Cell, Vol. 91, 335–345, October 31, 1997, Copyright ©1997 by Cell Press

Crystal Structure of the δ′ Subunit of the Clamp-Loader Complex of E. coli DNA Polymerase III

Brian Guenther,[†‖] Rene Onrust,[‡] Andrej Sali,[†] Mike O'Donnell,[*†‡] and John Kuriyan[*†§]
*Howard Hughes Medical Institute
[†]The Rockefeller University
New York, New York 10021
[‡]Cornell University Medical College
New York, New York 10021

Summary

The crystal structure of the δ′ subunit of the clamp-loader complex of E. coli DNA polymerase III has been determined. Three consecutive domains in the structure are arranged in a C-shaped architecture. The N-terminal domain contains a nonfunctional nucleotide binding site. The catalytic component of the clamp-loader complex is the γ subunit, which is homologous to δ′. A sequence-structure alignment suggests that nucleotides bind to γ at an interdomain interface within the inner surface of the "C." The alignment is extended to other clamp-loader complexes and to the RuvB family of DNA helicases, and suggests that each of these is assembled from C-shaped components that can open and close the jaws of the "C" in response to ATP binding and hydrolysis.

Introduction

The chromosomal replicases of eukaryotes, prokaryotes, and bacteriophage T4 are distinguished functionally from other DNA polymerases by their very high processivity (Nossal and Alberts, 1984; Kornberg and Baker, 1991; McHenry, 1991; Kelman and O'Donnell, 1995). These replicases are multisubunit holoenzymes that achieve a tight coupling to DNA by utilization of ring-shaped sliding DNA clamps (Huang et al., 1981). The central catalytic component of the holoenzymes is the DNA polymerase/exonuclease, which attaches tightly to DNA by binding to the noncatalytic processivity factor, the DNA-encircling ring. Additionally, since a closed ring cannot load onto DNA efficiently, a third component acts as a clamp loader that utilizes the binding and hydrolysis of ATP to load the processivity rings onto DNA (reviewed in Kelman and O'Donnell, 1995).

The primary chromosomal replicase of E. coli is the DNA polymerase III holoenzyme (pol III), which utilizes as its processivity factor the dimeric ring formed by the β subunit (reviewed in McHenry, 1991; Kelman and O'Donnell, 1995). The β ring is loaded onto DNA by the clamp-loader complex (γ complex) of pol III, which consists of five subunits (γ, δ, δ′, χ, and ψ; containing 2–4 molecules of γ, with one molecule each of the other subunits) (McHenry, 1991; Onrust et al., 1991). In eukaryotes, the processivity factor PCNA (proliferating cell nuclear antigen) is a trimeric ring (Krishna et al., 1994) that is loaded onto DNA by the five subunit replication factor C (RFC) or activator-1 complex (Lee et al., 1991). Likewise, the processivity factor of bacteriophage T4, the gene 45 protein, is also a trimeric ring, and is loaded onto DNA by a complex formed by the proteins encoded by genes 62 and 44 (Jarvis et al., 1989).

The bacterial, eukaryotic, and T4 bacteriophage processivity factors are all structurally similar, even though they do not share significant sequence similarity (Kong et al., 1992; Krishna et al., 1994; I. Moarefi and J. K., unpublished data). Each of the proteins forms a stable closed ring in solution, in the absence of DNA. The mechanism of action of clamp-loader complexes is therefore expected to be similar in each case, and is likely to involve a mechanical opening of the ring-shaped processivity factors since there is no evidence for DNA cleavage during loading of the clamps. A common mechanism is also suggested by the fact that subunits of the clamp-loader complexes from different organisms share some sequence similarity (Carter et al., 1993; Dong et al., 1993; O'Donnell et al., 1993).

The clamp-loader complex of E. coli binds to the β ring and to primed template (single-strand/double-strand junctions or nicked duplex DNA) and then places the β ring upon DNA in an ATP-dependent manner. The key players in the clamp-loading mechanism are the γ (47 kDa), δ (38.7 kDa), and δ′ (36.9 kDa) subunits, with the χ (16.6 kDa) and ψ (15.2 kDa) subunits contributing to the stability of the complex (Kelman and O'Donnell, 1995). In E. coli, translational frameshifting of the DnaX gene produces two protein products, γ and τ, with the sequence of γ being contained entirely within τ (Flower and McHenry, 1986, 1990; Yin et al., 1986; Tsuchihashi and Kornberg, 1990). The longer τ subunit (71 kDa) is also functional in clamp-loading, but has additional functions in the holoenzyme. We refer to both γ and τ in this paper as "γ" since both proteins completely encompass the three domains of δ′. The γ and δ′ subunits are unrelated in sequence to the δ, χ, and ψ subunits.

The γ subunit is the only subunit of the clamp-loader complex that is capable of binding and hydrolyzing ATP. The particular roles of the other subunits are not yet known with certainty, but the following model is consistent with the available biochemical evidence (Naktinis et al., 1996). The binding of the δ subunit to β induces an opening of the β ring. The interaction of δ with β is normally prevented by interactions of δ with the δ′ subunit. The binding or hydrolysis of ATP by the γ subunit results in a conformational change that allows the δ subunit to bind to the β subunit. The competition of δ′ for the δ subunit may also be involved in the release of the γ complex from the β subunit after the clamp is placed on DNA (Naktinis et al., 1996).

A detailed understanding of the mechanism of clamp loading requires that the three-dimensional structures of the subunits of the clamp-loader complexes be known, ideally in an active assembly. So far, no structural information is available on any of these subunits. In this

[§]To whom correspondence should be addressed.
[‖]Present Address: Biophysics Research Division, University of Michigan, 930 N. University Avenue, Ann Arbor, Michigan 48109-1055.

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

	Resolution (Å)	Number of Reflections (total/unique)	Completeness (%)	R_{sym}[a] (%)	R_{deriv}[b] (%)	Sites[c]	Power Phasing[d]	Figure of Merit[e]
Native Data								
Native1	2.2	78348/18034	89.6 (78.7)	6.3 (19.9)				
Native2	2.7	23711/9686	83.4 (63.9)	5.8 (22.1)				
MIR Analysis								0.50
Pt	2.8	52429/9596	97.8 (96.1)	7.8 (20.1)	13.2	3	1.02	
U	3.2	30492/6319	94.3 (85.4)	8.9 (17.1)	12.8	2	0.88	
W	3.0	31400/7002	87.0 (67.3)	5.7 (17.7)	15.7	2	0.63	
Os	3.5	9295/4375	84.6 (66.9)	7.0 (16.8)	15.4	2	0.67	
Ir	3.5	10415/4372	87.4 (72.4)	6.3 (12.2)	10.5	2	0.75	
Refinement				R Value	R Free			
All	30.0–2.2	18759	93.1 (88.0)	21.7	27.7			
$\|F\| > 2\sigma(\|F\|)$	30.0–2.2	17778	88.2 (74.4)	20.5	26.5			

[a] $R_{sym} = 100 \times \Sigma_h \Sigma_i |I_{h,i} - \langle I_h \rangle| / \Sigma_h \Sigma_i I_{h,i}$

[b] $R_{deriv} = \Sigma_h |F_{nat,h} - F_{deriv,h}| / \Sigma_h F_{nat,h}$

[c] Free R factor was calculated with 10% of the data.

[d] Phasing Power is rms ($|F_h| / E$), where the subscript "h" represents "heavy-atom" and "E" is the residual lack of closure.

[e] FOM (mean figure of merit) $= \langle |\Sigma P(\alpha) e^{i\alpha} / \Sigma P(\alpha)| \rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution. Numbers in parentheses refer to statistics for the outer shell of data.

paper, we present the three-dimensional structure of the δ′ subunit of E. coli pol III. The structure of this isolated subunit of the complex cannot, by itself, explain the mechanism of clamp loading. However, analysis of the δ′ structure, particularly in light of sequence-structure comparisons with other clamp-loader proteins, provides the first insights into the architecture and possible mechanisms of action of the basic components of these molecular machines.

The δ′ subunit is a C-shaped molecule comprised of three domains, with a RecA-like mononucleotide binding fold for the N-terminal domain. The δ′ subunit is not, however, a nucleotide triphosphate hydrolase since the nucleotide binding and catalytic sites in δ′ are nonfunctional. The catalytic component of the clamp-loader complex is the homologous γ subunit, which is a functional DNA-dependent ATPase. Sequence-structure comparisons indicate that the structure of the γ subunit will resemble that of the δ′ subunit, with amino acid substitutions in γ that are consistent with nucleotide binding. The alignment also indicates that the ATP binding site of γ is located in the central region of the C-shaped molecule, and that the binding or hydrolysis of ATP could be coupled to changes in the relative orientations of the three domains that make up the C. The structure-based alignment further suggests that the gene 44 protein of bacteriophage T4 and the subunits of the RFC complex are also likely to have the same general architecture, particularly for the first two domains. Finally, there is suggestive evidence that the RuvB family of hexameric DNA helicases have a fold related to that of δ′.

Results and Discussion

General Architecture

The crystal structure of the δ′ subunit of E. coli pol III has been solved and refined at 2.2 Å resolution (R value = 20.5%, R_{free} = 26.5%; Table 1). The structure consists of three sequential domains, with each domain defined by a discrete hydrophobic core (Figure 1). The crystallographic model contains 324 of the 334 amino acids of δ′. Four residues at the C terminus as well as six residues in a loop (residues 259–264) are not well-ordered and are not included in the model.

An insightful discussion of the δ′ structure results if two conceptual connections are made. First, we relate the sequence and structure of δ′, which does not bind or hydrolyze nucleotides, to that of the γ subunit of E. coli pol III, which is a functional ATPase. Second, the structure of δ′ and the implied relationship with γ is used to generate a structure-based alignment of the sequences of the eukaryotic clamp-loader subunits (RFC subunits 1–5) and the T4 bacteriophage gene 44 protein. A possible connection to the RuvB family of helicases was discovered and is included in the analysis. For clarity, discussion and evaluation of the sequence alignment is presented later (see "Sequence-Structure Alignments"), but the results of the alignment (Figure 3) are used throughout the following discussion of the δ′ structure.

The amino-terminal domain of δ′ (Domain 1, Met-1 to Pro-168) has at its center a β sheet containing five parallel strands that are surrounded by six helices (Figure 1). The structure of this domain resembles that of the core of the nucleotide-binding domain of RecA (Story et al., 1992). The five strands of the β sheet in Domain 1 of δ′ as well as helix α2 (colored red in Figures 1 and 4) have the same topological connectivity as in RecA (rms deviation of 1.1 Å over ~55 C^{α} atoms). The three α helices on one side of the β sheet (α4–α6; see Figure 4a) are also structurally similar between δ′ and RecA, but are displaced by 2–3 Å. On the other side of the β sheet, two helices in δ′ (α1 and α3) do not have corresponding elements in RecA. Domains 2 and 3 of δ′ are unrelated to any segments of the RecA structure.

Domain 1 of δ′ contains a zinc binding module, also found in γ, that is an insertion into an external loop between α2 and β2 (Figure 1). Four cysteine ligands for the zinc define a motif $(C\{X\}_8CXXCXXC$; Figure 2), that

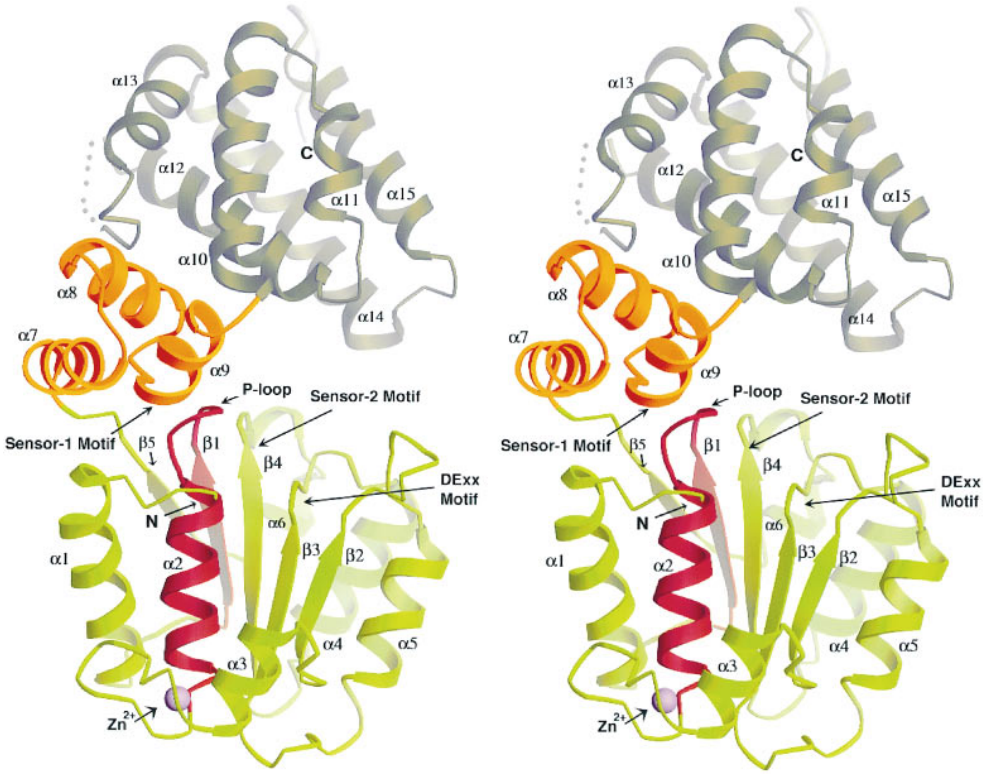

Figure 1. Ribbon Diagram of δ′

Domains 1, 2, and 3 are colored green, orange, and grey-green, respectively. The P-loop segment is colored red.

is conserved in bacterial sequences of δ′ and γ. This pattern of zinc coordinating ligands has not been observed previously in zinc modules of known structure (Klug and Schwabe, 1995). The presence of zinc in the bacterial clamp loaders was first discovered during the crystallographic analysis of δ′ and has subsequently been confirmed for E. coli δ′ and γ by atomic absorption spectroscopy (M. O'D., unpublished data). The zinc module is not present in the eukaryotic clamp loaders, nor in T4 gene 44 protein or RuvB (Figure 2).

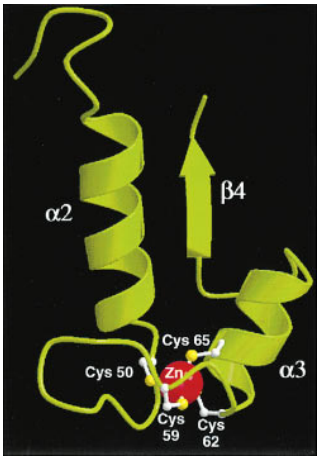

Figure 2. The Zinc Binding Module

The zinc atom is colored red and the cysteine ligands are shown in stick representation.

The function of this zinc module is not known, but a role in interactions with DNA is possible. In E. coli δ′, the zinc module forms a surface groove that is in a region of positive electrostatic potential (Figure 6). The γ subunit is a DNA-dependent ATPase, and it is intriguing that the zinc module is located at the base of helix α2, which emanates directly out of the phosphate binding P loop (Figure 1). Thus, interaction of DNA with the zinc module might provide a mechanism for coupling DNA binding to ATP hydrolysis.

The second domain (Domain 2, Glu-169 to Gln-206) is a small three helix bundle. The third domain (Domain 3, Gly-207 to Leu-334) contains six helices, with the up-down connectivity that is characteristic of helix bundles. The overall dimensions of the δ′ molecule are 75 Å by 40 Å by 30 Å, with Domains 1 and 3 separated by Domain 2, resulting in a C-shaped molecule.

The region of the γ subunit that can be aligned with δ′ (Figure 3) corresponds to the N-terminal 80% of γ. A third of the residues in the unique C-terminal segment of γ are either proline (19%) or glutamine (14%), whereas the average amino acid composition of globular proteins includes approximately 4.8% ± 2.4% and 3.8% ± 2.0% each of proline and glutamine, respectively (A. S., unpublished data). Hydrophobic residues are sparsely distributed in the C-terminal region of γ, with isoleucine, tyrosine, and tryptophan residues being completely absent. These factors suggest that the C-terminal region may not be folded into a globular domain. It appears therefore that the structure of γ consists of an N-terminal segment that is folded similarly to Domains 1, 2, and 3

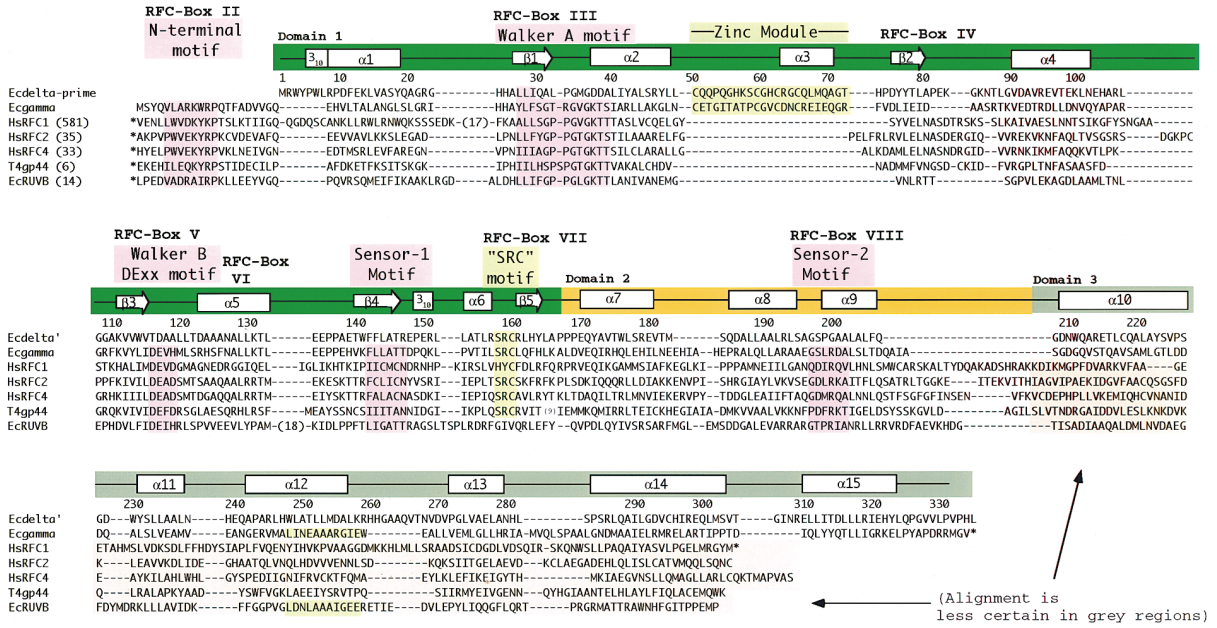

Figure 3. Sequence-Structure Alignment

The sequences of E. coli δ′ and γ; human RFC subunits 1, 2, and 4; bacteriophage T4 gene 44 protein; and E. coli RuvB are shown. The sequences shown in the figure are a subset of the >30 sequences utilized in generating the structure-based alignment described in the text. Numbers in parentheses after the names of the sequences refer to the number of residues preceding the first residue shown in the alignment. Numbers in parentheses within the alignment refer to the number of inserted residues that are not shown in the alignment. Sequence motifs identified in the text, as well as motifs that have been identified previously in the RFC proteins and referred to as RFC boxes, are highlighted in pink (Cullman et al., 1995). The zinc module and the region of local sequence similarity between γ and RuvB in helix α12 are highlighted in green.

of δ′ (21% sequence identity, Table 2), and an ~80 residue C-terminal segment of unknown function that is perhaps flexible or disordered in structure.

Walker A Motif: Phosphate Binding P Loop

Many nucleotide binding domains with α/β folds contain a conserved phosphate binding segment ("P loop") that interacts with the phosphate tail of the nucleotide and contains a sequence motif known as "Walker A" (Gxxxx GKT, with some variation) (Walker et al., 1982; Schulz, 1992; Cronet et al., 1995). In each of these proteins, the structural segment that is conserved comprises a β strand, the P loop, and an α helix (Figure 4). While the γ subunit of pol III has a canonical Walker-A motif (GTRGVGKT), this motif is absent in the δ′ sequence (the corresponding sequence segment is ALPGMGDD; see Figure 5). Nevertheless, the δ′ structure contains a strand-loop-helix (β1-α2) segment that closely resembles the corresponding segments of nucleotide binding proteins that contain P loops (colored red in Figures 1 and 4).

Nucleotide triphosphate hydrolases (NTPases) that contain P loops bind the phosphate groups of ATP or GTP in a similar manner (Schulz, 1992) (Figure 4). Superimposing the structure of the bound form of ATP, taken from adenylate kinase, onto the E. coli δ′ structure reveals two reasons for the inability of δ′ to bind nucleotides (Figure 5). First, the presence of Asp-37 and Asp-38 in the Walker-A motif instead of the canonical lysine and threonine residues is likely to result in unfavorable electrostatic and van der Waals interactions with the three phosphate groups of ATP (Figure 5). An electrostatic potential map of the E. coli δ′ surface shows that the P loop is in a region of negative electrostatic potential, as a consequence of Asp-37 and Asp-38 (Figure 6). Although acidic residues that coordinate metal ions are commonly located near phosphate binding sites, Asp-37 and Asp-38 are not sufficiently distant from the inferred phosphate binding sites for metal-mediated phosphate ligation. Secondly, the binding of ATP is further prevented by the structure of the N-terminal region of δ′. The side chain of Tyr-4 occupies the binding site for the purine ring in nucleotide binding domains, and is held in place by packing interactions with Arg-2 and Trp-6 (Figure 4b). These side chains are in turn held in place by ion pairing and packing interactions, effectively occluding the ATP binding site.

The sequence-structure alignment shows that Domain 1 of the γ subunit of pol III shares 29% sequence identity with δ′, and is consequently expected to adopt the same three-dimensional fold (Figure 3, Table 2). The canonical Walker-A motif of the γ subunit is expected to be part of a functional phosphate binding site (the two aspartate residues that would impede phosphate binding to δ′ are replaced by lysine and threonine in γ). The N-terminal region of Domain 1 in γ has a sequence that is quite distinct from that observed in δ′ (denoted "N-terminal motif/RFC Box II" in Figure 3). Specifically, the residues that are seen to interfere with the adenine binding site are unique to δ′. Intriguingly, the N-terminal motif seen in γ is conserved across the eukaryotic clamp-loader subunits, the T4 gp44 protein, and RuvB (Figure 3). The structure of δ′ suggests that this region

Table 2. Sequence Identities and PROSA Z Scores

Protein	Sequence Identity (%, Full length/Domain 1)	PROSA Z score (σ)
Ecδ′	—	−11.0
Hiδ′	33/47	−10.1
Paδ′	28/38	−10.1
BsYAAS	20/28	−9.3
Ecγ	21/28	(−8.2, −7.6, −7.0, 0.2)
Hiγ	20/28	−9.0
Bsγ	19/27	−8.7
HsRFC4	17/22	−6.0
ScRFC4	14/19	−7.0
HsRFC5	18/19	−6.8
ScRFC5	18/19	−6.8
HsRFC2	18/23	(−7.7,−7.1,−6.6, 0.3)
ScRFC2	16/20	−7.6
HsRFC3	18/21	−7.0
ScRFC3	18/22	−7.4
T4g44p	11/16	−6.9
ScYN02	13/12	−6.8
HiYCAJ	13/13	−7.4
EcRuvB	12/8	(−7.5,−6.5,−5.7, 0.3)
Correct PCNA	7	−5.8
Incorrect PCNA	6	−1.2

The sequence identity relative to E. coli δ′ in the final sequence-structure alignment is shown for Domain 1 and for the entire alignment. The PROSA Z scores (see text) for models generated using these alignments and MODELLER are shown. For E. coli γ, Human RFC-2, and E. coli RuvB, which are shown in Figure 7, the numbers in parentheses are the minimum (best), mean, maximum (worst), and standard deviation of the Z scores for 100 models generated from the same final sequence alignment. For other sequences, only single models were calculated. Abbreviations, Ec, Escherichia coli; Hi, Haemophilus influenzae; Bs, Bacillus subtilis; Hs, Homo sapiens; Sc, Saccharomyces cerevisiae; T4g44p, Bacteriophage T4 gene 44 protein. YAAS, YN02, and YCAJ are uncharacterized open reading frames.

of the sequence might be in the vicinity of the adenine group of ATP, and consequently, the conserved sequence motif might be involved in the recognition of the adenine. A similar motif is conserved in certain sequences containing helicase motifs, such as E. coli recG, human recQ, and human WRN, but not in other related sequences such as recA, recF, or recN (not shown).

DExx Motif

In addition to the Walker A motif, NTPases contain one or more acidic residues that are involved in metal binding and the catalytic mechanism. The "DEAD" sequence motif was first identified in helicases (Linder et al., 1989), and is found in four of the RFC subunits (Figure 3). The last two residues of the motif are less conserved, and the motif is more appropriately referred to as "DExx" (Hodgman, 1988; Gorbalenya et al., 1989; Koonin et al., 1991). A "DExx" (Walker B) motif is conserved in the sequences shown in Figure 3 and Table 2, except for the δ′ sequences and the RFC-5 subunits. The location of the DExx motif in δ′, on a strand one removed from the P loop, is similar to that seen in monomeric or dimeric DNA- and RNA-dependent helicases (Subramanya et al., 1996; Yao et al., 1997).

Possible Sensors for ATP Binding or Hydrolysis

The structures of well-characterized NTPases have shown how interactions between the protein and the phosphate groups of the nucleotide act as sensors or switches that mediate conformational changes. The location and nature of these residues can be quite different in the various NTPases, but two possible ATP sensors in the clamp-loading subunits are suggested by the δ′ structure. One is inferred by analogy to the structure of Ras (Brünger et al., 1990; Pai et al., 1990), and the other by analogy to adenylate kinase (Müller and Schulz, 1992).

In Ras, the β strand adjacent to the P loop (corresponding to β4 in the δ′ structure) and the segment immediately following it adopt distinct conformations in the presence and absence of GTP (Brünger et al., 1990; Pai et al., 1990). The γ phosphate of GTP interacts with the peptide backbone in the C-terminal region of the β strand, which is known as the "Switch II" region (Figure 4d) (Brünger et al., 1990; Pai et al., 1990). A corresponding backbone–phosphate interaction is less likely in the clamp-loader subunits and RuvB, which lack a glycine residue in strand β4 (present in Ras) that allows β4 in Ras to approach the terminal phosphate closely. Instead, these proteins have an asparagine or threonine residue in this region (denoted "Sensor-1 motif" in the sequence alignment, Figure 3). This residue would be in a position to form hydrogen bonding interactions with the terminal phosphate of ATP, and thus respond to nucleotide binding or hydrolysis. An analogous sensor of ATP binding/hydrolysis has been proposed for RecA (Story and Steitz, 1992) and DExx helicases (Subramanya et al., 1996).

Adenylate kinase undergoes a large segmental conformational change upon binding its substrates ATP and ADP (Müller and Schulz, 1992; Gerstein et al., 1993). The net effect is to close the active site by moving a small (∼25 residue) segment (colored orange in Figure 4c) over the active site upon ATP binding, thereby sequestering the catalytic site from water (Müller and Schulz, 1992). Although unrelated in sequence or conformation, Domain 2 in δ′ shares the same spatial relationship to the P loop as the lid segment in adenylate kinase: strand β5 in δ′ leads into Domain 2, and the corresponding strand in adenylate kinase leads into the lid segment. Arginine residues in the lid segment of adenylate kinase are important mediators of the conformational change, since they bind to the phosphate groups of ATP (Müller and Schulz, 1992). In δ′, the first turn of helix α9 in Domain 2 packs against the P loop (Figure 5), and all the clamp-loader and the RuvB sequences, except δ′, contain an arginine in the first turn of this helix in Domain 2 (Figure 5). This arginine is part of a conserved "Sensor-2 motif," G/PxΦRxΦ, where Φ denotes a hydrophobic residue (Figures 3 and 4). A mutation in the S. cerevisiae *cdc44* gene (the yeast homolog of RFC-1) that results in a defect in DNA replication has been localized to the Sensor-2 region (McAlear et al., 1994). This defect can be rescued by mutations in PCNA, suggesting a direct connection between the Sensor-2 motif and clamp loading (McAlear et al., 1994).

Sequence-Structure Alignments

Sequence homology between the δ′ sequence and that of other clamp-loader subunits had been suggested earlier based only on sequence comparisons (Carter et al.,

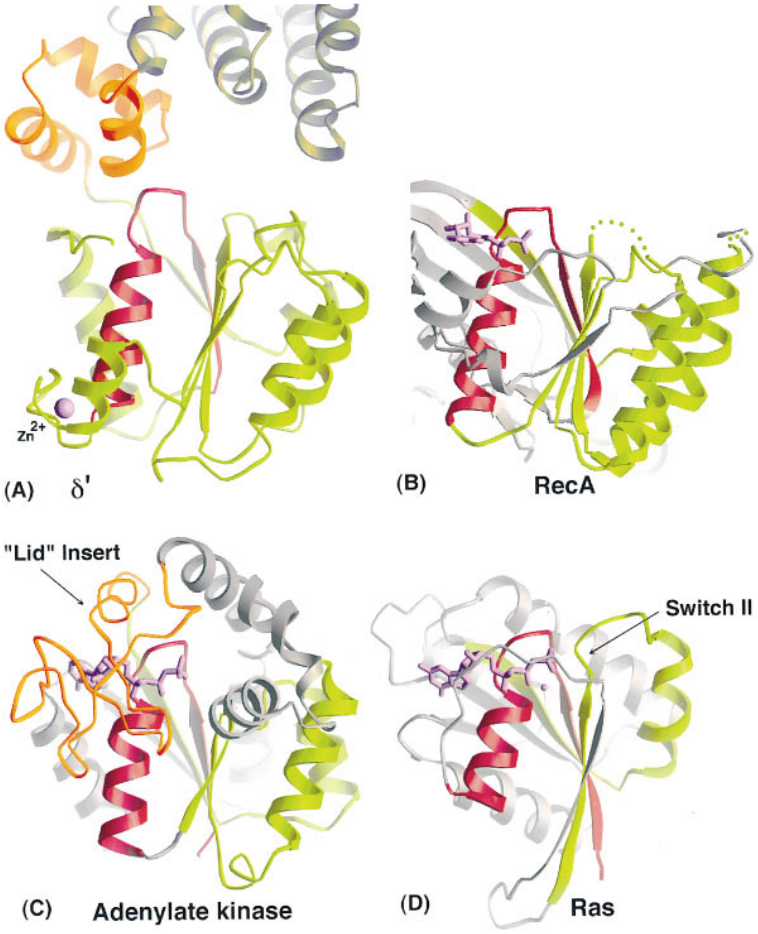

Figure 4. Nucleotide Binding Domains

The structure of δ′ (a) is compared to that of RecA ([b]; PDB entry 1rea, Story and Steitz, 1992), adenylate kinase ([c]; PDB entry 1ake, Müller and Schulz, 1992), and Ras ([d]; PDB entry 5p21, Pai et al., 1990). The structures were superimposed on the P-loop segment, shown in red. The δ′ structure is colored as in Figure 1. In adenylate kinase and Ras, structural elements that are located in a similar context to that in δ′ are colored green; other elements are colored grey. The lid domain of adenylate kinase is shown in orange, and the polypeptide chain is drawn in thinner lines in this region. ATP and GTP are colored violet in adenylate kinase and Ras, respectively, and a magnesium ion in Ras is indicated as a small violet sphere. The "lid" domain (orange) in adenylate kinase is shown in the closed conformation. It undergoes a large structural transition to an open conformation in the absence of ATP (Müller and Schulz, 1992).

1993; O'Donnell et al., 1993; Cullman et al., 1995). However, the levels of sequence identity are low, ranging from 21% between E. coli δ′ and E. coli γ, 9%–18% between E. coli δ′ and various human RFC subunits, and 11% between E. coli δ′ and T4 gene 44 protein (sequence identities are from the structure-based alignment, Table 2). BLAST searches (Altschul et al., 1990) using the δ′ and other clamp-loader sequences also revealed a possible connection between these proteins and the RuvB family of hexameric helicases (see Experimental Procedures). This connection was further substantiated by using ALIGN, a sequence alignment program that uses a local dynamic programming algorithm with a new gap penalty function (Altschul, submitted). ALIGN generated a locally optimal sequence alignment between E. coli γ and E. coli RuvB over an ~250 residue range that spans all three domains with few gaps. A notable feature of this alignment, which was generated

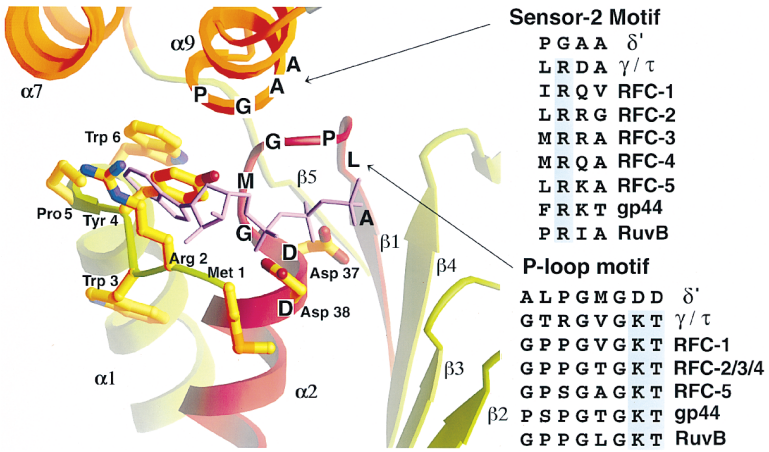

Figure 5. Inactive Nucleotide-Binding Region of δ′

The region surrounding the P loop of δ′ is shown, with side chains of certain residues indicated in stick representation. P-loop sequences and sensor sequences in δ′ and other proteins are indicated. The structure of ATP is taken from adenylate kinase, after the P-loop segments of adenylate kinase and δ′ were superimposed. Residues that are actual or potential ligands of the phosphate groups of ATP are highlighed in blue in the alignment of sequence motifs.

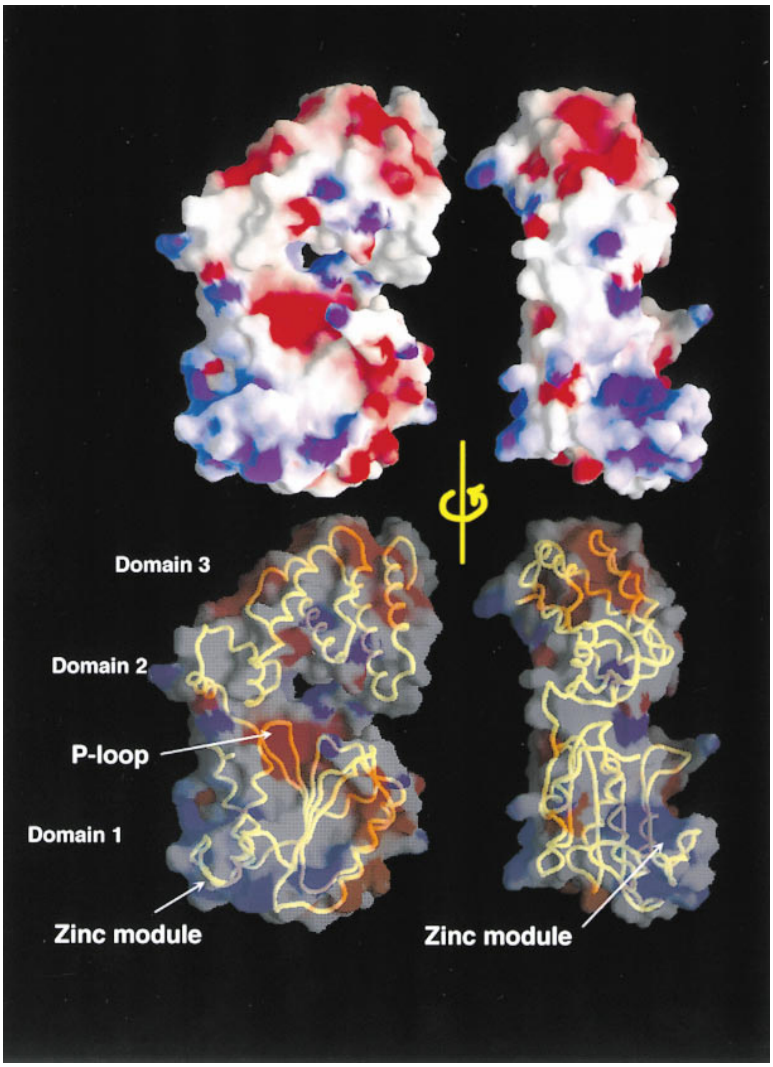

Figure 6. GRASP Diagrams of δ′

The molecular surface of δ′ is shown in two orientations, colored according to local electrostatic potential. A transparent representation of the surface with the polypeptide backbone revealed underneath is also shown. Note that there are no contacts between Domains 1 and 3, and that the appearance of such contacts in the GRASP diagram is a consequence of the orientation of the molecule. Figure prepared using GRASP (Nicholls et al., 1991).

automatically without reference to structural information, is that it includes Sensor 2 in Domain 2 (the conserved GxΦRxΦ motif) as well as a region of high local sequence similarity between γ and RuvB in helix α12 in Domain 3 (LINEAAARGIEWE in γ, LDNLAAAIGEERE in RuvB, see Figure 3). Similar results are obtained upon using δ′ instead of γ, but the lack of the functional ATP binding site in δ′ weakens the sequence similarity.

The RecA-like core (Story et al., 1992) is common to δ′ and to other ATPases such as monomeric or dimeric helicases (Subramanya et al., 1996) and the F_1 ATPase (Abrahams et al., 1994). The significance of the ALIGN result is that it suggests a similarity between γ, δ′, and RuvB that extends well beyond the RecA-like core, into Domains 2 and 3 of δ′. However, the overall sequence identity between δ′ and the other proteins is low, except for γ, and reliable sequence-based alignments are not possible (Johnson and Overington, 1993). We therefore turned to an iterative process (Guenther, 1996; Sanchez and Sali, 1997a, 1997b) of manual sequence-structure alignment, three-dimensional model generation (Sali and Blundell, 1993), and model evaluation using the PROSA program (Sippl, 1993). This iterative process provides a means of establishing and evaluating possible structural relationships between δ′ and proteins of unknown structure. Details of the process and a test of its validity are described further in the Experimental Procedures. Briefly, the method involves generating three-dimensional models based on the proposed sequence alignments, and evaluating the quality of the model using empirical energy functions.

An overall measure of the accuracy of the alignment is the PROSA Z score of the corresponding model, which is the difference between the empirical energy of the model and the mean energy of many unrelated folds onto which the same sequence is "threaded," expressed in units of the standard deviation of the energy distribution (Sippl, 1993). The Z score is proportional to the length of the protein. For experimentally determined protein structures with ~100 and ~300 residues, the Z scores are in the range of −8 to −4 σ and between −10 and −6 σ, respectively (Sippl, 1993).

More than 30 clamp-loader and RuvB sequences as well as sequences derived from several uncharacterized open reading frames were aligned with that of δ′ using this procedure. Sequence identities and PROSA Z scores

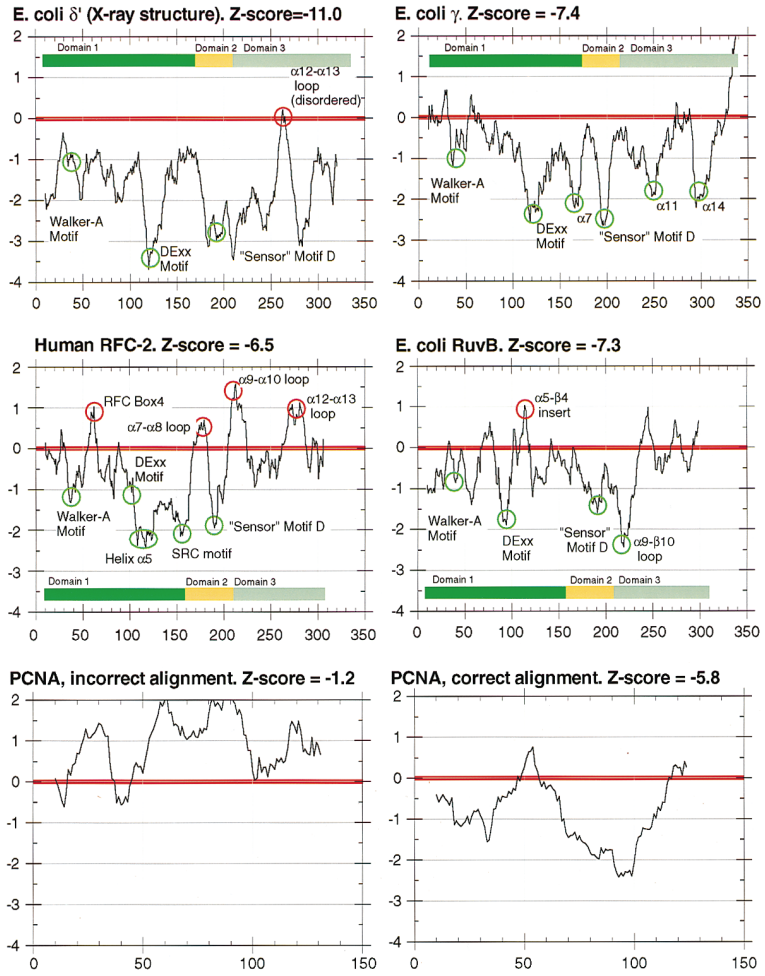

Figure 7. PROSA Energy Profiles of Various Structures and Models

(a) E. coli δ′
(b) E. coli γ subunit
(c) Human RFC-2
(d) E. coli RuvB
(e) S. cerevisiae PCNA, N-terminal domain, modeled on the structure of E. coli β subunit, using the manual (incorrect) alignment of Kong et al. (1992).
(f) The same domain of PCNA, modeled on β but with the correct structure-based alignment of Krishna et al. (1994). The residue numbers for δ′ refer to the actual structure. For the other proteins, the residue numbers are relative to the first residue used in the alignment, and the total number of residues differ because of insertions and deletions. The graphs cannot therefore be directly overlaid. Certain landmarks in the sequences are indicated, as is the domain structure of δ′. Sensor Motif D refers to the ''Sensor 2'' motif of Figure 3. The PROSA Z scores are given above each graph. All models were calculated with MODELLER.

for a representative set of 18 are shown in Table 2, and sequence alignments for 6 of them are shown in Figure 3. The overall sequence identities with respect to E. coli δ′ range from 33%, for the δ′ sequence from Hemophilus influenzae, to 11% for T4 gene 44 protein (Table 2). The Z scores range from −10 σ for the δ′ sequences to −6.8 σ for certain RFC sequences. These Z scores are within the range obtained for experimentally determined structures (Sippl, 1993), supporting the proposed structural relationship. This contrasts with the Z score of −1.2 σ for the incorrect manual aligment of PCNA and β (see Experimental Procedures and Table 2).

The per-residue PROSA energy profile is negative throughout most of the modeled structure of γ, indicating that the entire structure of δ′ may be a good model for the structure of the corresponding region of the bacterial γ subunits (Figure 7b). The results also suggest that Domains 1 and 2 of δ′ are likely to be reliable models for the general chain fold of the corresponding regions of the RFC proteins, gene 44 protein, and the RuvB family of proteins. This structural correspondence is also suggested by the presence of conserved sequence motifs that span Domains 1 and 2. The lack of such sequence motifs in Domain 3 makes the reliability of the alignments much more difficult to assess in this region. PROSA profiles for the models and the secondary structure predictions that led to the sequence alignment both suggest that Domain 3 in the RFC proteins, gene 44 protein, and RuvB is likely to be helical, but substantial structural rearrangements are expected since one of the helices appears to be missing or truncated in these proteins (Figure 3).

Conclusions

Examination of the structure of δ′ suggests that the most likely consequence of ATP binding or hydrolysis by the γ subunit and other active clamp-loader proteins would be a conformational change that causes the mouth of the "C" to open or close. Molecular surface representations (Figure 6) of the δ′ subunit reveal a structure that seems poised to undergo structural transitions that would alter the relative dispositions of Domains 1 and 3. The possibility of such conformational flexing is suggested by the shape of the molecule, by the limited interactions between residues at the interfaces between the domains, and by the predicted location of the ATP binding site in γ. In particular, a degree of flexibility between Domains 2 and 3 is suggested by the fact that the hydrophobic cores of these domains are separated by a network of hydrogen-bonded polar side chains that are partially buried at the interdomain interface, and by the complete lack of direct contacts between residues in Domains 1 and 3.

The connection between the clamp-loading subunits and RuvB was unanticipated. The bacterial RuvA and RuvB proteins are components of the inducible SOS systems of DNA repair, and are involved in homologous recombination (for review, see West, 1996). The RuvAB complex is a molecular motor that binds Holliday junctions and promotes branch migrations in an ATP-dependent manner (Shiba et al., 1991; Tsaneva et al., 1993). The E. coli RuvB protein (336 amino acids) is about the same size as δ′ (334 amino acids). The implied structural similarity between RuvB and δ′ suggests that RuvB most likely does not belong to the family of DExx box helicases represented by the structure of the PcrA helicase from Bacillus stearothermophilus (Subramanya et al., 1996). The PcrA structure consists of two domains that are related to each other and that in turn resemble the nucleotide binding domain of RecA. One of these domains of PcrA binds ATP. The other domain, in a situation reminiscent of δ′, has lost ATP binding activity. Topologically, the location of the second domain in PcrA corresponds to that of Domain 2 of δ′. Our sequence alignment (Figure 3) would suggest that instead of this α/β domain, RuvB contains a three helix bundle (Domain 2) that also acts as a sensor of ATP binding. In addition, RuvB contains the N-terminal motif (RFC-Box II), a signature of the clamp-loader subunits that is distinct from the RecA scaffold. Note that the proposed structural relationship with Domains 2 and 3 of δ′ is specific to RuvB, and that other hexameric helicases such as the SV40 middle T protein do not appear to be related to δ′ in these regions.

It is interesting that the δ′ subunit, an essential component of the clamp-loader complex, is not itself an ATPase. The presence of an inactive subunit in a protein complex that transduces ATP energy into mechanical motion is not without precedent. The PcrA helicase has two domains that have similar nucleotide-binding folds, only one of which actually binds nucleotides (Subramanya et al., 1996). Likewise, in F_1 ATPase, three of the subunits of the hexameric ring structure are nucleotide binding proteins that are catalytically inactive, and appear to act as passive mechanical couplers (Boyer, 1993; Abrahams et al., 1994). From an evolutionary point of view, it may be that a process of gene duplication and specialization has been utilized to generate protein subunits that are necessary for the proper functioning of various ATP-dependent machineries, but which have subsequently lost the now redundant function of binding nucleotides. Further clarification of the role of δ′ in such a mechanism awaits the determination of the structures of intact clamp-loading complexes.

Experimental Procedures

Crystallization and X-Ray Data Collection

Recombinant E. coli δ′ was purified as described (Dong et al., 1993). Briefly, E. coli cells overexpressing δ′ were lysed by freeze-thaw, and the supernatant was fractionated by ammonium sulfate precipitation. δ′ was then purified by heparin-agarose, Q-Sepharose, and epoxyaminohexane Sepharose chromatography. The purified protein was stored at −70°C in 20 mM Tris (pH 7.5), 20 mM NaCl, 10% glycerol, 0.5 mM EDTA, and 2 mM DTT. The protein was concentrated to 20 mg/ml for crystallization, with the best crystals obtained using vapor diffusion at 4°C, with a reservoir solution containing 20%–27% PEG 400, 100 mM HEPES (pH 6.8), 100 mM $MgCl_2$, 1%–3% glycerol, 10 mM $MgSO_4$ at 4°C. Plate-like crystals are obtained that grow in dense clusters, with individual plates of dimensions of up to $\sim 1.0 \times \sim 0.4 \times \sim 0.2$ mm^3.

The crystals were stabilized in a cryosolvent (45% PEG 400, 100 mM HEPES [pH 6.8], 100 mM $MgCl_2$, 20% glycerol, 10 mM $MgSO_4$) and flash frozen. Crystals frozen using this cryosolvent were used for the heavy atom screening process, and yielded diffraction patterns that extended to 2.7 Å at best, using a laboratory X-ray source. Significantly improved diffraction patterns were obtained when this cryosolvent was augmented by adding δ′ protein (final protein concentration, 4 mg/ml), leading to the measurement of data at 2.2 Å resolution. However, the addition of protein to the cryosolvent led to severe nonisomorphism from crystal to crystal, and this procedure was not used for heavy atom data collection. The crystals are in space group $C222_1$, with a = 98.9/98.9 Å, b = 104.0/103.7 Å, c = 75.6/75.5 Å for the crystals frozen with/without protein in the cryosolvent. There is one molecule in the asymmetric unit.

Initial phases were obtained from five heavy-atom derivatives (Table 1). The MIR phases are of relatively poor quality, and SQUASH (Zhang and Main, 1990) was utilized to improve the MIR electron density map. As suggested by Smith et al., 1994, the best results were obtained by determining a reliable molecular envelope at low resolution (4.5 Å), and using SQUASH to improve the phases from this resolution outward. This procedure resulted in an interpretable electron density map, into which a model was built using the program "O" (Jones et al., 1991). The model was refined using X-PLOR (Brünger, 1992), resulting in final R values of 20.5% and 21.7%, with free R values of 26.5% and 27.7% for data with $|F| > 2\sigma(|F|)$ and for all data, respectively. The final model includes 2512 nonhydrogen protein atoms, 1 zinc atom, and 202 water molecules. The average temperature factors of protein backbone and side-chain atoms and water molecules are 24, 26, and 34 Å^2, respectively. PROSA profile for the final structure is shown in Figure 7. The PROSA Z score ($-11.0\ \sigma$) and the favorable energy profile for the entire structure are consistent with an accurate structure determination (Sippl, 1993).

Structure-Based Sequence Alignment

The BLAST program (Altschul et al., 1990) on the National Center for Biotechnology Information (NCBI) web site (www.ncbi.nlm.nih.gov) was used to detect proteins that are possibly related to E. coli δ′. The BLAST search was used only for the detection of possible matches, and not for the generation of sequence alignments, which was done using the δ′ structure as a reference. The BLAST searches progressed through pairwise identifications with cutoffs of $P < 10^{-8}$. Utilization of the E. coli δ′ sequence readily identified the bacterial γ/τ sequences, and searches with the γ sequences in turn identified the eukaryotic RFC subunits and RuvB.

The sequences obtained from the BLAST search were then utilized for the generation of a structure-based alignment. The PHD program (Rost and Sander, 1994) was used to generate secondary structure predictions for each sequence, which in turn led to an initial sequence alignment, based on the amphipathicity and extents of the secondary structural elements. Each of the sequence alignments was then used to generate three-dimensional models based on the δ′ structure using a fully automated procedure implemented in MODELLER (Sali and Blundell, 1993). The models were evaluated using empirical energy profiles generated by PROSA (Sippl, 1993). This model evaluation program indicated both global and local problems with particular alignments. Such alignments were manually adjusted, and the model building and evaluation procedures were repeated. This cycle of alignment, model building, and evaluation was repeated up to about 20 times for some sequences. Such a procedure has previously proven useful in resolving several difficult sequence-structure alignment problems (Sanchez and Sali, 1997a, 1997b). The overall reliability of the final structure-based alignments of E. coli δ′ with 18 target sequences and the overall quality of the corresponding 3-D models are described in Table 2, which lists the pairwise sequence identities and 3-D model Z scores.

Manual alignments between the sequences of two proteins that share less than ∼30% sequence identity can lead to gross errors in the relative positioning of secondary structural elements (Johnson and Overington, 1993). For example, upon determination of the

structure of the dimeric β ring of E. coli pol III, it was postulated that eukaryotic PCNA would form a trimeric ring (Kong et al., 1992). Each monomer of PCNA, which has a monomer molecular weight ~2/3 that of β, was predicted to comprise two topologically identical structural domains, each of which was similar to the three repeated domains of the β monomer. Determination of the crystal structure of PCNA revealed that this prediction was accurate in general terms: the architecture of the trimeric PCNA ring is strikingly similar to that of the dimeric β ring (Krishna et al., 1994). However, the manual sequence alignment used in the earlier paper (Kong et al., 1992) turned out to be almost entirely incorrect (Krishna et al., 1994).

To validate the iterative alignment procedure used here, we carried out the following test on β and PCNA. The sequence identity between the various domains of β and PCNA is low (~10% in both correct and incorrect alignments), and the manual alignment had errors in the register of many equivalent secondary structural elements. We used the PROSA program (Sippl, 1993) to evaluate two comparitive models for PCNA that were generated by an automated procedure using MODELLER (Sali and Blundell, 1993) and the structure of the β subunit as a template. In one, the structure of one domain of PCNA was generated using the β structure, based on the incorrect manual alignment (Kong et al., 1992). In the other, the β structure was again used as a template, but the accurate structure-based alignment between the two sequences (Krishna et al., 1994) was utilized instead. Figures 7e and 7f show that the PROSA potential is able to distinguish clearly between the incorrect and correct alignments (Figures 7e and 7f and Table 2).

For the PCNA models generated in the test (~130 residues), the Z scores are −1.2 σ and −5.8 σ for the incorrect and correct alignments, respectively (Table 2). A more detailed analysis is provided by the per-residue values of the PROSA empirical energy (Figure 7). The PROSA energies are positive for almost the entire span of the PCNA model generated using the incorrect alignment, whereas the energies are generally negative for the model based on the correct sequence alignment (Figures 7d and 7e).

The results with β and PCNA, considered along with previously reported tests of the PROSA potential (Sippl, 1993), suggest that grossly incorrect sequence alignments are likely to be detected by this method. However, it cannot be ruled out that significantly inaccurate regions of a model might still have a good PROSA profile. Consequently, the structure-based sequence alignments shown here should be considered only as indicative of plausible structural relationships that have to be tested by mutagenesis and further structure determination.

Acknowledgments

We are grateful to Stephen Altschul for the ALIGN program and to Manfred Sippl for the PROSA program. A. S. is a Sinsheimer Scholar. This work was supported in part by grants from the National Institutes of Health (GM 45547 to J. K., GM 38839 to M. O'D., and GM 54762 to A. S.) and the National Science Foundation (BIR 9601845 to A. S.).

Received July 7, 1997; revised September 24, 1997.

References

Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. Nature *370*, 621–628.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Boyer, P.D. (1993). The binding change mechanism for ATP synthase—some probabilities and possibilities. Biochim. Biophys. Acta *1140*, 215–250.

Brünger, A.T. (1992). X-PLOR (Version 3.1) Manual (New Haven: The Howard Hughes Medical Institute and Yale University).

Brünger, A.T., Milburn, M.V., Tong, L., deVos, A.M., Jancarik, J., Yamaizumi, Z., Nishimura, S., Ohtsuka, E., and Kim, S.H. (1990). Crystal structure of an active form of RAS protein, a complex of a GTP analog and the HRAS p21 catalytic domain. Proc. Natl. Acad. Sci. USA *87*, 4849–4853.

Carter, J.R., Franden, M.A., Aebersold, R., and McHenry, C.S. (1993). Identification, isolation, and characterization of the structural gene encoding the δ′ subunit of Escherichia coli DNA polymerase III holoenzyme. J. Bacteriol. *175*, 3812–3822.

Cronet, P., Bellsolell, L., Sander, C., Coll, M., and Serrano, L. (1995). Investigating the structural determinants of the p21-like triphosphate and Mg^{2+} binding site. J. Mol. Biol. *249*, 654–664.

Cullman, G., Fien, K., Kobayashi, R., and Stillman, B. (1995). Characterization of the five replication factor C genes of *Saccharomyces cerevisiae*. Mol. Cell. Biol. *15*, 4661–4671.

Dong, Z., Onrust, R., Skangalis, M., and O'Donnell, M. (1993). DNA polymerase III accessory proteins. I. holA and holB encoding δ and δ′. J. Biol. Chem. *268*, 11758–11765.

Flower, A.M., and McHenry, C.S. (1986). The adjacent dnaZ and dnaX genes of Escherichia coli are contained within one continuous open reading frame. Nucl. Acids Res. *14*, 8091–8101.

Flower, A.M., and McHenry, C.S. (1990). The γ subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. Proc. Natl. Acad. Sci. USA *87*, 3713–3717.

Gerstein, M., Schulz, G., and Chothia, C. (1993). Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers. J. Mol. Biol. *229*, 494–501.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair, and expression of DNA and RNA genomes. Nucl. Acids Res. *17*, 4713–4730.

Guenther, B.D. (1996). PhD thesis. Structural studies on the DNA replication apparatus: X-ray crystal structures of the delta-prime subunit of Escherichia coli DNA polymerase III (New York: The Rockefeller University).

Hodgman, T.C. (1988). A new superfamily of replicative proteins. Nature *333*, 22–23.

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.

Jarvis, T.C., Paul, L.S., and von Hippel, P.H. (1989). Structural and enzymatic studies of the T4 DNA replication system. I. Physical characterization of the polymerase accessory protein complex. J. Biol. Chem. *264*, 12709–12716.

Johnson, M.S., and Overington, J.P. (1993). A structural basis for sequence comparisons. An evaluation of scoring methodologies. J. Mol. Biol. *233*, 716–738.

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.

Kelman, Z., and O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. Annu. Rev. Biochem. *64*, 171–200.

Klug, A., and Schwabe, J.W. (1995). Protein motifs 5. Zinc fingers. FASEB J. *9*, 597–604.

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

Koonin, E.V., Choi, G.H., Nuss, D.L., Shapira, R., and Carrington, J.C. (1991). Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses. Proc. Natl. Acad. Sci. USA *88*, 10647–10651.

Kornberg, A., and Baker, T.A. (1991). DNA Replication, Second Edition (New York: W.H. Freeman).

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

Lee, S.-H., Kwong, A.D., Pan, Z.-H., and Hurwitz, J. (1991). Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase δ. J. Biol. Chem. *266*, 594–602.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielson, P.J., Nishi, K., Schinier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. Nature *337*, 121–122.

McAlear, M.A., Howell, E.A., Espenshade, K.K., and Holm, C. (1994). Proliferating cell nuclear antigen (pol30) mutations suppress cdc44 mutations and identify potential regions of interaction between the two encoded proteins. Mol. Cell. Biol. *14*, 4390–4397.

McHenry, C.S. (1991). DNA polymerase III holoenzyme. J. Biol. Chem. *266*, 19127–19130.

Müller, C.W., and Schulz, G.E. (1992). Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap_5A refined at 1.9 Å resolution: a model for a catalytic state. J. Mol. Biol. *224*, 159–177.

Naktinis, V., Turner, J., and O'Donnell, M. (1996). A molecular switch in a replication machine defined by an internal competition for protein rings. Cell *84*, 137–145.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Nossal, N.G., and Alberts, B.M. (1984). Mechanism of DNA replication catalyzed by purified T4 replication proteins. In Bacteriophage T4, C. Mathews, ed. (Washington, D.C.: American Society for Microbiology), pp. 71–81.

O'Donnell, M., Onrust, R., Dean, F.B., Chen, M., and Hurwitz, J. (1993). Homology in accessory proteins of replicative polymerases—*E. coli* to humans. Nucl. Acids Res. *21*, 1–3.

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.

Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W., and Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. EMBO J. *9*, 2351–2359.

Rost, B., and Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. Prot. Struct. Funct. Genet. *19*, 55–72.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Sanchez, R., and Sali, A. (1997a). Advances in comparative protein-structure modeling. Curr. Opin. Struct. Biol. *7*, 206–214.

Sanchez, R., and Sali, A. (1997b). Evaluation of comparative structure modelling by MODELLER-3. Prot. Struct. Funct. Genet., in press.

Schulz, G.E. (1992). Binding of nucleotides by proteins. Curr. Opin. Struct. Biol. *2*, 61–67.

Shiba, T., Iwasaki, H., Nakata, A., and Shinagawa, H. (1991). SOS-inducible DNA repair proteins, RuvA and RuvB, of *Escherichia coli*: functional interactions between RuvA and RuvB for ATP hydrolysis and renaturation of the cruciform structure in supercoiled DNA. Proc. Natl. Acad. Sci. USA *88*, 8445–8449.

Sippl, M. (1993). Recognition of errors in three-dimensional structures of proteins. Prot. Struct. Funct. Genet. *17*, 355–362.

Smith, J.L., Zaluzec, E.J., Wery, J.P., Nin, L., Switzer, R.L., Zalkin, H., and Satow, Y. (1994). Structure of the allosteric regulatory enzyme of purine biosynthesis. Science *264*, 1427–1433.

Story, R.M., and Steitz, T.A. (1992). Structure of the recA protein–ADP complex. Nature *355*, 374–376.

Story, R.M., Weber, I.T., and Steitz, T.A. (1992). The structure of the E. coli recA protein monomer and polymer. Nature *355*, 318–325.

Subramanya, H.S., Bird, L.E., Brannigan, J.A., and Wigley, D.B. (1996). Crystal structure of a DExx box DNA helicase. Nature *384*, 379–383.

Tsaneva, I.R., Muller, B., and West, S.C. (1993). RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity in vitro. Proc. Natl. Acad. Sci. USA *90*, 1315–1319.

Tsuchihashi, Z., and Kornberg, A. (1990). Translational frameshifting generates the γ subunit of DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA *87*, 2516–2520.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the α- and β- subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. *1*, 945–951.

West, S.C. (1996). DNA helicases: new breeds of translocating motors and molecular pumps. Cell *86*, 177–180.

Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A.D., Le, H.V., and Weber, P.C. (1997). Structure of the hepatitis C virus RNA helicase domain. Nat. Struct. Biol. *4*, 463–467.

Yin, K.C., Blinkowa, A., and Walker, J.R. (1986). Nucleotide sequence of the Escherichia coli replication gene dnaZX. Nucl. Acids Res. *14*, 6541–6549.

Zhang, K.Y.J., and Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Crystallogr. *A46*, 377–381.