Cell, Vol. 97, 75–84, April 2, 1999, Copyright ©1999 by Cell Press

# Crystal Structures of Complexes of PcrA DNA Helicase with a DNA Substrate Indicate an Inchworm Mechanism

**Sameer S. Velankar,† Panos Soultanas,† Mark S. Dillingham,† Hosahalli S. Subramanya, and Dale B. Wigley***
Sir William Dunn School of Pathology
University of Oxford
South Parks Road
Oxford OX1 3RE
United Kingdom

## Summary

**We have determined two different structures of PcrA DNA helicase complexed with the same single strand tailed DNA duplex, providing snapshots of different steps on the catalytic pathway. One of the structures is of a complex with a nonhydrolyzable analog of ATP and is thus a "substrate" complex. The other structure contains a bound sulphate ion that sits in a position equivalent to that occupied by the phosphate ion produced after ATP hydrolysis, thereby mimicking a "product" complex. In both complexes, the protein is monomeric. Large and distinct conformational changes occur on binding DNA and the nucleotide cofactor. Taken together, these structures provide evidence against an "active rolling" model for helicase action but are instead consistent with an "inchworm" mechanism.**

## Introduction

Helicases are found in all living organisms and participate in almost every process that involves nucleic acids (Lohman and Bjornson, 1996). Sequence analysis of bacterial genomes has revealed that even in these organisms there is a requirement for around a dozen different helicases. The biochemical activity of the enzymes is to couple the free energy of hydrolysis of ATP to the separation of a DNA (or RNA) duplex into its component strands. Although the physiological role of some helicases has been determined, the functions of many others remain unclear. One such example is PcrA helicase that, although shown to be an essential enzyme in *Bacillus subtilis* (Petit et al., 1998) and *Staphylococcus aureus* (Iordanescu, 1993) involved in repair and rolling circle replication (Petit et al., 1998; Soultanas et al., 1999), has an imprecisely defined physiological role in cells. In order to learn more about this enzyme, we have initiated a study of PcrA helicase from the moderate thermophile *Bacillus stearothermophilus* and have reported the preliminary characterization of the enzyme (Bird et al., 1998a) and its crystal structure (Subramanya et al., 1996). The enzyme comprises four domains termed 1A, 1B, 2A, and 2B, with domains 1A and 2A having very similar folds. The ATP-binding site is situated in a cleft between domains 1A and 2A that is lined with a number of conserved sequence motifs that are characteristic of helicases (Gorbalenya and Koonin, 1993). PcrA is a member of a large family of helicases that have a 3′–5′ directionality (Bird et al., 1998a) and share a number of structural features (Bird et al., 1998b). Other members of this family include the Rep and UvrD helicases, which have been the subject of a great deal of study in recent years (reviewed in Lohman and Bjornson, 1996), and the NS3 RNA helicase from hepatitis C virus (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998; Porter et al., 1998).

The crystal structure of Rep helicase complexed with single-stranded DNA (Korolev et al., 1997) provided the first insights into the interaction of the protein with DNA. Intriguingly, the enzyme crystallized in two different conformations (termed "open" and "closed"), with the two molecules sitting adjacent to each other on the single-stranded $dT(pT)_{15}$ oligonucleotide. The conformational difference between the molecules comprised a large rigid body rotation of the 2B domain by approximately 130°. This "domain swiveling" was proposed to be an important aspect of the mechanism of the enzyme. Surprisingly, although Rep has been reported to be a dimer in the presence of single-stranded DNA (Chao and Lohman, 1991), the protein proved to be monomeric in the crystal structure. The only other helicase for which there are structural data is the NS3 RNA helicase (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998). The structure shows a tandem repeat of domains, each with folds similar to domains 1A and 2A but with a slightly different connectivity (Bird et al., 1998b). The third domain of the protein has no structural homology with Rep or PcrA but sits in a position roughly equivalent to that occupied by domains 1B and 2B in the closed conformation of Rep. The structure of a complex of the protein with a $(dU)_8$ oligonucleotide (Kim et al., 1998) showed the ssRNA-binding site to be located in a position equivalent to the binding site for ssDNA in the Rep helicase. Three different models for the mechanism of the enzyme have been proposed, one to accompany each of the three structures. Biochemical and structural data show the enzyme to be monomeric under a range of different conditions (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998; Porter et al., 1998).

There are two popular models for a general mechanism for helicases (reviewed in Bird et al., 1998b) (Figure 1), termed the "inchworm" (Yarranton and Gefter, 1979) or "active rolling" models (Wong and Lohman, 1992), respectively. Experiments to distinguish conclusively between these models have been difficult to design, largely because the two models actually share a number of similarities in terms of the biochemical events taking place, and most experimental observations are in fact consistent with either mechanism. There are, however, a few features distinct to each mechanism. One of the most important of these is the absolute requirement of the rolling model for (at least) a dimeric protein. By contrast, the inchworm model is consistent with any oligomeric state for the protein, including monomeric.

*To whom correspondence should be addressed (e-mail: wigley@eric.path.ox.ac.uk).
† These authors contributed equally to this work.

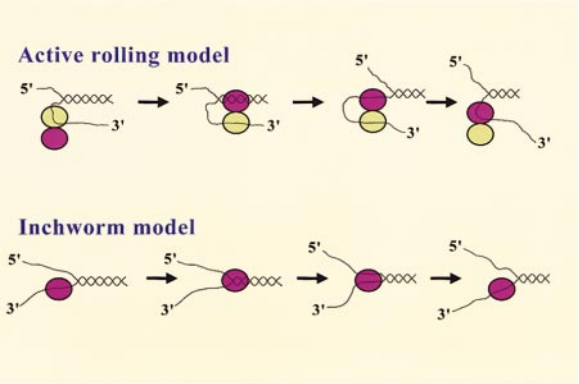


Figure 1. Active Rolling and Inchworm Models

The two most popular mechanisms for unwinding of nucleic acid duplexes by helicases (Yarranton and Gefter, 1979; Wong and Lohman, 1992). Both mechanisms require the hydrolysis of ATP, but it is not certain at which step this hydrolysis takes place, although ATP binding appears to be associated with an increased affinity of the enzyme for duplex DNA or RNA. In the scheme for the active rolling model, the two subunits of the dimer are colored differently in order to distinguish between them. Initially, the subunits of the dimer are both bound to ssDNA. As a consequence of binding ATP, one of the subunits releases the ssDNA and binds to the duplex region at the fork. This is followed by helix destabilization and the release of one of the DNA strands in a process that accompanies the hydrolysis of ATP. For the inchworm model, the enzyme monomer is bound to ssDNA and then translocates along the DNA strand to the fork region, probably upon binding ATP. Helix destabilization and release of one of the ssDNA strands takes place as ATP is hydrolyzed.

A second distinction between the models concerns the binding of DNA. A central concept of the rolling model is that each subunit of the dimer can bind to either ssDNA or duplex DNA, but never both at the same time. For the inchworm model, each protein monomer has to bind simultaneously to both ssDNA and duplex DNA during at least one point in the reaction cycle. A third difference concerns the step size for each cycle of the reaction. An inchworm mechanism, although also consistent with larger step sizes, could involve progression by as little as 1 base pair at a time, while a rolling model implies a much larger step size, at least as large as the individual binding site of each of the monomers. Estimates of the step size for different helicases vary from 4 or 5 base pairs for the UvrD helicase (Ali and Lohman, 1997), to 2 base pairs for the NS3 RNA helicase (Porter et al., 1998), to as little as 1 base pair for the Rep helicase (Yarranton and Gefter, 1979).

In order to try to distinguish between these models at the structural level, we have determined two different crystal structures of PcrA helicase complexed with a 10 base pair DNA duplex with a seven base single-stranded 3′ tail. In addition, one complex also includes a bound nonhydrolyzable analog of ATP, thus trapping a "substrate" complex. The other complex has a bound sulphate ion in the position normally occupied by a phosphate ion after hydrolysis of ATP, thereby representing a "product" complex. Large conformational changes associated with the binding of the DNA and nucleotide provide insight into the likely mechanism of helicases, supporting an inchworm rather than rolling model for these enzymes.

## Results and Discussion

### Structure of a Complex with DNA and a Sulphate Ion—A Product Complex

Initial attempts to solve the structure of the PcrA/DNA/sulphate complex by molecular replacement using the apo enzyme structure as a starting model proved to be unsuccessful, the first indication that a conformational change had taken place. The best starting model proved to be one that combined just domains 1A and 2A—domains 1B and 2B had to be located separately. The overall fold of the protein is the same as that described previously for the apo protein (Subramanya et al., 1996), but the domain orientations are very different (Figure 2). Although domains 1A and 2A are in similar orientations to that observed in the apo protein, there has been a considerable rigid body movement of domains 1B and 2B to a position more like that observed in the closed conformation of the Rep helicase complexed with ssDNA (Korolev et al., 1997). The two molecules in the asymmetric unit had slightly different orientations of the domains that, although relatively small, required individual noncrystallographic symmetry restraints to be applied to each domain during refinement. However, compared to the much larger conformational differences observed for the various complexes described in this paper, these changes are relatively small and are probably not of significance in the overall mechanism of the enzyme.

Crystallization of the PcrA/DNA complex, in the absence of a nucleotide cofactor, shows an absolute requirement for sulphate ions. There is obvious density for a sulphate ion located in the site that is occupied by a phosphate ion when the apo enzyme crystals are soaked in phosphate (our unpublished data), which is similar to the position occupied by the $\gamma$ phosphate in the ADPNP complex (see below).

The asymmetric unit of these crystals contained two PcrA/DNA complexes, raising the possibility that there might be a dimer interface evident from the molecular packing. No such interface was present. Protein contacts made between molecules involved very few side chains and were typical of those expected between molecules in a crystal lattice. The largest region of contact between any pair of complexes was mediated by base-stacking interactions between the ends of the DNA duplexes. Electron density corresponding to the bound DNA varied considerably in quality over the molecule. Although several bases in the ssDNA tail in both noncrystallographically related complexes were evident in very early stages of refinement, the duplex regions were much less clear due to disorder that is present even in the final structure. The best ordered region of the duplex is in the contacts made between the ends of duplexes from each of the noncrystallographically related monomers, but even these bases have much higher temperature factors than the protein. Consequently, we have only been able to model a few bases of the duplex DNA in the region of the contacts between symmetry related molecules. The poor electron density for the duplex DNA suggests that the contacts made with the protein are at best weak and at worst simply a consequence of crystal packing. Therefore, we believe that the details of these contacts should be regarded with extreme caution.

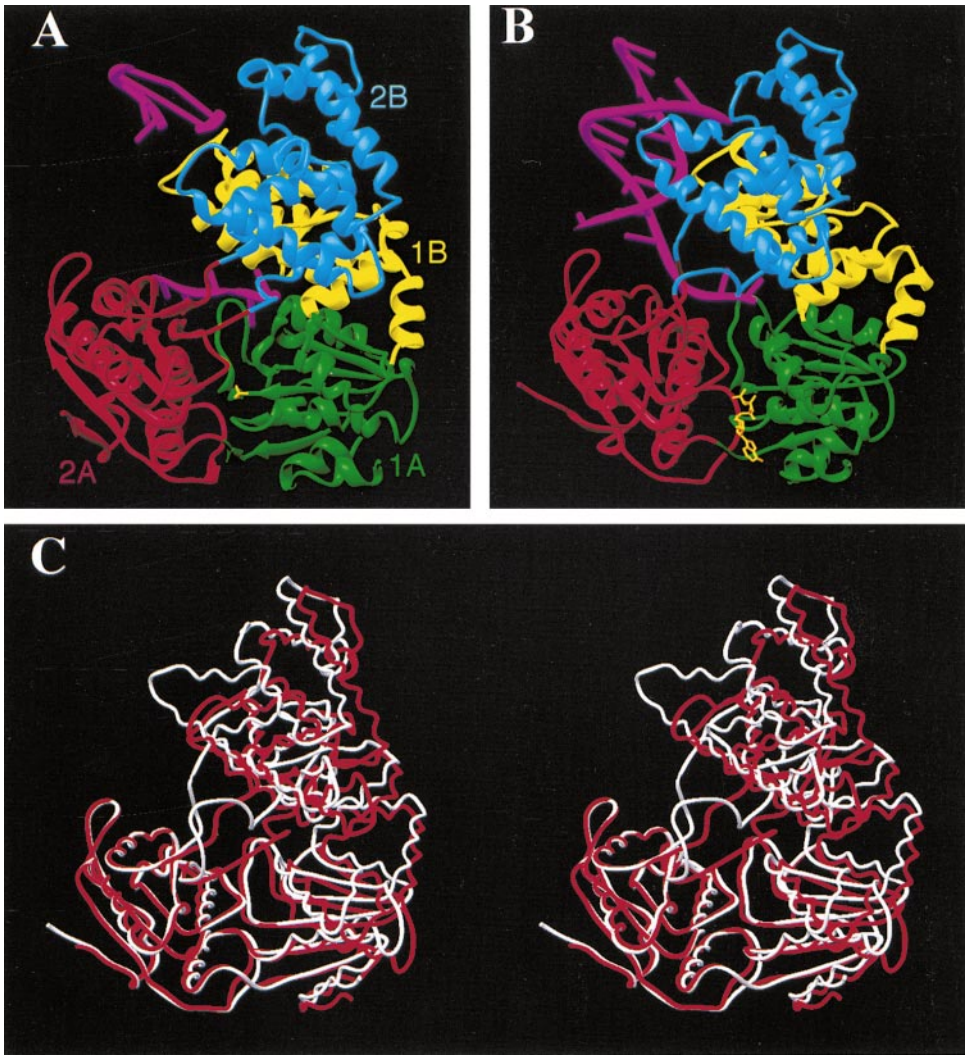


Figure 2. Domain Orientations of the Protein in the Complexes

Domain structure of (A) the product complex and (B) the substrate complex, with domain 1A colored green, domain 1B in yellow, domain 2A in red, and domain 2B in blue. The bound DNA is colored magenta, with ADPNP and sulphate in gold. These figures were produced using RIBBONS (Carson, 1991). (C) Stereo diagram of an overlay of the Ca backbone of each of the two complexes illustrating the domain movements. The molecules are superimposed on domain 2A. The substrate complex is colored white, and the product complex is red. DNA and ADPNP have been omitted for clarity.

Details of the interactions between the single-stranded tail of the DNA and the protein are shown in Figure 3. Of the seven bases in the tail, only the last five at the 3′ end are clearly visible. These five bases bind in a groove along the top of domain 1A and across onto domain 2A. The overall binding site is broadly similar to that observed for the Rep/ssDNA complex (Korolev et al., 1997). Although there are many good contacts between the protein and the single-stranded tail of the duplex, there are few contacts with the duplex DNA itself, as evidenced by the disorder of the DNA in this region. Coupled with the requirement for occupation of the phosphate ion–binding site for crystallization, this suggests that the structure of this complex mimics that of the product complex of the helicase. Therefore, in addition to the apo enzyme, this structure provides a second snapshot of the structures in the catalytic mechanism of PcrA helicase. In order to observe another intermediate in the catalytic cycle, we determined the structure of a substrate complex using the same DNA substrate and a nonhydrolyzable analog of ATP (adenylylimidodiphosphate, ADPNP).

### Structure of a Complex with DNA and ADPNP—A Substrate Complex

Initial attempts failed to solve the structure of the PcrA/DNA/ADPNP complex by molecular replacement using either the apo structure or the product complex, indicating a third, distinct conformation for this complex. The best search model proved to be one that combined domains 1A and 1B from the product complex. The other domains were then located using the product complex as a guide but required significant rigid body refinement of the domains before a satisfactory starting model was obtained from which refinement could proceed. The

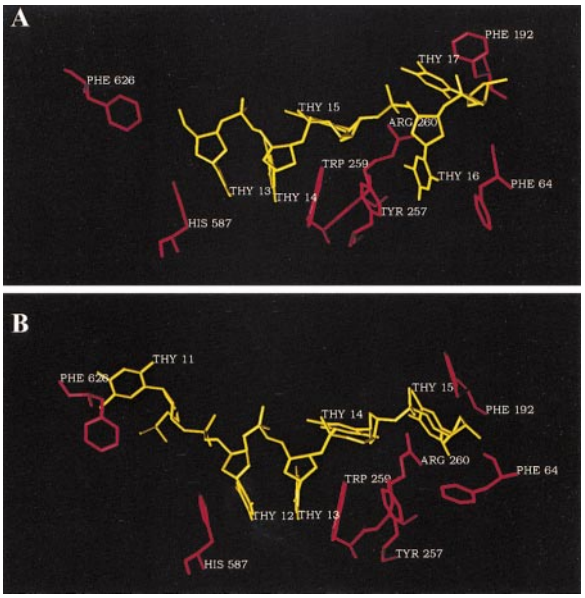


Figure 3. The ssDNA-Binding Site in Each Complex

Residues that comprise the single-stranded DNA-binding sites in (A) the product complex and (B) the substrate complex. DNA is colored yellow and protein red. There are three bases in equivalent positions in the two structures, but they are displaced by one residue in the sequence. The bases at the 3′ end of the DNA are in different binding pockets due to conformational changes in the protein that includes the rotation of the side chain of the conserved F64 to occupy one of the binding pockets in the substrate complex.

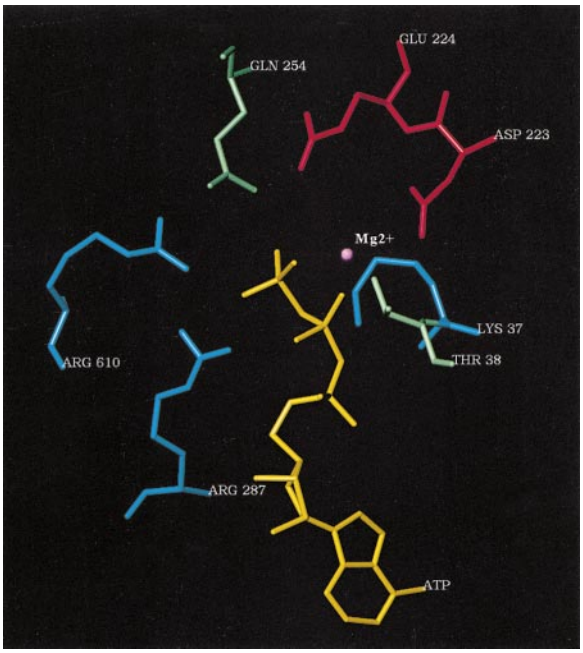


Figure 4. Details of the ADPNP-Binding Site

Residues that make contact with the γ phosphate of the ADPNP or the bound magnesium ion in the substrate complex.

overall fold of the final model of the PcrA/DNA/ADPNP complex was similar to the other structures but again with large rigid body domain movements that showed distinct differences to either the apo enzyme or the product complex (Figure 2). Although more similar to the DNA/sulphate complex, it became evident that the positions of domains 1B and 2B were altered as a consequence of the closing of the cleft between domains 1A and 2A (discussed below). These conformational changes trap the bound ADPNP in the site shown previously to be occupied by ADP in PcrA (Subramanya et al., 1996) and Rep helicases (Korolev et al., 1997), with conserved residues in domain 2A coming into much closer contact with the bound nucleotide than in the open structures. The closed form of the protein appears to be stabilized entirely by the closer interactions that are mediated by the bound ATP with no additional protein–protein contacts being made between domains 1A and 2A across the cleft. The most important additional contacts involve either the γ phosphate (mainly with domain 2A) or the bound magnesium ion (with domain 1A), suggesting how domain closure and the subsequent reopening might be coupled to the hydrolysis of ATP to ADP and Pi. Three residues make direct contacts with the γ phosphate of the ADPNP, namely Q254, R287, and R610, from conserved motifs III, IV, and VI, respectively (Figure 4). Replacement of each of these residues with alanine results in mutant PcrA proteins with impaired ATPase and helicase activities (our unpublished data). There are other conformational changes that involve residues from motif I (the "Walker A" motif). There is a small shift in the position of the glycine-rich loop that contacts the triphosphate tail, apparently as a consequence of the movement of T38 as it interacts with the bound magnesium ion that is also contacted by the conserved D223 of motif II (the DExx box). The position occupied by the bound magnesium ion is occupied by the side chain of K37 in the apo and ADP soak structures of PcrA (Subramanya et al., 1996) and in the product complex described above. Upon binding ADPNP and magnesium in the closed complex, the side chain of K37 vacates the cation-binding site and moves to contact the β phosphate of the nucleotide (Figure 4). From this position it would be able to stabilize the transition state during ATP hydrolysis by interacting with the β-phosphoryl oxygens. Replacement of this residue with alanine results in a mutant PcrA protein with impaired ATPase and helicase activities (our unpublished data).

Although electron density corresponding to the bound DNA was clearly evident even in the early stages of refinement, the limited resolution of this structure created difficulties with the correct assignment of the register of the base sequence. This problem was overcome by the use of a 5-iodo-uracil (5-iodoU) substitution at position 3 on the 10-mer strand (see Experimental Procedures). We are able to locate all 10 base pairs in the duplex region and the first five bases of the seven in the tail. Although the density becomes very weak beyond the fifth T of the tail, there is some additional electron density on the outer surface of the protein beyond an opening on the protein surface that we interpret as the disordered tail of the DNA emerging from a narrow channel that connects the ssDNA-binding site to the outside of the protein. The channel opens into a positively charged cleft across the surface of domain 1A and is therefore the likely exit of one of the ssDNA strands produced by the enzyme. DNA substrates with ssDNA tails of between seven and ten bases give apparently

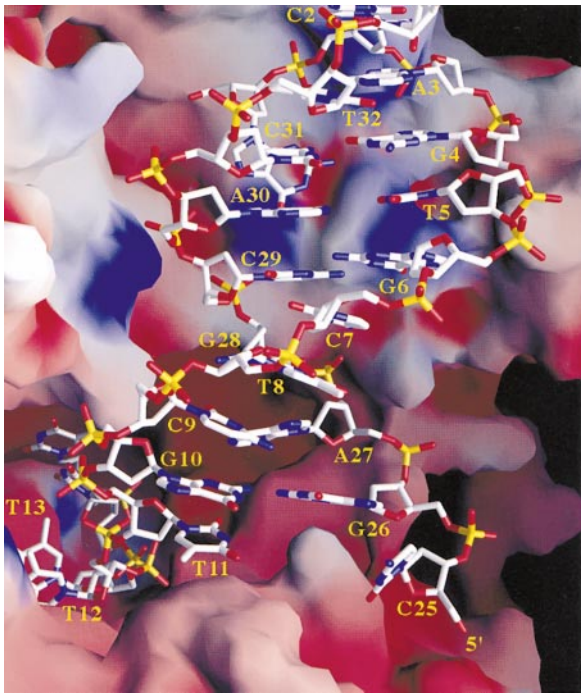


Figure 5. Electrostatic Surface Potential in the Substrate Complex

The positive potential is colored blue and negative potential red, calculated with the GRASP program (Nicholls and Honig, 1991). The bound DNA substrate is overlaid in stick representation with phosphorus atoms colored yellow, carbons white, nitrogens blue, and oxygens red. The complementarity between the molecular surface and the bound DNA is evident, as the duplex fits into a groove on the surface of the protein. Separation of the two DNA strands at the junction is also clearly visible in this view, which is a rotation of approximately 90° compared to Figure 2. Numbering of residues in the 17-mer oligonucleotide (5′-GCAGTGCTCGTTTTTTT-3′) is 1–17, and the 10-mer (5′-CGAGCACTGC-3′) is 25–34.

similar crystals, but one with a six base ssDNA tail does not crystallize. The path taken by the other ssDNA strand is more difficult to determine but presumably continues in some manner from the end of the duplex region of the DNA across the outer surface of domain 2A (see Figure 5), although it may well be that the protein makes no direct contacts with this strand. These different pathways taken by the two strands provide a convenient mechanism to prevent reannealing of the duplex. In common with the PcrA/DNA/sulphate complex, the ends of different duplexes make crystal contacts in this structure, but these are very different in the two complexes. In the PcrA/DNA/ADPNP complex, there is a non-Watson-Crick base pair between the duplexes involving two cytosine residues, while their guanosine partners fold back and make hydrogen bonds with base pairs in the major groove of the duplex region. Evidently, this unusual conformation at the ends of the duplex regions is a crystallization artifact. However, the packing of the ends of the duplexes creates an approximate continuation of the phosphodiester backbone. Symmetry related contacts between the DNA duplex region and protein molecules might, therefore, be a cautious indication of a larger duplex-binding region than that represented by the contacts within an individual protein DNA complex observed with the relatively short DNA duplex region used for this study. Examination of the molecular packing reveals that there is no protein dimer in these crystals.

Details of the contacts made between the protein and the DNA substrate are shown in Figure 3. Unlike the other PcrA/DNA complex, the thermal factors for the bound DNA are similar to those of the protein, showing that it is better ordered and indicating that the contacts are stronger. The DNA duplex region is closer to the surface of the protein in the substrate complex, suggesting that a tighter interaction is formed with duplex DNA when ADPNP is also bound to the protein, consistent with biochemical data for Rep helicase (Wong and Lohman, 1992; Wong et al., 1992). Although the first few base pairs of the bound duplex are close to the conformation of regular B-form DNA, the region from about halfway down the duplex to the junction starts to adopt a more distorted conformation (Figure 5). Finally, at the junction itself, the last few base pairs have begun to separate as the 3′ tail of the duplex twists away into the center of the protein and the 5′ end of the other strand directs a course across the outer surface of domain 2A. This strained conformation of the duplex reveals that it is binding of the duplex on the surface of the enzyme that causes unwinding of a small region of DNA at the junction. The exposed base pairs at the junction appear to be stabilized by an interaction with an exposed phenylalanine residue (F626). The unpaired 3′ tail of the DNA substrate twists around the protein and into the same groove that is occupied by the ssDNA tail in the product complex (Figure 3). There are significant differences in the binding of the ssDNA in the two complexes that suggest a mechanism for DNA translocation that is described in more detail below.

### Comparison of the Substrate and Product Complexes

As noted above, a comparison of the structures reveals that there are significant ligand-induced conformational changes that we believe to be important for the overall mechanism of the enzyme. The rigid body movements of the domains are most easily illustrated by superimposing domain 2A to act as a reference point (Figure 2). In addition to the closure of the cleft between domains 1A and 2A around the bound ADPNP, there are significant movements of domains 1B and 2B that alter their positions relative to each other. It has been shown that the ATP complex of the closely related Rep helicase has a higher affinity for duplex DNA than does the ADP complex that favors binding of single-stranded DNA (Wong and Lohman, 1992; Wong et al., 1992). The structure that we present here begins to explain this observation. It seems that the conformational changes that occur on binding ATP set up the protein surface to bind duplex DNA. These conformational changes move domains 1B and 2B into positions that form a surface that is complementary to the shape and charge of duplex DNA (Figure 5) but that is incorrectly formed prior to these conformational changes. Therefore, prior to binding of ATP, the appropriate conformational changes in domains 1B and 2B are prevented from taking place, resulting in a low affinity of this surface for duplex DNA, explaining how one direct function of the coupling of ATP binding and hydrolysis to enzyme activity is to modulate the affinity of the complex for duplex DNA, consistent with the biochemical data (Wong and Lohman, 1992).

The differences between the complexes are not confined to conformational changes of the protein; there are also some important changes in the structure of the ssDNA tail of the substrate (Figure 3). In the substrate complex, there is a region of four bases that extends across the center of the protein, but in the product complex there are five bases across the equivalent region. Although there are bases in four of the five equivalent positions, the additional base in the product complex occupies a pocket on domain 1A that is not accessible in the substrate complex. The reason for this is that in the substrate complex the side chain of a conserved Phe residue (F64) within motif Ia has rotated around the Cα–Cβ bond and is now filling the pocket. We believe that this difference between the complexes plays a crucial part in the mechanism of the enzyme (see below).

**Evidence for an Inchworm Mechanism**

The rolling model has been popularized by the extensive kinetic data that have been obtained for the closely related Rep and UvrD helicases (reviewed in Lohman and Bjornson, 1996). However, the applicability of this model as a general mechanism for helicases has been questioned (Bird et al., 1998a, 1998b; Porter et al., 1998). The problems arise from a number of different sources.

Experimental data have been amassed on the Rep helicase that seem to suggest cooperativity between subunits of a dimeric enzyme molecule (Wong and Lohman, 1992; Wong et al., 1992). Even so, it is known that the protein is a monomer in the absence of a DNA substrate (Chao and Lohman, 1991). Surprisingly, the crystal structure of the enzyme proved to be monomeric (Korolev et al., 1997) despite being complexed with a single-stranded oligonucleotide that was used previously to demonstrate dimerization in solution (Chao and Lohman, 1991). Moreover, the structure revealed that the $dT(pT)_{15}$ oligonucleotide used for these experiments was long enough to allow adjacent binding of two monomers of Rep to the same oligonucleotide, raising some doubts about the validity of cross-linking experiments that utilize this DNA ligand. That being said, other biochemical experiments, such as those that demonstrate cooperativity of the ATPase activity between Rep subunits (Wong and Lohman, 1992; Wong et al., 1992), are difficult to interpret in any way other than in terms of a Rep dimer. The experimental data are therefore confusing for Rep, but the situation is much clearer for other helicases. For the NS3 RNA helicase, ultracentrifugation experiments under a variety of conditions and of different complexes have shown that the protein is monomeric (Porter et al., 1998). For PcrA, we have repeated many of the experiments undertaken with Rep helicase but can find no evidence whatsoever for the existence of a dimer (Bird et al., 1998a). Furthermore, we have now determined five crystallographically independent crystal structures of PcrA including structures of the apo enzyme as well as complexes with ADP and now ternary complexes with DNA and nucleotides; all of these structures are monomeric (Subramanya et al., 1996; Bird et al., 1998a; this work). In addition, there are two crystallographically independent Rep structures (Korolev et al., 1997) and three crystal structures of the hepatitis virus NS3 RNA helicase (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998), all of which are monomeric. It is difficult to ignore the message conveyed by this mass of structural information from a range of different enzymes.

A second problem arises when we examine the DNA-binding properties of the protein. Work on the Rep helicase has been interpreted as showing that there are separate binding sites for ssDNA and duplex DNA on the dimer that can be occupied simultaneously but only on different subunits (Wong and Lohman, 1992), although work on PcrA helicase suggests an increased affinity of the enzyme for duplex DNA carrying single-stranded tails when compared to either duplex or single-stranded DNA (Bird et al., 1998a). The crystal structures that we present here demonstrate that the PcrA helicase monomer is competent to bind single-stranded and duplex DNA simultaneously, but only when ATP is also bound to the protein. After hydrolysis of ATP, the resulting conformational changes allow the enzyme to release the duplex DNA but retain affinity for the single-stranded DNA.

A third problem concerns the step size of the enzymes. Estimates of the step size for helicases vary from 4–5 base pairs to as small as 1 base pair (Yarranton and Gefter, 1979; Ali and Lohman, 1997; Porter et al., 1998). However, the adjacent positioning of the two Rep monomers on the 16 base oligonucleotide in the crystal structure provides an estimate of the "footprint" for binding of each molecule to single-stranded DNA of 8–10 bases (Korolev et al., 1997), a number that is in good agreement with the structure of NS3 helicase complexed with a $(dU)_8$ oligonucleotide (Kim et al., 1998) and the PcrA/DNA complexes presented herein. For the rolling model, it is evident that the step size has to be at least equal to the binding site size on the product single strand. The only way to reconcile a step size that is smaller than the binding site size on the product is to invoke slippage of the protein backwards along the DNA during each cycle of the mechanism. It is unclear how (or indeed why) this might be achieved in a controlled manner. An inchworm mechanism, on the other hand, could be consistent with any step size.

Despite these difficulties, a model has been proposed for DNA translocation by the Rep helicase based upon the crystal structure (Korolev et al., 1997). As a consequence of the observation of two conformational states of the Rep helicase bound to single-stranded DNA and the assumption that the protein is dimeric, a domain-swiveling model was proposed for DNA translocation by the enzyme. Unfortunately, there is no evidence for the proposed dimerization interface in Rep, which was suggested to be formed between the 2B domains of different subunits. Based upon sequence homologies, it has been suggested previously (Subramanya et al., 1996; Bird et al., 1998b) that helicases are likely to be modular in structure, with conservation of domains 1A and 2A and the ATP-binding site between them, either as a part of the same protein subunit or coming from different subunits. Domain 2B is not conserved, and is sometimes absent, in different helicases and is therefore unlikely to form a dimerization interface in a general model for helicase activity. Consequently, the validity of domain swiveling as a general model is doubtful particularly for helicases that are not dimeric.

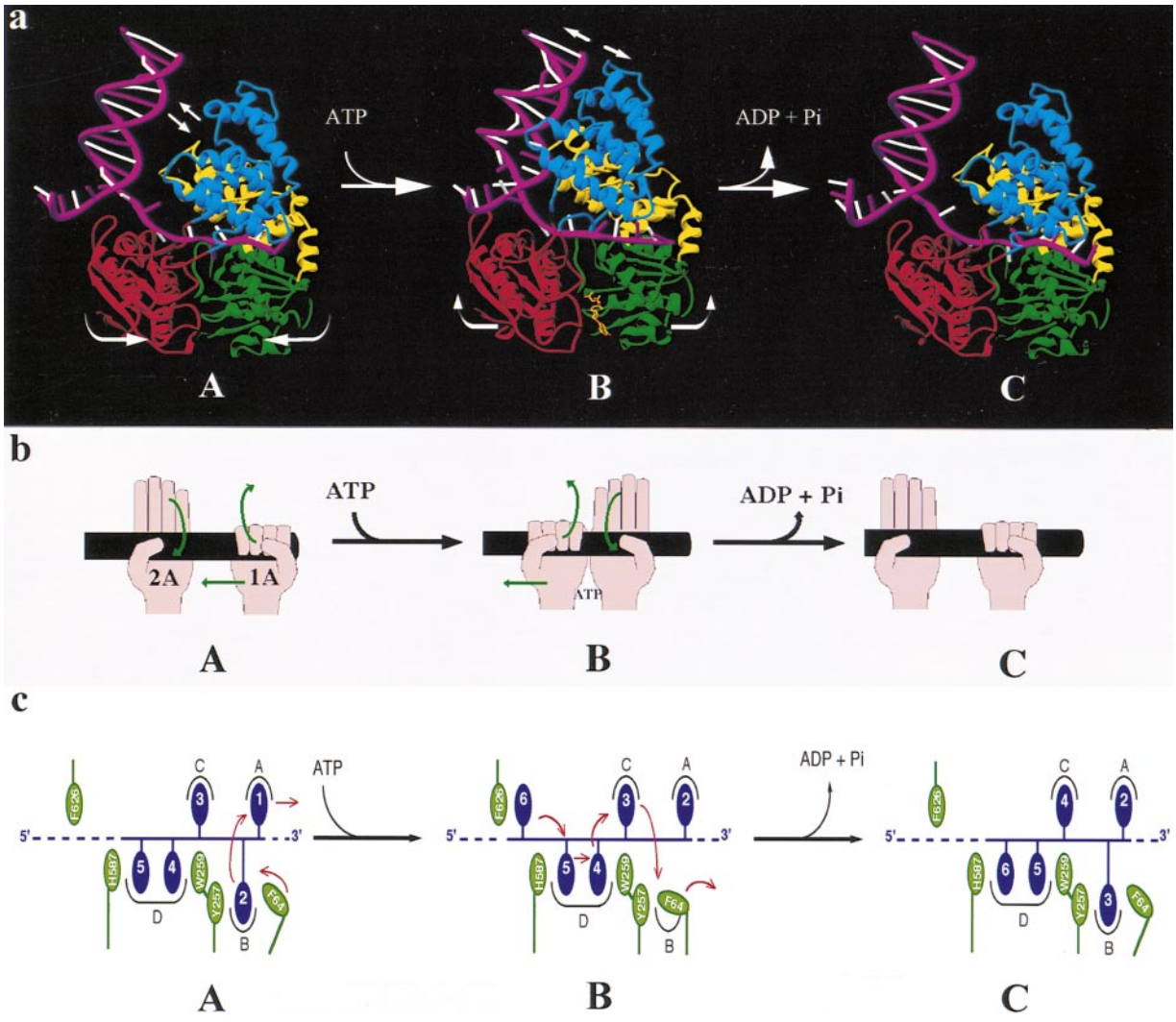


Figure 6. A Model for the Mechanism of Helicases

(a) A model for helicase activity with regard to the large conformational changes in the protein and the DNA (or RNA) substrate. The intermediates are based upon our structures, but the DNA has been extended at both ends to illustrate how a larger substrate might bind. The colors of the protein are the same as in Figure 2, but to assist in following the translocation process the base pairs have been colored alternately magenta and white. At the initial step in the reaction (A), the protein is bound to the ssDNA tail but does not bind the duplex region of the DNA. Upon binding ATP (B), there is a conformational change in the protein, and the duplex region binds to domains 1B and 2B with a concomitant unwinding of several base pairs at the junction. Finally (C), following the hydrolysis of ATP, the protein conformation returns to that in (A) as the protein translocates along the ssDNA tail by one base and releases the DNA duplex.

(b) Cartoon demonstrating the alternation in affinity for ssDNA of domains 1A and 2A during translocation. An open hand represents a loose grip on the DNA, and a closed hand is a tighter grip. (A–C) correspond to those in (a).

(c) Cartoon of the ssDNA-binding region at each stage of the reaction to illustrate the conformational changes that occur in this site as bases flip between binding pockets during translocation along single-stranded DNA. Again, (A–C) correspond to those in (a). The bases are numbered arbitrarily in the 3′ to 5′ direction.

In order to reconcile these problems with the structures that we present here, we have devised a model for a general mechanism for helicases that is consistent with most of the current data. The model that we propose (Figure 6) is based loosely on the original inchworm model proposed by Yarranton and Gefter (1979), although modified extensively from their original proposals to accommodate more recent biochemical and structural data that were not available when their model was first proposed. The model we propose can be divided into two processes, DNA translocation and duplex destabilization, and it is the coupling of these properties that gives rise to helicase activity.

DNA translocation is initiated when the protein binds to ssDNA. This induces the domain swiveling that sets up the potential for binding a DNA duplex. This is what we believe has been trapped in our product complex and in the closed molecule in the Rep crystal structure (Korolev et al., 1997). At this initial stage, the ssDNA is bound to both domains 1A and 2A. However, when ATP

binds to the complex, the cleft between domains 1A and 2A closes. In order for this to happen while retaining a hold on the ssDNA, it is evident that one of the domains must release its grip on the DNA and slide along it. Comparison of the structures suggests how this might take place (discussed below). The result of the cleft closure is the structure that we see as our substrate complex. The bound ATP is now hydrolyzed to release the protein from this conformational state. Hydrolysis results in a destabilization of the cleft closure because contacts mediated through the $\gamma$ phosphate are now broken. As the cleft opens, domain 2A weakens its hold on the ssDNA. Translocation along the DNA is now effected across domain 2A as the cleft springs open again with domain 1A retaining a tight grip on the DNA and pulling the ssDNA across the surface of domain 2A. This mechanism explains how ATP-dependent unidirectional translocation can be accomplished by helicases, but it does not explain how the helix is destabilized. The crystal structures, however, also suggest a mechanism for this process.

The ATP-dependent DNA translocation presumably occurs until the helicase meets a duplex. At this point, when the protein binds ATP and the cleft closes, the protein has an increased affinity for double-stranded DNA and now binds to the duplex region adjacent to the fork. The duplex region of the DNA is pulled onto the surface presented by domains 1B and 2B, creating strain in the base pairing at the fork as the DNA substrate is bent across the surface of the protein. The role of the exposed F626 appears to be to stack with the DNA at the fork, thereby helping to stabilize the unwound form of the DNA. The structure of the duplex DNA in the substrate complex shows that the duplex is regular B-form DNA to begin with but becomes distorted when closer to the junction, with 4–5 base pairs near the junction having been destabilized and with the strands already beginning to separate. Thus, binding energy is used to create additional single-stranded DNA along which the helicase can move, and consequently, helix destabilization is coupled to DNA translocation to give the combined activity that is characteristic of a helicase. The free energy of hydrolysis of ATP is therefore utilized by the enzyme in two ways, for unidirectional translocation and for strand separation, both of which are energetically unfavorable processes.

### Molecular Details of the DNA Translocation Process

A central aspect of the above mechanism is that the ssDNA must slide across the surface of domains 1A and 2A separately at different steps in the cycle. A close examination of the differences in conformation of the ssDNA bound to the protein in each of the complexes suggests a mechanism for this process (Figure 6). At the initial step in the cycle, which we presume to be most similar to the product complex, domain 1A has a firm grip on the DNA with bases in all of the acceptor pockets (A, B, and C) in this domain. The complex then binds ATP, and a number of conformational changes take place as the cleft between the domains closes. First, the hold of domain 1A on the ssDNA has to be released to allow translocation across this domain. This is accomplished as the side chain of F64 moves into pocket B, thereby displacing the base within it to pocket A, while the base formerly in pocket A is pushed outside the protein as the DNA slides across the surface of domain 1A, resulting in the structure represented by the substrate complex. At this point in the cycle, translocation of the ssDNA has taken place, but only across domain 1A. The next stage of the reaction follows ATP hydrolysis and accompanies the subsequent reopening of the cleft between domains 1A and 2A. ATP hydrolysis releases the F64 side chain from pocket B. The base in pocket C, above the conserved W259, is then able to flip into pocket B, with the result that domain 1A now has a tighter grip on the ssDNA, allowing the ssDNA to be pulled over the surface of domain 2A as the cleft opens. This movement forces a base to flip from the stacked pair in pocket D on domain 2A into pocket C. At the same time, the other base of the stacked pair in pocket D moves along one position as the next base along flips from being base stacked with F626 and moves into the pocket. Translocation, therefore, results from a wave of base flipping moving along the bound ssDNA tail with the power stroke being the relaxation of the protein as the cleft opens. Although this simple mechanism is based upon just the two conformational states of the protein and bound DNA that we have crystallized, we would not wish to rule out additional conformations at other stages in the catalytic cycle, such as that after phosphate release, but this mechanism does at least provide a framework for the overall process.

The mechanism that we propose implies a step size of one base for the translocation process as the bases flip between adjacent pockets along the ssDNA-binding site. Although we believe this to be the most likely mechanism, it is possible that each unwinding step could unwind more than 1 base pair from the duplex (the structure actually suggests that as many as 4 or 5 base pairs are separated at the junction when the duplex binds to the protein), and it is not necessarily the case that every translocation step would have to be coupled to the DNA unwinding phase of the reaction. Such a proposal could explain the estimates of 4–5 base pairs for the step size in unwinding of DNA by UvrD helicase (Ali and Lohman, 1997) or 2 base pairs of RNA for NS3 helicase (Porter et al., 1998), yet still be consistent with a step size of one base for translocation. The observation that the ATPase properties of the PriA helicase are different for the translocation and helicase activities of the protein (Lee and Marians, 1990) might be evidence in favor of such a scheme.

### Implications of the Model

Our model has several major implications for understanding the mechanism of helicases. First, helicase activity is an active rather than passive process. Some proposals for a mechanism for helicases suggest that they operate by simply translocating along ssDNA as it is produced by a transient “fraying” of the base pairs at the fork (Kim et al., 1998). Such a mechanism requires no destabilization of the duplex region of the substrate or even an interaction with it, such as that demonstrated in our substrate complex, and it seems intrinsically unlikely that the free energy of hydrolysis of ATP is not

utilized by the enzyme to facilitate the energetically unfavorable strand separation. In fact, Rep and UvrD helicases have been shown to operate by an active mechanism (Lohman and Bjornson, 1996). A second feature of our model is that all of the components that are required for DNA translocation are contained within domains 1A and 2A, whose sequences include motifs that are conserved across a wide range of DNA and RNA helicases (Gorbalenya and Koonin, 1993). Coupling of this translocation activity to helix destabilization and/or substrate specificity in order to produce helicase activity would be conferred by the much more variable domains 1B and 2B, suggesting that our model could be a general model for helicases. Moreover, in addition to a mechanism for helicases, our model provides a mechanism for the ATP-dependent translocation along DNA that is a property not only of helicases (Lee and Marians, 1990) but also other enzymes that have the conserved helicase sequence motifs (e.g., type I and type III restriction endonucleases [Gorbalenya and Koonin, 1991]). Another important aspect is that this inchworm mechanism could be consistent with any of the estimates of the step size for unwinding of DNA by helicases. Fourth, domain swiveling serves to create the binding site for duplex DNA once the protein has bound to ssDNA, rather than being an intrinsic part of the translocation mechanism. In this way, the protein is prevented from interacting with duplex DNA until it is activated by the presence of single-stranded DNA and is therefore unable to initiate strand separation from within a sealed duplex, an activity that would be highly undesirable in a cell. Finally, each helicase monomer is perfectly competent to bind both ssDNA and dsDNA at the same time, which negates the requirement for the enzyme to be oligomeric to be able to function as a helicase. We note that a dimeric enzyme could still function by this mechanism, but each subunit would have an intrinsic ability to function independently as a helicase. Although formally possible, it seems unlikely that two enzymes with such closely related structures as those of Rep and PcrA will operate by vastly different mechanisms. One pleasing aspect of the model that we present here is that it is equally applicable to each of the subunits of a dimeric Rep enzyme, with the simple requisite that dimerization is not essential for the mechanism per se, although we would not wish to rule out a requirement for cooperativity between the subunits that could relate, for example, to a need for coordinated progression of two forks on the same (or different) DNA substrates. However, until a subunit interface in Rep has been identified, it is inappropriate to speculate further about any potential advantage that dimerization might confer.

Although this model improves our understanding of the mechanism of helicases, it does not, of course, provide detailed answers to all of the questions. For example, we still do not understand the molecular details of how other proteins modulate the activity of helicases or how helicases function within larger macromolecular complexes such as the primosome. Finally, although we note that even hexameric helicases could adopt a mechanism similar to that which we propose here, with individual subunits taking the place of domains 1A and 2A, evidence in support of this suggestion awaits a crystal structure.

Table 1. Crystallographic Statistics

| | Product Complex | Substrate Complex |
|---|---|---|
| Space group | $P2_1$ | $P6_122$ |
| Resolution (Å) | 15–2.9 | 15–3.3 |
| Completeness (%) | 97.9 | 92.0 |
| $R_{sym}$ (%) | 6.8 | 6.0 |
| R factor (%) | 24.0 | 23.6 |
| $R_{free}$ (%) | 29.7 | 31.5 |
| Rmsd bond length (Å) | 0.011 | 0.014 |
| Rmsd bond angle (°) | 1.69 | 2.60 |

## Experimental Procedures

### Structure Determination of the PcrA/DNA/Sulphate Complex

In the original apo PcrA structure, residues 651–724 were disordered. In addition, because this region is absent in the closely related Rep helicase (Korolev et al., 1997), we presumed that it was likely to be dispensable for activity. We therefore produced a truncated version of PcrA that was engineered to terminate at residue 650. Truncated protein was prepared as described previously for the full-length protein (Bird et al., 1998a). The truncated protein was assayed for helicase and ATPase activities, which were the same as those obtained for the full-length protein (data not shown).

Synthetic oligonucleotides, a 10-mer (5′-CGAGCACTGC-3′) and a 17-mer (5′-GCAGTGCTCGTTTTTTT-3′), and the tailed heteroduplex formed between them were purified by ion exchange chromatography. The 5-iodoU oligonucleotides were purchased from Oswel (UK) and purified in the dark. A 5-iodoU modification was made in the 10-mer (5′-CGAGCACXGC-3′) at the position marked with an X.

The complex of PcrA and the DNA substrate was prepared by incubating PcrA (90 μM) and DNA (166 μM) in 37 mM Tris (pH 7.5), 1.48 mM EDTA, 1.74 mM DTT, 148 mM NaCl, 3.1% glucose, 4.9 mM $MgCl_2$, and 7.4% glycerol for 1 hr. Crystals were grown using the hanging drop vapor diffusion method above wells containing 100 mM Tris (pH 8.5), 50 mM $LiSO_4$, 20% PEG 400, and 20%–23% PEG 4000. The protein–DNA solution was mixed 1:1 with well solution, and microseeds were added immediately in an additional small volume of well solution. Crystallization showed an absolute requirement for sulphate ions. Crystals were of the monoclinic space group $P2_1$ with unit cell dimensions a = 85.0 Å, b = 62.6 Å, c = 141.8 Å, β = 95.8°. There are two PcrA/DNA complexes in the asymmetric unit. Data were collected from flash-frozen crystals at 100 K and processed using the HKL programs DENZO and SCALEPACK (Otwinowski and Minor, 1997) (Table 1). The structure was solved by molecular replacement using the CCP4 programs ALMN and TFFC (Collaborative Computing Project No. 4, 1994) with the apo PcrA coordinates as a model. The initial model used just domains 1A and 2A. Domains 1B and 2B had to be located separately because of the large conformational changes between the two structures. The validity of the structure was confirmed by location of selenium positions using data collected from crystals of selenomethionine-derivatized protein (data not shown). Rigid body refinement was carried out in CNS (Brünger et al., 1998). Initial positional refinement was carried out using the maximum entropy method as implemented in CNS and included a bulk solvent correction. NCS restraints were applied separately to each of the individual domains during refinement. Rounds of manual model building were undertaken between refinement cycles using TurboFrodo (Roussel and Cambillau, 1989), but the DNA was omitted until the final stages of refinement. Statistics concerning the quality of the final model are presented in Table 1. The coordinates for the product complex have been deposited at the Protein Data Bank under ID code 2PJR.

### Structure Determination of the PcrA/DNA/ADPNP Complex

Crystals of the PcrA/DNA/ADPNP complex were obtained using full-length PcrA protein, prepared as described previously (Bird et al., 1998a). PcrA (83–100 μM) was incubated for 1 hr at room temperature with the DNA substrate (175 μM) in 4.8% glucose, 5 mM $MgCl_2$, 2.5 mM ADPNP, 2 mM DTT, 26.5 mM Tris (pH 7.5), 1 mM EDTA,

and 106 mM NaCl. Crystals were obtained using the hanging drop method by mixing protein solution 1:1 with well solution and improved by microseeding. The wells contained 100 mM Tris (pH 8.5), 50 mM $LiSO_4$, and 12%–20% PEG 4000. Crystallization showed an absolute requirement for ADPNP, but sulphate ions were dispensable. Crystals were of the hexagonal space group $P6_122$, with unit cell dimensions a = b = 105.1 Å, c = 380.0 Å and contained one PcrA/DNA/ADPNP complex in the asymmetric unit. Conditions to freeze crystals could not be found, so data were collected from several conventionally mounted crystals at room temperature. Typically, five degrees of data could be collected from each crystal before radiation damage became too severe. Data were processed and merged using the HKL programs DENZO and SCALEPACK (Otwinowski and Minor, 1997) (Table 1). The structure was solved by molecular replacement using the CCP4 programs ALMN and TFFC (Collaborative Computing Project No. 4, 1994). The initial model used just domains 1A and 1B from the monoclinic PcrA/DNA complex; domains 2A and 2B had to be located separately because of conformational changes compared to the other PcrA/DNA complex and the apo enzyme. Rigid body refinement was carried out in CNS (Brünger et al., 1998). Positional refinement was carried out using the CCP4 program REFMAC and included a bulk solvent correction. Rounds of manual model building and refinement were carried out to improve the model, but DNA and ADPNP were not included until the very final stages of refinement. The correct register of the bases in the DNA was checked using crystals grown using the 5-iodoU-substituted DNA described above (data not shown). In common with our structure of the apo enzyme (Subramanya et al., 1996), the last 70 residues of the protein were disordered in the final structure and have not been modeled. Statistics concerning the quality of the final model are presented in Table 1. The coordinates for the substrate complex have been deposited at the Protein Data Bank under ID code 3PJR.

**Acknowledgments**

We thank Eleanor Dodson for advice on the use of REFMAC; Steve Smerdon for advice on purification of iodo-substituted DNA; Val Cooper for oligonucleotide synthesis; Jennifer Byrne for technical assistance; and the Daresbury and Hamburg synchrotrons (including support from the TMR/LSF program to the EMBL Hamburg Outstation—ERBFMGECT980134) for access to data collection facilities. This project was supported by the Wellcome Trust.

Received February 11, 1999; revised March 12, 1999.

**References**

Ali, J.A., and Lohman, T.M. (1997). Kinetic measurement of the step size of DNA unwinding by *Escherichia coli* UvrD helicase. Science *275*, 377–380.

Bird, L.E., Brannigan, J.A., Subramanya, H.S., and Wigley, D.B. (1998a). Characterisation of *Bacillus stearothermophilus* PcrA helicase: evidence against an active rolling mechanism. Nucleic Acids Res. *26*, 2686–2693.

Bird, L.E., Subramanya, H.S., and Wigley, D.B. (1998b). Helicases: a unifying structural theme? Curr. Opin. Struct. Biol. *8*, 14–18.

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 and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 904–925.

Carson, M. (1991). Ribbons 2.0. J. Appl. Crystallogr. *24*, 958–961.

Chao, K.L., and Lohman, T.M. (1991). DNA-induced dimerisation of the *Escherichia coli* Rep helicase. J. Mol. Biol. *221*, 1165–1181.

Cho, H-S., Ha, N-C., Kang, L-W., Chung, K.M., Back, S.H., Jang, S.K., and Oh, B-H. (1998). Crystal structure of RNA helicase from genotype 1b hepatitis C virus: a feasible mechanism for unwinding duplex RNA. J. Biol. Chem. *273*, 15045–15052.

Collaborative Computing Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.

Gorbalenya, A.E., and Koonin, E.V. (1991). Endonuclease (R) subunits of type-I and type-III restriction modification enzymes contain a helicase-like domain. FEBS Lett. *291*, 277–281.

Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. *3*, 419–429.

Iordanescu, S. (1993). Characterisation of the *Staphylococcus aureus* chromosomal gene pcrA, identified by mutations affecting plasmid pT181 replication. Mol. Gen. Genet. *241*, 185–192.

Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C., and Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure *6*, 89–100.

Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G. (1997). Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. Cell *90*, 635–647.

Lee, M.S., and Marians, K.J. (1990). Differential ATP requirements distinguish the DNA translocation and DNA unwinding activities of the *Escherichia coli* PriA protein. J. Biol. Chem. *265*, 17078–17083.

Lohman, T.M., and Bjornson, K.P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. Annu. Rev. Biochem. *65*, 169–214.

Nicholls, A., and Honig, B.J. (1991). A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J. Comput. Chem. *12*, 435–445.

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

Petit, M-A., Dervyn, E., Rose, M., Entian, K-D., McGovern, S., Ehrlich, S.D., and Bruand, C. (1998). PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. Mol. Microbiol. *29*, 261–274.

Porter, D.J.T., Short, S.A., Hanlon, M.H., Preugschat, F., Wilson, J.E., Willard, D.H., and Consler, T.G. (1998). Product release is the major contributor to kcat for the hepatitis virus helicase-catalyzed strand separation of short duplex DNA. J. Biol. Chem. *273*, 18906–18914.

Roussel, A., and Cambillau, C. (1989). TURBO-FRODO. In Silicon Graphics Geometry Partner Directory (Mountain View, CA: Silicon Graphics), pp. 77–78.

Soultanas, P., Dillingham, M.S., Papadopoulos, F., Phillips, S.E.V., Thomas, C.D., and Wigley, D.B. (1999). Plasmid replication-initiator protein RepD increases the processivity of PcrA DNA helicase. Nucleic Acids Res. *27*, 1421–1428.

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

Wong, I., and Lohman, T.M. (1992). Allosteric effects of nucleotide cofactors on *Escherichia coli* Rep helicase-DNA binding. Science *256*, 350–355.

Wong, I., Chao, K.L., Bujalowski, W., and Lohman, T.M. (1992). DNA-induced dimerisation of the *Escherichia coli* Rep helicase: allosteric effects of single stranded and duplex DNA. J. Biol. Chem. *267*, 7596–7610.

Yao, N.H., 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.

Yarranton, G.T., and Gefter, M.L. (1979). Enzyme-catalyzed DNA unwinding: studies on *Escherichia coli* rep protein. Proc. Natl. Acad. Sci. USA *76*, 1658–1662.

**Protein Data Bank ID Codes**

The product and substrate complexes in this paper have been deposited in the Protein Data Bank under ID codes 2PJR and 3PJR, respectively.