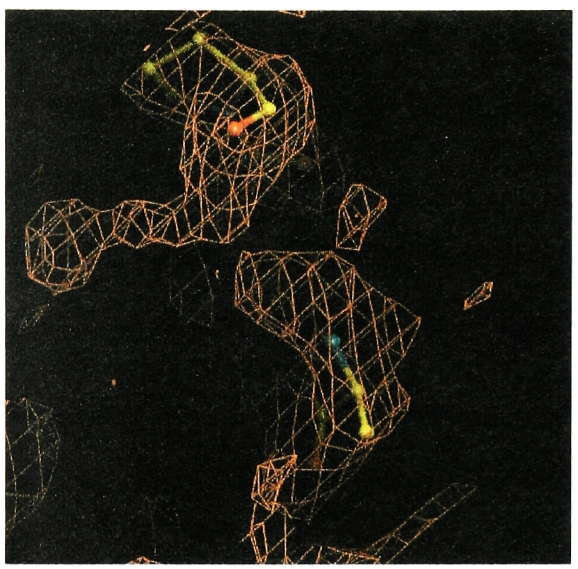

81 CA
5 CB

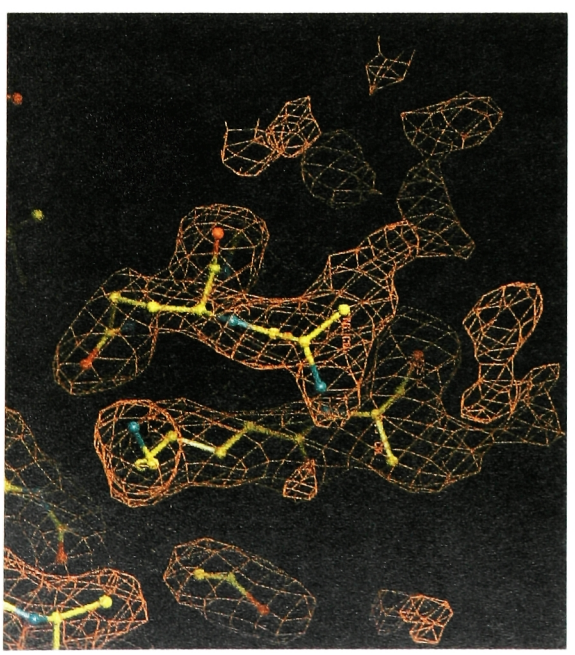
Figure 3.5 Electron Density for Previously Disordered Region in p21-PCNA Complex. Electron density from a difference |Fo|-|Fc| map is shown for regions disordered in the starting molecular replacement structure solved from the untwined data. (A) Electron density for missing segment between residues 186 and 191 in PCNA molecule. (B) Electron density for missing segment between residues 201 and 203 in PCNA. (C) Electron density unaccounted for in the model at the N-terminal of p21 peptide. Protein atoms in the region of illustration are shown in bond-and-stick presentation.



A



B



C

structure. Further refinement to reduce the R value will definitely help to fill in the details in this important interaction region. This extension in resolution is particularly important because of the highly polar nature of the interaction between p21 peptide and PCNA. The precise details of hydrogen bonding and water mediated interactions will be interesting and useful for understanding the elements of specificity and affinity at this site.

3.3.3 Conclusion

The 2.0 Å data set of the human PCNA/p21 peptide complex has perfect hemihedral twinning in the P3 space group, which resulted in an apparent 6/m symmetry in diffraction data. A detwinning algorithm has been implemented in XPLOR, using the human p21/PCNA structure determined at 2.6 Å from another crystal form as a starting model. Detwinning proved to be successful and the structure of the complex is improved in terms of resolution and the presence of interpretable density for several previously disordered loops in the structure. The significance of this work is that it will provide a more accurate picture of the highly polar interface between PCNA and p21 peptide, which will be helpful in designing small molecules to mimic the interaction. This project is also a very valuable training in crystallography to understand space group complexity and structure refinement.

References

Abrahams, J. P., and Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial F_1 ATPase. Acta Cryst *D52*, 30-42.

Beese, L. S., and Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. EMBO J *10*, 25-33.

Berdis, A. J., Soumillion, P. and Benkovic, S.J. (1996). The carboxyl terminus of the bacteriophage T4 DNA polymerase is required for holoenzyme complex formation. Proc Natl Acad Sci U S A *93*, 12822-12827.

Bottger, A., Bottger, V., Garcia-Echeverria, C., Chene, P., Hochkeppel, H., Sampson, W., Ang, K., Howard, S., Picksley, S., and Lane, D. (1997). Molecular characterization of the HDM2-P53 interaction. J Mol Biol *269*, 744-756.

Braithwaite, D. K., and Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nuc Acids Res *21*, 787-802.

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

Brautigam, C. A., and Steitz, T. A. (1998b). Structural principles for the inhibition of the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates. J Mol Biol *277*, 363-377.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., Swaminathan, S., and Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem *4*, 187-217

Bruck, I., and O'Donnell, M. (2000). The DNA replication machine of a gram-positive organism. J Biol Chem *275*, 28971-28983.

Brünger, A. T. (1992). *X-PLOR (Version 3.1) A System for X-Ray Crystallography and NMR*, (New Haven, Connecticut: Yale University)).

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

CCP4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D *52*, 760-763.

Chen, J., Jackson, P. J., Kirschner, M. W., and Dutta, A. (1995). Separate domains of p21 involved in the inhibition of Cdk kinase and PCNA. Nature *374*, 386-388.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc *117*, 5179-5197.

Darden, T., York, D., and Pedersen, L. (1995). Particle mesh ewald:An n log(n) method for ewald sums in large systems. J Chem Phys *98*, 10089-10092.

De-La-Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. In Macromolecular Crystallography, R. M. Sweet, and C. W. Carter, eds. (Academic Press, New York), pp. 472-493.

Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P (1990). An attempt to unify the structure of polymerases. Protein Eng *3*, 461-467.

Doublie, S., Sawaya, M. R., and Ellenberger, T. (1999). An open and closed case for all polymerases. Structure Fold Des *7*, R31-R35.

Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature *391*, 251-258.

Ellison, V., and Stillman, B. (2001). Opening of the clamp: an intimate view of an ATP-driven biological machine. Cell *106*, 655-660.

Eom, S. H., Wang, J., and Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. Nature *382*, 278-281.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. (1995). A smooth particle mesh Ewald method. J Chem Phys *103*, 8577-8593.

Feller, S. E., Zhang, Y. H., Pastor, R. W., and Brooks, C. L., 3rd (1995). Constant pressure molecular dynamics simulation - the Langevin piston method. J Chem Phys *103*, 4613-4621.

Franklin, M. C., Wang, J., and Steitz, T. A. (2001). Structure of the replicating complex of a pol alpha family DNA polymerase. Cell *105*, 657-667.

Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from Thermus thermophilus complexed with tRNAPhe. Structure *15*, 59-68.

Gomes, X. V., and Burgers, P. M. (2001). ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. Journal of Biological Chemistry *276*, 34768-34775.

Gomes, X. V., Schmidt, S. L., and Burgers, P. M. (2001). ATP utilization by yeast replication factor C. II. Multiple stepwise ATP binding events are required to load proliferating cell nuclear antigen onto primed DNA. Journal of Biological Chemistry *276*, 34776-34783.

Goodrich, L. D., Lin, T.C., Spicer, E.K., Jones, C. and Konigsberg, W.H. (1997). Residues at the carboxy terminus of T4 DNA polymerase are important determinants for interaction with the polymerase accessory proteins. Biochemistry *36*, 10474-10481.

Griffith, J. D., Lindsey-Boltz, L. A., and Sancar, A. (2002). Structures of the human rad17-replication factor C and checkpoint rad 9-1-1 complexes visualized by glycerol spray/low voltage microscopy. J Biol Chem *277*, 15233-15236.

Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M., and Kuriyan, J. (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. Cell *87*, 297-306.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K., and Jansonius, J. N. (1995). 2.0 A structure of indole-3-glycerol phosphate synthase from the hyperthermophile Sulfolobus solfataricus: possible determinants of protein stability. Structure *3*, 1295-1306.

Herendeen, D. R., Kassavetis, G. A., and Geiduschek, E. P (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. Science *256*, 1298-1303.

Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. Proteins *33*, 417-429.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol *233*, 123-138.

Huang, C. C., Hearst, J. E., and Alberts, B. M. (1981). Two types of replication proteins increase the rate at which T4 DNA polymerase traverses the helical regions in a single-stranded DNA template. J Biol Chem *256*, 4087-4094.

Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669-1675.

Im, W., Beglov, D., and Roux, B. (1998). Continuum solvation model: electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Comput Phys Comm *111*, 59-75.

Jacobo-Molina, A., Ding, J., Nanni, R., Clark, A. J., X., L., Tantillo, C., Williams, R L., Kamer, G., Ferris, A. L., Clark, P., and al., e. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc Natl Acad Sci USA *90*.

Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2001a). Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. Cell *106*, 429-441.

Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2002). Clamp Loaders and Sliding Clamps. Curr Opin Struct Biol *12*, 217-224.

Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M., and Kuriyan, J. (2001b). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of E. coli DNA polymerase III. Cell *106*, 417-428.

Jeruzalmi, D. a. S., T.A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. Embo J *17*.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr *A47*, 110-119.

Jorgensen, W. L. (1981). Transferable intermolecular potential functions for water, alchohols and ethers. Application to liquid water. J Am Chem Soc *103*, 335-340.

Joyce, C. M., and Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. Annu Rev Biochem *63*, 777-822.

Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. Nature *391*, 304-307.

Kiefer, J. R., Mao, C., Hansen, C.J., Basehore, S.L., Hogrefe, H.H., Braman, J.C. and Beese, L.S. (1997). Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 A resolution. Structure *5*, 95-108.

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., and T.A., S. (1995). Crystal structure of Thermus aquaticus DNA polymerase. Nature *376*, 612-616.

Kleywegt, G. J., and Jones, T. A. (1994). In Proceedings of the CCP4 Study Weekend (Warrington, U.K., SERC Daresbury Laboratory), pp. 59-66.

Koch, E. (1992). In International Tables for Crystallography, A. J. C. Wilson, ed. (Dordrecht: Kluwer Academic Publishers), pp. 10-14.

Kollman, P. (1993). Free Energy Calculations Applications to Chemical And Biochemical Phenomena. Chem Rev *93*, 2395-2417.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res *33*, 889-897.

Kong, H., Kucera, R. B., and Jack, W. E. (1993). Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. J Biol Chem *268*, 1965-1975.

Kong, X. P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992). Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. Cell *69*, 425-437

Kornberg, A., and Baker, T A. (1991). DNA Replication, 2 edn (New York, W.H. Freeman).

Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E., and Waksman, G. (1995). Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5-A resolution: structural basis for thermostability. Proc Natl Acad Sci U S A *92*, 9264-9268.

Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell *79*, 1233-1243.

Kunkel, T. A. (1992). DNA replication Fidelity. J Biol Chem *267*, 18251-18254.

Lasken, R. S., Schuster, D. M., and Rashtchian, A. (1996). Archaebacterial DNA polymerases tightly bind uracil-containing DNA. J Biol Chem *271*, 17692-17696.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr *26*, 283-291.

Lee, M. H., Reynisdottir, I., and Massague, J. (1995). Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution. Genes & Devel *9*, 639-649.

Li, R., Waga, S., Hannon, G. J., Beach, D., and Stillman, B. (1994). Differential effects by the p21 CDK inhibitor on PCNA-dependent DNA replication and repair. Nature *371*, 534-537.

Li Y, Korolev S, and G., W. (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*, 7514-7525.

Lindahl, M., Svensson, L. A., Liljas, A., Sedelnikova, S. E., Eliseikina, I. A., Fomenkova, N. P., Nevskaya, N., Nikonov, S. V., Garber, M. B., and Muranova, T. A. (1994). Crystal structure of the ribosomal protein S6 from Thermus thermophilus. Embo J *13*, 1249-1254.

Lopez de Saro, F. J., and O'Donnell, M. (2001). Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I. Proceedings of the National Academy of Sciences of the United States of America *98*, 8376-8380.

Luo, Y., Hurwitz, J., and Massagué, J. (1995). Cell-cycle inhibition by independent CDK and PCNA binding domains in $p21^{Cip1}$. Nature *375*, 159-161.

MacKerell, A. D. J., D. Bashford, M. Bellott, R. L. Dunbrack, J., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B *102*, 3586-3616.

Matsumiya, S., Ishino, Y., and Morikawa, K. (2001). Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from Pyrococcus furiosus. Protein Sci *10*, 17-23.

McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. Nature *267*, 585-590.

Miyachi, K., Fritzler, M. J., and Tan, E. M. (1978). Autoantibody to a nuclear antigen in proliferating cells. Journal of Immunology *121*, 2228-2234.

Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M., and Kuriyan, J. (2000). Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. J Mol Biol *296*, 1215-1223.

Nina, M., Beglov, D., and Roux, B. (1997). Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. J Phys Chem B *101*, 5239-5248.

O'Donnell, M., Jeruzalmi, D., and Kuriyan, J. (2001). Clamp loader structure predicts the architecture of DNA polymerase III holoenzyme and RFC. Curr Biol *11*, R935-946.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985). Structure of large fragment of Escherichia coli polymerase I complexed with dTMP. Nature *313*, 762-766.

Onrust, R., Stukenberg, P. T., and O'Donnell, M. (1991). Analysis of the ATPase Subassembly Which Initiates Processive DNA Synthesis by DNA Polymerase. J Biol Chem *266*, 21681-21686.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Meth Enzymol *276*, 307-326.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature *372*, 432-438.

Oyama, T., Ishino, Y., Cann, I., Ishino, S., and Morikawa, K. (2001). Atomic structure of the clamp loader small subunit from Pyrococcus furiosus. Mol Cell *8*, 455-463.

Pavlov, A. R., and Karam, J. D. (1994). Binding specificity of T4 DNA polymerase to RNA. J Biol Chem *269*, 12968-12972.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., Fergusen, D. M., Seibel, G. L., Singh, U. C., Weiner, P., and Kollman, P. (1995). AMBER 4.1 (San Francisco, CA, UCSF).

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase E, a DNA template-primer and ddCTP. Science *264*, 1891-1903.

Peter, M., and Herskowitz, I. (1994). Joining the complex: cyclin-dependent inhibitory proteins and the cell cycle. Cell *79*, 181-184.

Podust, V. N., Podust, L. M., Goubin, F., Ducommun, B., and Hübscher, U. (1995). Mechanism of inhibition of proliferating cell nuclear antigen-dependent DNA synthesis by the cyclin-dependent kinase inhibitor p21. Biochemistry *34*, 8869-8875.

Polyak, K., Lee, M.-H., Erdjument-Bromage, H., Koff, A., Roberts, J. M., Tempst, P., and Massagué, J. (1994). Cloning of $p27^{Kip1}$, a cyclin-dependent kinase inhibitor and a potential mediator or extracellular antimitogenic signals. Cell *78*, 59-66.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Cryst *A42*, 140-149.

Redinbo, M., and Yeates, T. (1993). Structure determination of Plastocyanin from a speciman with a hemihedral twinning faction of one-half. Acta Crystallogr D *49*, 375-380.

Roux, B. (1996). Valence selectivity of the gramicidin channel: a molecular dynamics free energy pertubation study. Biophys J *71*, 3177-3185.

Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the Cartesian equation of motions of a system with constraints: molecular dynamics of n-alkanes. J Comp Chem *23*, 327-341.

Schmidt, S. L., Gomes, X. V., and Burgers, P. M. (2001a). ATP utilization by yeast replication factor C. III. The ATP-binding domains of Rfc2, Rfc3, and Rfc4 are essential for DNA recognition and clamp loading. Journal of Biological Chemistry *276*, 34784-34791.

Schmidt, S. L., Pautz, A. L., and Burgers, P. M. (2001b). ATP utilization by yeast replication factor C. IV. RFC ATP-binding mutants show defects in DNA replication, DNA repair, and checkpoint regulation. Journal of Biological Chemistry *276*, 34792-34800.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R., and Steitz, T A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nat Struct Biol *4*, 215-222.

Shamoo, Y., and Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell *99*, 155-166.

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

Srinivasan, J., Cheatham III, T. E., Cieplak, P., and Kollman, P. (1998). Continnum solven studies of the stability of DNA, RNA and Phosphoramidate-DNA helices. J Am Chem Soc *120*, 9401-9409.

Stanley, E. (1972). The identification of twins from intensity statistics. J Appl Cryst *5*, 191-194.

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

Steitz, T. A., Smerdon, S. J., Jäger, J., and Joyce, C. M. (1994). A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. Science *266*, 2022-2025.

Stewart, J., Hingorani, M. M., Kelman, Z., and O'Donnell, M. (2001). Mechanism of beta Clamp Opening by the delta Subunit of Escherichia coli DNA Polymerase III Holoenzyme. J Biol Chem *276*, 19182-19189.

Stukenberg, P. T., Studwell-Vaughan, P. S., and O'Donnell, M. (1991). Mechanism of the Sliding E clamp of DNA Polymerase III Holoenzyme. J Biol Chem *266*, 11328-11334.

Su, J.-Y., Rempel, R. E., Erikson, E., and Maller, J. L. (1995). Cloning and characterization of the Xenopus cyclin-dependent kinase inhibitor $p27^{XIC1}$. Proc Natl Acad Sci (USA) *92*, 10187-10191.

Sugino, A., Hirose, S., and Okazaki, R. (1972). RNA-linked nascent DNA fragments in Escherichia coli. Proc Natl Acad Sci U S A *69*, 1863-1867.

Sutton, M. D., Smith, B. T., Godoy, V. G., and Walker, G. C. (2000). The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. Annu Rev Genet *34*, 479-497

Tama, F., Wriggers, W., and Brooks, C. L., 3rd (2002). Exploring Global Distortion of Biological Macromolecules and Assemblies from Low-Resolution Structural Information and Elastic Network Theory. J Mol Biol *321*, 297-305.

Tuerk, C., Eddy, S., Parma, D., and Gold, L. (1990). Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. J Mol Biol *213*, 749-761.

Turner, J., Hingorani, M. M., Kelman, Z., and O'Donnell, M. (1999). The internal workings of a DNA polymerase clamp-loading machine. EMBO J *18*, 771-783.

Uemori, T., Ishino, Y., Toh, H., Asada, K., and Kato, I. (1993). Organization and nucleotide sequence of the DNA polymerase gene from the archeon *Pyrococcus furiosus*. Nucleic Acids Res *21*, 259-265.

van Gunsteren, W. F., and Berendsen, H. J. (1984). Computer simulations as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. Journal of Molecular Biology *176*, 559-564.

Varani, G., and Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. Annu Rev Biophys Biomol Struct *27*, 407-445.

Waldman, T., Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1996). Uncoupling of S phase and mitosis by anticancer agents in cells lacking p21. Nature *381*, 713-716.

Wang, C. C., Pavlov, A., and Karam, J. D. (1997a). Evolution of RNA-binding specificity in T4 DNA polymerase. J Biol Chem *272*, 17703-17710.

Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, J. D., and Steitz, T. A. (1997b). Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. Cell *89*, 1087-1099.

Warbrick, E., Lane, D. P., Glover, D. M., and Cox, L. S. (1995). A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21waf1 and the proliferating cell nuclear antigen. Curr Biol *5*, 275-282.

Watson, J. D., and Crick, F. H. C. (1953a). Genetic Implications of the Structure of Deoxyribonucleic Acid. Nature *171*, 964-967.

Watson, J. D., and Crick, F. H. C. (1953b). Molecular Structure of Nucleic Acid. A Structure for Deoxyribonucleic Acid. Nature *171*, 737-738.

Wesson, L., and Eisenberg, D. (1992). Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci *1*, 227-235.

Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997). Xu, R.M., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A.R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure *5*, 559-570.

Yao, N., Turner, J., Kelman, Z., Stukenberg, P. T., Dean, F., Shechter, D., Pan, Z. Q., Hurwitz, J., and O'Donnell, M. (1996). Clamp loading, unloading and intrinsic stability of the PCNA, beta and gp45 sliding clamps of human, E. coli and T4 replicases. Genes Cells *1*, 101-113.

Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., and Consalvi, V. (1995). The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure *3*, 1147-1158.