J Mol Model (2013) 19:4997–5006
DOI 10.1007/s00894-013-2008-4

ORIGINAL PAPER

# The study of interactions between DNA and PcrA DNA helicase by using targeted molecular dynamic simulations

**Hao Wang · Jiajia Cui · Wei Hong · Ian C. Paterson · Charles A. Laughton**

Received: 16 April 2013 / Accepted: 8 September 2013 / Published online: 26 September 2013


**Abstract** DNA helicases are important enzymes involved in all aspects of nucleic acid metabolism, ranging from DNA replication and repair to recombination, rescue of stalled replication and translation. DNA helicases are molecular motors. Through conformational changes caused by ATP hydrolysis and binding, they move along the template double helix, break the hydrogen bonds between the two strands and separate the template chains, so that the genetic information can be accessed. In this paper, targeted molecular dynamic simulations were performed to study the important interactions between DNA and PcrA DNA helicase, which can not be observed from the crystal structures. The key residues on PcrA DNA helicase that have strong interactions with both double stranded DNA (ds-DNA) and single stranded DNA (ss-DNA) have been identified, and it was found that such interactions mostly exist between the protein and DNA backbone, which indicates that the translocation of PcrA is independent of the DNA sequence. The simulations indicate that the ds-DNA is separated upon ATP rebinding, rather than ATP hydrolysis, which suggests that the two strokes in the mechanism have two different major roles. Firstly, in the power stroke (ATP hydrolysis), most of the translocations of the bases from one pocket to the next occur. In the relaxation stroke (ATP binding), most of the 'work' is being done to 'melt' the DNA at the separation fork. Therefore, we propose a mechanism whereby the translocation of the ss-DNA is powered by ATP hydrolysis and the separation of the ds-DNA is powered by ATP binding.

**Keywords** PcrA DNA helicase · Targeted molecular dynamic simulation

H. Wang (✉) · J. Cui
School of Pharmacy, Ningxia Medical University, 1160 Shengli Street, Yinchuan, Ningxia 750004, People's Republic of China
e-mail: paxhw@yahoo.co.uk

W. Hong
School of Chemistry and Chemical Engineering, Beifang University of Nationalities, Yinchuan, Ningxia 750021, People's Republic of China

I. C. Paterson
Dental Research and Training Unit and Oral Cancer Research Coordinating Centre, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia

C. A. Laughton
School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK

## Introduction

DNA helicases are important molecular motors that, powered by ATP hydrolysis and binding, unwind DNA duplexes by moving along the template double helix and breaking the hydrogen bonds between the two strands. They are involved in many processes of DNA metabolism, such as replication, recombination, repair and transcription [1, 2]. Defects in helicase function in one or more of these processes can result in human genetic disorders and predisposition to cancer [3], and it has been reported that the defects of different helicases can cause a number of human syndromes [4]. Since helicases are important in most aspects of nucleic acid metabolism and they play prominent roles in human disease, understanding the function and mechanism of this class of enzymes is an attractive target for study [5]. Based on the direction of separation of the ds-DNA, DNA helicases can be divided into two types, namely, those that unwind the ds-DNA in 3′ to 5′ direction and those that unwind in 5′ to 3′ direction.

The first helicase structure, solved and published in 1996 [6, 7], was that of the *Bacillus stearothermophilus* PcrA DNA helicase, which belongs to superfamily 1 (SF-1) and displays a 3′ to 5′ polarity in DNA unwinding [8, 9]. The translocation speed of PcrA along the ss-DNA is around 50 nucleotides (nt)

per second, consuming one ATP per 1-nt distance [10, 11]. PcrA has been proposed to work as a monomer [7] and contains four subdomains termed 1A, 1B, 2A, and 2B. The ATP binding site is located in the cleft between subdomains 1A and 2A. Three structures of PcrA have been solved by X-ray crystallography [6, 7] and these are the apo-form (PDB code 1PJR), the substrate complex form (Fig. 1a, PDB code 3PJR) and the product complex form (Fig. 1a, PDB code 2PJR). Before binding to DNA, the PcrA helicase is in the apo-form. The binding of ss-DNA on the top of subdomains 1A and 2A induces a rigid-body rotation of subdomain 2B by 130°. In molecular modeling studies, it was found that the flexibility of the loop between I541 and A558 is important for this rotation [12]. Then, the binding of ATP in the active site leads to the closure of this cleft changing the relative positions of subdomains 1B and 2B, which causes the formation of a surface complementary to the shape and charge of the ds-DNA [12, 13]. Now the conformation of PcrA changes from the apo-form to the substrate complex. In the substrate complex, a magnesium ion binds in the position between ATP and subdomain 1A [14], and no additional protein–protein interactions between subdomains 1A and 2A across the cleft are formed. In the crystal structure of the substrate complex (resolution: 3.30 Å), ten base pairs in the duplex region and five bases of the seven in the single strand region have been identified. The other two bases on the tail of the ss-DNA could not be recognized because of the very weak electron density. A molecular dynamic simulation predicted that the channel formed by THR91, SER94, LEU540, ASN66, and THR65 of PcrA could be the way for the ss-DNA to pass to the outside [15].

After the ATP in the binding site is hydrolyzed, the ADP and $\gamma$ phosphate group will leave the binding site. This process drives the change in conformation from the substrate complex to the product complex. One new ATP molecule will rebind to the product complex, and as a consequence the conformation of the product complex will change back to the substrate complex. Through this unwinding cycle, one ds-DNA base pair has been separated and the ss-DNA moves one base forward. It is now generally accepted that the five T bases at the 3′ tail of the ss-DNA bind in a groove along the top of subdomains 1A and 2A, and translocate via three different pockets (pocket 3 (composed by PHE626 and HIS587), pocket 2 (HIS587 and TRP259), and pocket 1 (TYR257 and PHE64)) on the surface of PcrA [16]. Over four sequential unwinding cycles, each DNA base translocates across the surface of PcrA by passing from pocket 3, to pocket 2, to pocket 1 sequentially. In the crystal structure of the substrate complex, five thymine bases occupy the ss-DNA binding cleft. T5 is outside pocket 3, T3 and T4 occupy pocket 2, and T2 lies above pocket 1. When the conformation changes to the product complex, T5 drops into pocket 2 and takes the position originally occupied by T4, T4 moves forward to take T3's position, T3 moves to the T2's position, and T2 drops into pocket 1 (Fig. 1b). If we consider the substrate complex and take into account that each base must move forward by one position (so T5 becomes T4, T4 becomes T3 etc.), it can be seen that when ATP rebinds and the conformation changes back from the product complex to the substrate complex, T2 moves out of pocket 1 and, although the other bases shift, they remain within the same pockets [7].

During ss-DNA translocation, extensive rotation of the ss-DNA backbone is involved, so the backbone flexibility is thought to be important for PcrA unidirectional translocation. Based on this assumption, it was reported that

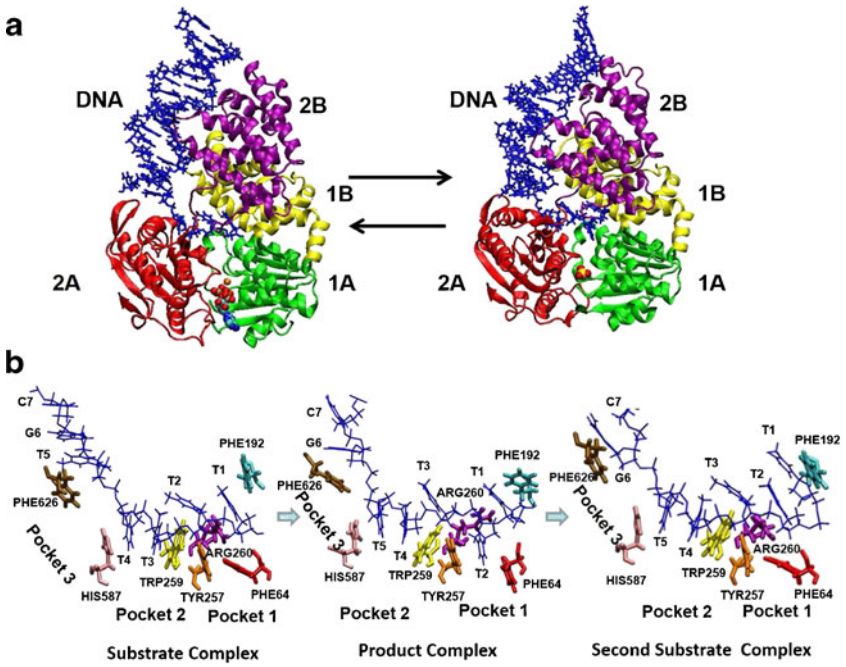


**Fig. 1** **a** The unwinding cycle of the PcrA DNA helicase. Subdomain 1A is colored *green*, subdomain 1B *yellow*, subdomain 2A *red*, and subdomain 2B *purple*. The bound DNA is colored *blue*. The ATP is colored by atom type and the magnesium ion is in *orange*. **b** DNA bases translocation via the binding pockets. This figure was adapted from reference [7]

vinylphosphonate-linked (T*T) dimers were designed and synthesized [17, 18] to restrict the free rotation of the DNA backbone. It was reported that the presence of a single T*T dimer in the translocating strand (3′-5′ chain) is sufficient to inhibit significantly the helicase activity of PcrA as long as it is situated either at the ss-ds-DNA junction or within the duplex, while four consecutive modifications inhibit it completely [17]. It was suggested by molecular modeling that the vinylphosphonate-linked (T*T) dimer has different effects on the energetics of DNA translocation through the protein as it reaches different sub-sites [19].

As described above, much effort has been directed toward understanding the PcrA reaction mechanism. However, the detailed conformational changes and interactions between DNA and PcrA are still unclear. In the present study, targeted molecular dynamics simulations were performed to drive the conformation changing between the substrate and product complexes in one unwinding cycle, whilst leaving sufficient freedom for the system to relax. Along the reaction path generated by the simulations, we present detailed structural analysis which cannot be observed based only on the crystal structures.

## Methods

### Targeted molecular dynamics

By continuously decreasing the root mean square distance (RMSD) to the target structure, targeted molecular dynamics (TMD) restrains the system to change the conformation from its initial state to the final conformation so that the reaction pathway can be achieved [20–25]. In AMBER, the TMD option adds an additional term to the energy function based on the mass-weighted RMSD of a set of atoms in the current structure compared to a reference structure. Below is the function calculating the TMD energy term using the AMBER forcefield [25].

$$E = 0.5 \times k \times N \times (R_c - R_t)^2, \quad (1)$$

where $k$ is the force constant added on the system which can be set by users. $N$ is the number of atoms which are constrained by TMD. $R_c$ is the mass weighted RMSD of current structure during the simulation with respect to the target structure. The $R_t$ is the RMSD value which the user wants to achieve at the end of simulation. By using the weight change technique [25], $R_t$ can decrease automatically so that the simulation can be constrained to find a path from the start to the end continuously.

The advantage of TMD is that the constraint potential on each atom could be quite small, so it allows flexibility of the system while it approaches the target structure [26]. TMD can simulate the reaction continuously, so the transition points can be found. The disadvantage of TMD is that it forces the reaction to follow the direction of decreasing RMSD, therefore large energy barriers inaccessible to the unforced system could be crossed when using TMD [27]. This may lead the simulation to follow an unrealistic pathway. Therefore, the careful choice of the force constant and relatively long simulations are always necessary.

### System setup

The original crystal structures of PcrA—the substrate complex (3PJR) and the product complexes (2PJR), were obtained from the Protein Data Bank. As the missing residues in the protein components of the crystal structures are not located in the DNA active site, if a residue was missing in one of these two structures, it was also deleted from all other ones. The missing DNA bases in the product complex were completed by transferring coordinates from restrained molecular dynamic simulations of the (more complete) substrate complex targeted at the product complex conformation. In order to simulate the second step in the unwinding cycle when the conformation changes back from that seen in the product complex to that corresponding to the substrate complex, a new model, called the second substrate complex, was created with shifted DNA bases. Finally, in order to show the whole interaction between the DNA and protein, the ds-DNA was extended by 5 base pairs in these three conformations [13]. As only the conformation of the protein and DNA is of interest and targeted molecular dynamics requires the start and target structures contain exactly the same topology, the bound ATP, magnesium ion and $\gamma$ phosphate group were removed. The final system for each of these three complexes contains 11,439 atoms, of which 637 residues are in the protein, 15 bases are in the 5′-3′ chain (short chain) of the DNA, and 20 bases are in the 3′-5′ chain (long chain). All three structures were explicitly solvated in a truncated octahedral box (at least 10 Å from complex to avoid periodic artifacts from occurring) of TIP3P model water (25,419 water molecules), and 52 $Na^+$ ions were added to neutralize the charges of each system under Amber-03 force field by using ‘addions’ command line with tleap module, which adds counterions around the complex using a Coulombic potential on a grid. The three structures (each one contains 87,619 atoms in total) were then optimized by energy minimizations and molecular dynamics simulations with our standard equilibration strategy [28] using Amber 8 package. For further details about the system setup, please refer to the Supplementary data of our previous publication [19].

### The protocol of targeted molecular dynamics

Our standard equilibration strategy is listed as follows:

1. The solvent is energy minimized by 50 steps of the steepest descent method and then followed by 10,000

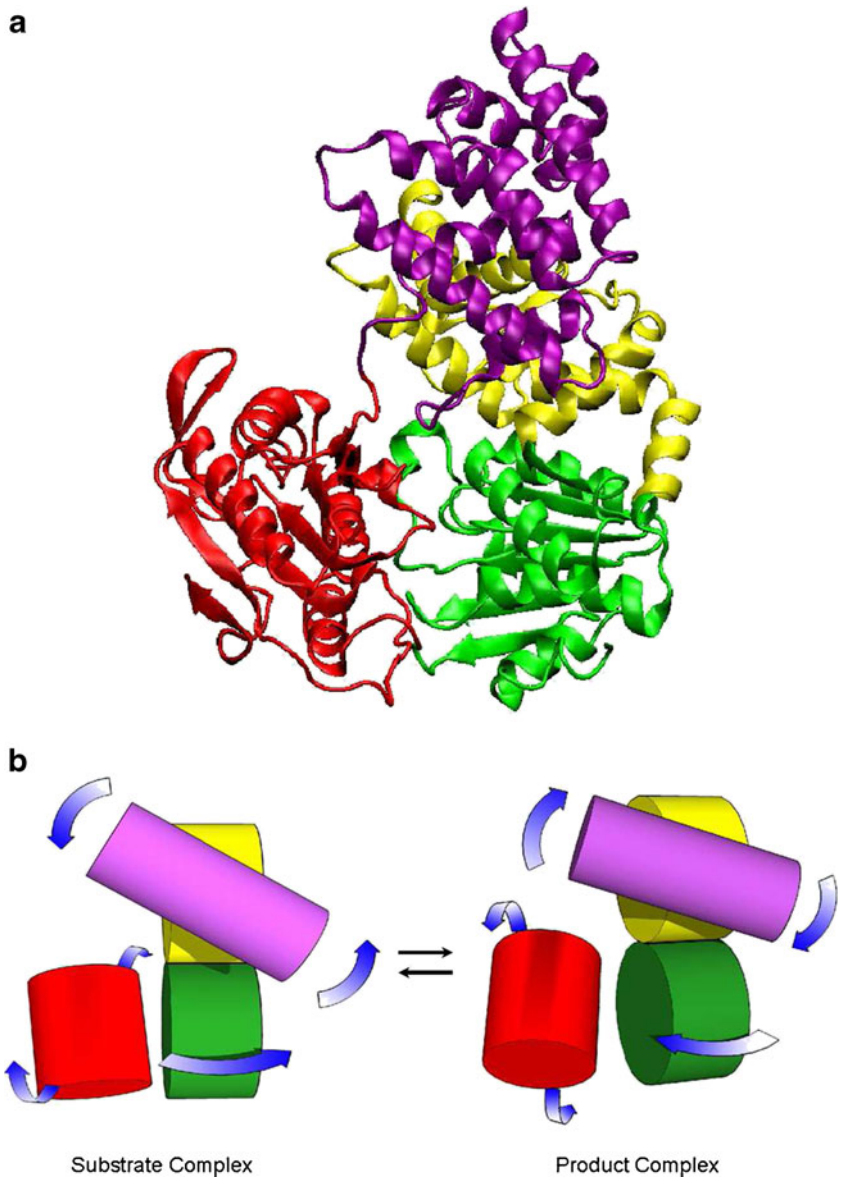


**Fig. 2** **a** The structure of the protein component of PcrA in the substrate complex. Subdomain 1A is colored in *green*, subdomain 1B in *yellow*, subdomain 2A in *red*, and subdomain 2B in *purple*. The ATP binding site is located in the cleft between subdomains 1A and 2A. **b** The cartoon diagram to show the major motions of the four subdomains

steps of the conjugate gradient method with the default nonbonded cutoff of 8 Å, while the solute is held fixed. For this and all following energy minimizations, minimization needs not be exhaustive, that a target RMS (root-mean-square) gradient of $0.1\,kcal\ mol^{-1}$ Å $^{-1}$ is sufficient.

2. The water molecules and counterions are minimized by 50 steps of the steepest descent method and then followed by 10,000 steps of the conjugate gradient method, while the solute is held fixed.
3. The entire system is energy minimized by 50 steps of the steepest descent method and then followed by 10,000 steps of the conjugate gradient method.
4. The solvent is subjected to a short (20 ps) MD simulation at a temperature of 100 K, under constant pressure conditions. The purpose of this is to remove voids in the solvent. The success of this stage is easily monitored by observing the change in the reported density of the system. For this and all following simulations, MD simulations are performed with explicit solvent models and in the NPT ensemble ($T$=300 K; $P$=1 atm). Periodic boundary conditions (PBC) and particle-mesh-Ewald method (PME) [29] are used to model long-range electrostatic effects, while the temperature is coupled to an external bath using a weak coupling algorithm [30]. The cutoff non-bonded interaction is set as 8 Å. The bond interactions involving H-atoms are constrained by using the SHAKE algorithm.
5. Over 20 ps, the solvent temperature is raised to 300 K. During both this phase and the last, position restraints on every solute atom (force constant $100\,kcal\ mol^{-1}$/Å$^2$) maintain it in its energy-minimized conformation.

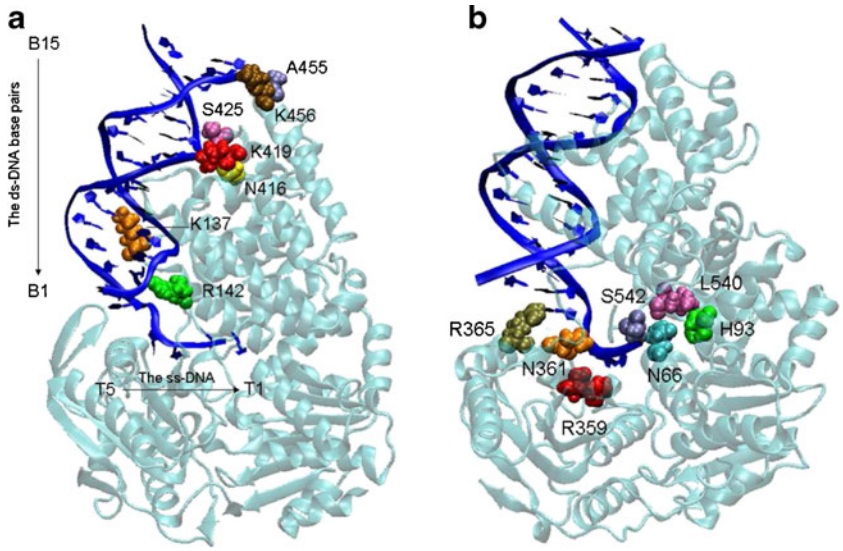


**Fig. 3** **a** The H-bond interactions between the protein and ds-DNA. Fifteen DNA base pairs along the ds-DNA are numbered from B15 to B1 in order from the tail of the ds-DNA to the separation fork. Five T bases along the ss-DNA are numbered from T5 to T1 from the separation fork to the 3′ end of the ss-DNA. Five residues in subdomain 2B of PcrA have strong H-bond interactions with the ds-DNA, in which ALA455 is in *iceblue*, LYS456 is in *ochre*, SER425 is in *mauve*, ASN416 is in *yellow* and LYS419 is in *red*. At the separation fork, LYS137 (in *orange*) and ARG142 (in *green*) have interactions with the 3′-5′ chain of the DNA. **b**, The H-bond interactions between the protein and ss-DNA. ARG365 (in *tan*) has strong H-bond interactions with the 5′ end of the 5′-3′ chain of the DNA. Six residues have strong H-bond interactions with the ss-DNA in the 3′-5′ chain, in which ARG359 is in *red*, ASN361 is in *orange*, SER542 is in *iceblue*, ASN66 is in *cyan*, LEU540 is in *mauve* and HIS93 is in *green*

6. Over a series of 20 ps constant-pressure simulations at 300 K, the restraints on the solute are gradually relaxed (100, 50, 25, 10, 5, 2 and then 1 $kcal\ mol^{-1}/Å^2$ respectively).
7. Over 300 ps, the entire system is subjected to an unrestrained MD simulation at 300 K.

After in total 480 ps equilibrations, TMD was first run for 10 ns to drive the conformation of the substrate complex (starting structure) to change to the one of the product complex (target structure). In this simulation, the force constant for TMD was set at 3.2 $kcal\ mol^{-1}/Å^2$, and the weight change technique was combined to constantly reduce the target RMSD from 3.42 Å to zero with respect to the product complex by 5,000,000 steps, so that the simulation was directed in the RMSD reducing direction automatically. Since some of the DNA bases were rebuilt, only the atoms contained in both the crystal structures of the substrate and the product complexes, 10,712 atoms, were restrained by TMD at constant temperature and pressure ($T$=300 K; $P$=1 atm.). The time step necessary to solve Newton's equations is chosen to be equal to 2 fs. The trajectory coordinates were recorded every 1 ps.

Once the TMD simulation from the substrate complex to the product complex was completed, the last snapshot, which can be treated as the product complex and whose RMSD is 3.39 Å with respect to the second substrate complex, was used as the input conformation for another TMD simulation from the product to the second substrate complex with the same protocol.

## Results

### The global conformational changes of PcrA

To understand the global conformational changes of PcrA along the trajectory, principal component analysis (PCA) was performed and the projections on the first eigenvectors were analyzed. It was observed that the major motion of PcrA is illustrated by the movements of the four rigid-body subdomains. In the power stroke (when ATP is hydrolyzed and the conformation changes from the substrate to product complex), subdomain 2A performs a rotation motion, such that the top of subdomain 2A rotates slightly into the page and the bottom rotates slightly out of the page as shown in Fig. 2b. Subdomains 1A and 1B move as one unit, rotating in the opposite direction to subdomain 2A. As a result, the cleft between subdomains 1A and 2A opens. Besides these motions, an anticlockwise rotation of subdomain 2B is also significant. In the relaxation stroke (when ATP rebinds to PcrA and the conformation changes from the product complex to the second substrate complex), the first eigenvector shows that, as ATP rebinds to the product complex, the rotation of subdomain 2A is performed in the direction that the top rotates slightly out of the page and the bottom rotates slightly into the page (Fig. 2b), and subdomains 1A and 1B rotate toward subdomain 2A. As a result, the cleft between subdomains 1A and 2A closes. At the same time, subdomain 2B rotates clockwise to its original position.

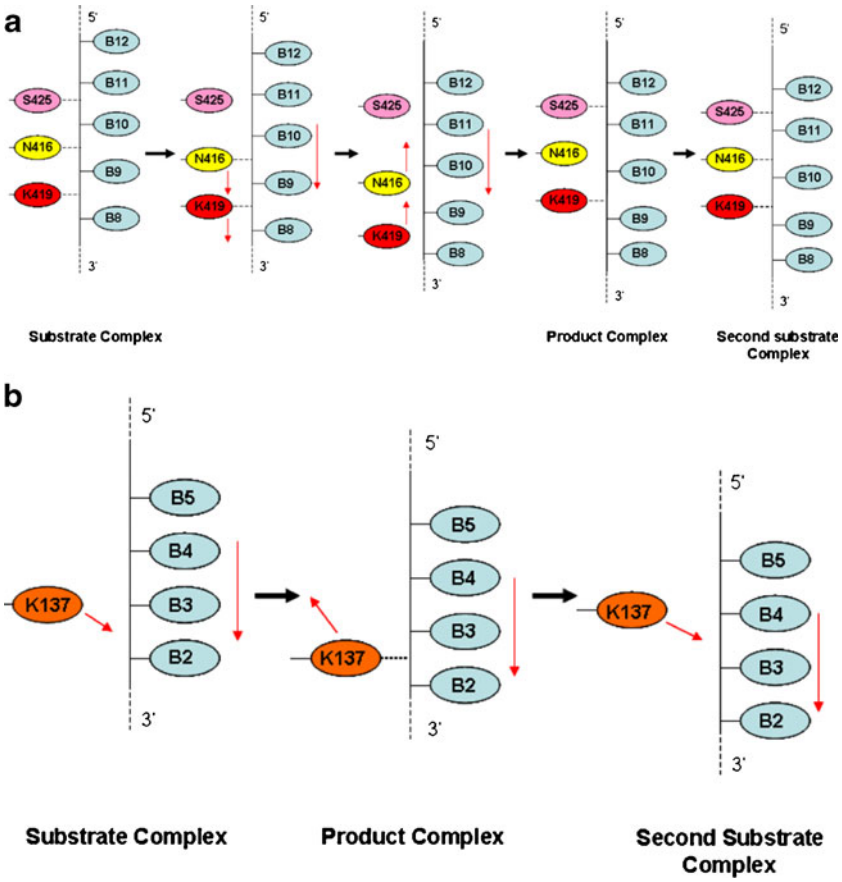


**Fig. 4** **a** The formation and breakage of the H-bonds between SER425, ASN416, LYS419 and the ds-DNA backbone. **b** The formation and breakage of the H-bond between LYS137 and the ds-DNA backbone. The *dotted lines* between the residues and the DNA backbone indicate the H-bonds between them, and the *red arrows* represent the motions of the DNA and residues. The *black arrows* represent the changes of states

The interactions between ds-DNA and PcrA

For clarity, 15 DNA base pairs along the ds-DNA are numbered from B15 (the tail of the ds-DNA) to B1 (the separation fork) in order as shown in Fig. 3(a). Five T bases along the ss-DNA are numbered from T5 to T1 from the separation fork to the 3′ end of the ss-DNA. During the reaction, subdomain 2B has strong interactions with the ds-DNA backbone. Five residues in subdomain 2B (ALA455, LYS456, SER425, ASN416 and LYS419 as shown in Fig. 3) form five H-bonds with the ds-DNA backbone. Two of them, ALA455 and LYS456, are at the top of subdomain 2B and form two H-bonds with the backbones of the first two DNA bases (B15 and B14) at the 3′ end of the 5′-3′ chain (short chain) of the DNA. The H-bond between ALA455 and the backbone of B15 only remains for a short period, as B15 moves down toward the separation fork in the power stroke. However, the H-bond between LYS456 and the backbone of B14 is maintained for the entire process and, during this time, LYS456 changes its conformation to follow the DNA's

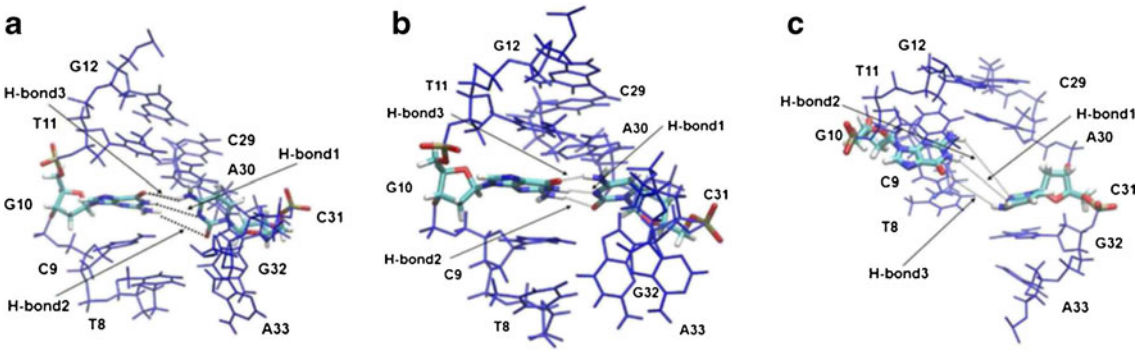


**Fig. 5** The separation of the ds-DNA base pair at the fork. **a** In the substrate complex, the base pair (G-C) at the fork is destabilized, and only one H-bond remains between them. **b** In the product complex, all three H-bonds between G-C base pair are reformed. **c** In the second substrate complex, G-C base pair is separated and three H-bonds are all broken

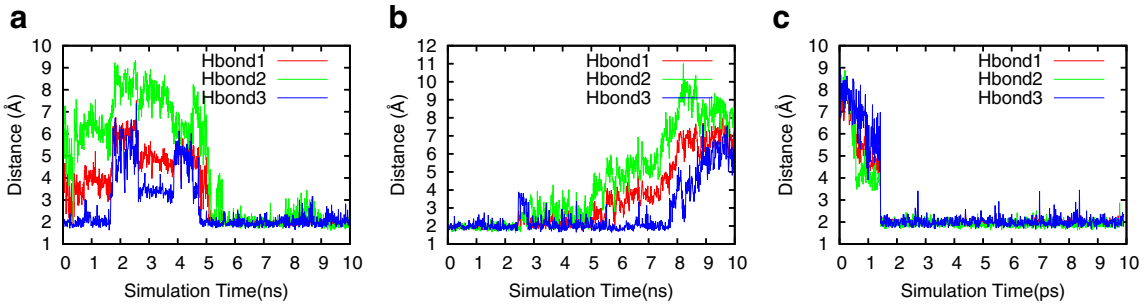


**Fig. 6** The distances of the H-bonds of the ds-DNA base pair at the fork. **a** The power stroke. **b** The relaxation stroke. **c** Along the free MD simulation on the product complex

movement. It is our expectation that these two residues should come back to their positions in the substrate complex during the relaxation stroke, and form H-bonds with successive bases.

The side chains of SER425, ASN416 and LYS419 have strong interactions with the backbones of three successive DNA bases on the 3′-5′ chain (long chain) of the ds-DNA. When the reaction starts from the substrate complex, SER425, ASN416 and LYS419 form three H-bonds with the DNA backbones of B10, B9 and B8 respectively. Then the H-bond between SER425 and the DNA backbone of B10 is broken first, but the H-bonds of ASN416 and LYS419 are maintained and 'help' the ds-DNA to move down toward the separation fork. After the ds-DNA has moved down by one base pair, these two H-bonds are broken, and ASN416 and LYS419 move back to their original positions. Then two new H-bonds are formed, one of which is between SER425 and the backbone of the base in B11 on the helix, and another one is between LYS519 and the backbone of B9. At this time the conformation of PcrA is in the product complex state. During the relaxation stroke, a H-bond is formed between ASN416 and the backbone of B10. The interactions of these three residues are represented in Fig. 4a.

At the DNA separation fork, one positively charged residue in subdomain 1B, LYS137, has strong interactions with the 3′-5′ chain of the DNA. LYS137 has no H-bond with the DNA in the substrate complex state, and the distance between LYS137 and the backbone of B2 is around 12 Å. In the power stroke, LYS137 moves close to the backbone of B2, and then a H-bond is formed between them. Followed by the relaxation stroke, this H-bond is broken and LYS137 moves back to the original position. The interaction is represented in Fig. 4b.

It was assumed that the free energy employed for the separation of ds-DNA should be obtained by the ATP hydrolysis, but based on the modeling, the separation was observed in the relaxation stroke rather than the power stroke. In the substrate complex, initially the G-C base pair at the fork is destabilized, and only one H-bond (O6 on G and H41 on C, Hbond3) remained between them, possibly because of the ATP binding. As Figs. 5 and 6(a, b) show that, after around 5.5 ns TMD simulation, three H-bonds between G-C are reformed, and all of them are maintained during the entire power stroke. During the relaxation stroke, after around 5 ns, these three H-bonds are all broken and the distances gradually increased so that the G-C base pair is separated.

In order to confirm that the ds-DNA fork separation we observed is not because of the TMD restraint, 10 ns free MD simulation was performed on the product complex. The equilibrium trajectory suggested that these three H-bonds were formed quickly after the start of the simulation, and were maintained during the whole simulation at around 2 Å (Fig. 6c). Based on the previous observations, we suggest that the separation of the ds-DNA happens because of the ATP binding rather than the ATP hydrolysis, which can not be observed through the crystal structures.

Our results are supported by the experiments described by Soultanas et al. By using Cu-phenanthrolate footprint to test

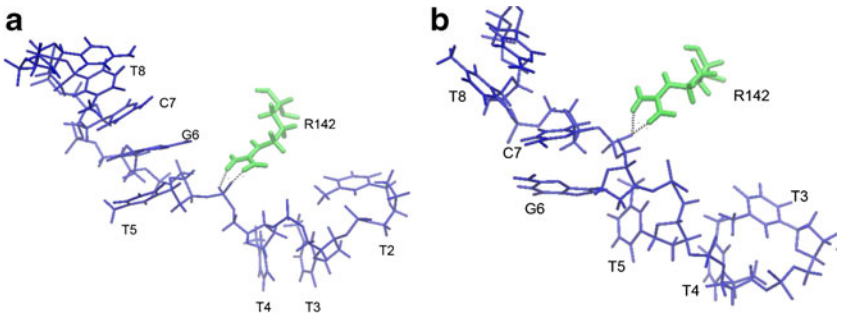


**Fig. 7** **a** ARG142 forms two H-bonds with the backbone of T4 during the period when the conformation changes from the substrate to the product complex. **b** ARG142 forms two H-bonds with the backbone of G6 as the conformation approaches close to the second substrate complex. The *dotted lines* between the residues and the DNA backbone indicate the H-bonds between them

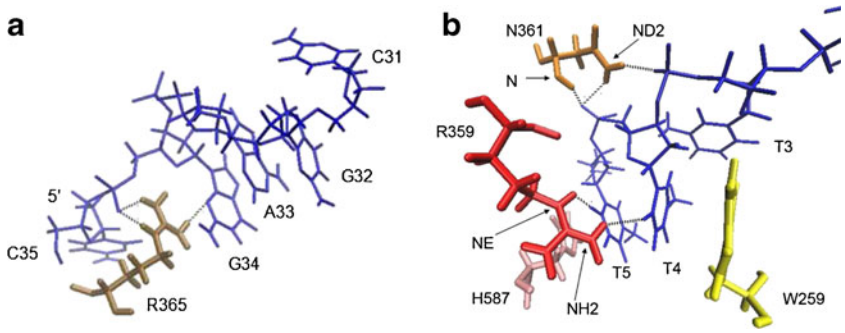


**Fig. 8** **a** The interaction between ARG365 and the 5′-3′ chain of the DNA. ARG365 forms three H-bonds with the second base (G) from the 5′ end of the 5′-3′ chain of the DNA. **b** ARG359 (in *red*) forms two H-bonds with the two bases (T4 and T5) in pocket 2 and ASN361 (in *orange*) forms three H-bonds with the backbones of T3 and T4

the distortion of DNA in the PcrA complex, they found that the ds-DNA is distorted dramatically when ADPNP, a non-hydrolyzable analogue of ATP, binds with PcrA complex. Therefore, they concluded that the energy associated with the binding of ATP is utilized by the enzyme to distort the DNA duplex and promote strand separation, which corresponds to the relaxation stroke in our simulation [13].

## The interactions between ss-DNA and PcrA

The ss-DNA translocation along the PcrA surface has been discussed in detail by us previously [19], and we now concentrate on the important interactions between key residues in Fig. 3a (only for ARG142) and b on PcrA and ss-DNA.

The position of ARG142 is close to the ds-DNA separation fork, but it has a strong interaction with the ss-DNA backbone. In the substrate complex, ARG142 does not form any H-bonds with the DNA. In the power stroke, ARG142 moves close to the backbone of T4, and when T5 moves toward pocket 3, it forms two H-bonds with the backbone of T4, as shown in Fig. 7a. As ARG142 moves back, these two H-bonds pull the DNA backbone upward to 'help' T5 to flip out of pocket 3. Then after these two H-bonds are broken, T5 drops into pocket 2. At this time, the conformation of PcrA is in the state of the product complex, in which ARG142 has no H-bond interaction with the DNA. During the relaxation stroke, ARG142 is stable and two H-bonds between it and the backbone of G6 are observed (Fig. 7b) when the conformation approaches close to the second substrate complex.

PHE626, one edge of pocket 3, plays the role of a door, which allows only one base to pass it each time. In the substrate complex, PHE626 forms a H-bond with T5 to stop it moving toward pocket 3. During the ATP hydrolysis step, PHE626 moves away, so that the H-bond is broken and T5 is then allowed to pass. Then when ATP rebinds to PcrA, PHE626 moves back to the original position and stops the next base (G6).

The 5′-3′ chain of the DNA passes over the surface of subdomain 2A after the ds-DNA has been separated. ARG365 has a strong interaction with this strand. As Fig. 8a shows, during the reaction, ARG365 forms three H-bonds with the second base (G) from the 5′ end of the 5′-3′ chain of the DNA. One of the H-bonds is between ARG365 and the G base, and the other two are formed between it and the backbone of the ss-DNA. When the ATP is hydrolyzed, ARG365 forms these three H-bonds to prevent the 5′-3′ chain from moving back. After one base pair of the ds-DNA has been separated, these H-bonds are broken and the 5′-3′ chain is pushed toward the 5′ end. Then ARG365 forms another three H-bonds with the successive base.

ARG359 can form two H-bonds with the two bases within the ss-DNA binding pocket 2. In this simulation, ARG359 forms two H-bonds with T3 and T4 separately in the substrate complex. As T3 flips out of pocket 2, the H-bond between the NH2 atom on ARG359 and the O2 atom on T3 is broken first

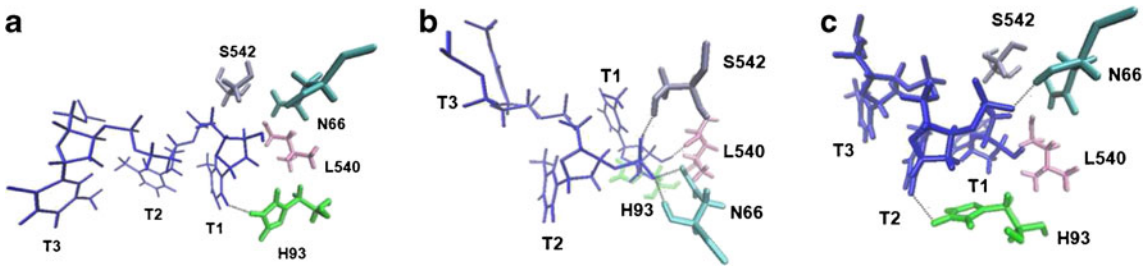


**Fig. 9** The H-bond interactions between the protein and 3′ end of the ss-DNA. **a** In the substrate complex, HIS93 (*green*) forms one H-bond with T1. **b** In the product complex, ASN66 (*cyan*), LEU540 (*mauve*) and SER542 (*iceblue*) form four H-bonds with T1. **c** In the second substrate complex, HIS93 (*green*) forms one H-bond with T2, and one of the H-bonds between T1 and ASN66 is maintained

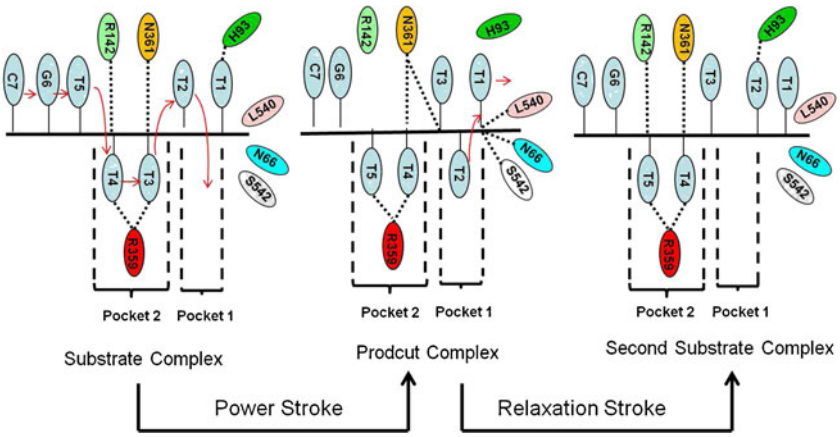


**Fig. 10** Schematic diagram to summarize the interactions between ss-DNA (T1 to C7) and key residues in PcrA along the unwinding cycle. The *dotted lines* indicate the H-bonds, and the *red arrows* represent the motions of the DNA

but the H-bond between NE on ARG359 and O2 on T4 is still maintained. After a short period, a new H-bond is formed between NH2 on ARG359 and O2 on T4. The formation of this new H-bond helps T4 move forward to take the position which was occupied by T3. When T5 moves into pocket 2, the H-bond between NE on ARG359 and O2 on T4 is broken and NE on ARG359 forms a new H-bond with O2 on T5 (Fig. 8b), and at this time the conformation has changed to the product complex. From the product to the second substrate complex, these two H-bonds are maintained all the time.

In the substrate complex, the N and ND2 atoms on ASN361 form two H-bonds with the backbone of T3. When T3 flips out of pocket 2, both of these H-bonds are broken. When T4 takes the position that was occupied by T3, the N and ND2 atoms form two H-bonds with the backbone of T4. When the conformation approaches the product complex, ND2 on ASN361 forms a H-bond with the backbone of T3, so now ASN361 forms H-bonds with the backbones of T3 and T4 at the same time (one H-bond with T3 and two with T4), as shown in Fig. 8b. During the process of the conformational changes from the product to the second substrate complex, after T2 flips out of pocket 1, ASN361 releases the H-bond with the backbone of T3, but the two H-bonds with T4 are maintained.

The interactions between the 3′ end of the ss-DNA and the protein involve four residues, which are HIS93, ASN66, LEU540 and SER542. In the substrate complex, HIS93 forms one H-bond with the base of T1 (Fig. 9). As T2 drops into pocket 1 when the conformation changes to the product complex, the H-bond between T1 and HIS93 is broken and then, ASN66, LEU540 and SER542 form four H-bonds with T1 to pull it forward. Then as the conformation changes to the second substrate complex, T2 takes the position that was occupied by T1 and forms a new H-bond with HIS93. One of the H-bonds between T1 and ASN66 is maintained in this process, and the other H-bonds are broken. The interactions between ARG142, ASN361, ARG359, HIS93, ASN66, LEU540 and SER542 with ss-DNA have been summarized in Fig. 10.

## Conclusions

In the study of conformational changes of the protein component of PcrA, by using PCA, the major motion of PcrA is illustrated by the movements of the four rigid-body subdomains. After the ATP is hydrolyzed in the substrate complex, subdomain 2A performs a rotation motion, subdomains 1A and 1B rotate in the opposite direction to subdomain 2A, and subdomain 2B performs an anticlockwise rotation. Then as the ATP binds to the product complex, all four subdomains rotate back to their original positions in the substrate complex.

In the study of the interactions between the protein and the ds-DNA, five residues (ALA455, LYS456, SER425, ASN416 and LYS419) found in subdomain 2B have strong interactions and form five H-bonds with the ds-DNA backbone. In general, they all perform similar functions. Initially, they form H-bonds with the ds-DNA backbone and help the ds-DNA to move down toward the strand separation fork. Then H-bonds between them and the ds-DNA backbone break somewhere in the middle of the reaction, and they come back to their original positions to form H-bonds with the backbones of successive base pairs. At the separation fork, LYS137 and ARG142 in subdomain 1B have strong interactions with the 3′-5′ chain of the DNA. LYS137 forms one H-bond with the ds-DNA backbone at the separation fork. The position of ARG142 is close to the separation fork, but it has strong H-bond interactions with the ss-DNA backbone to help T5 to flip out of pocket 3 and drop into pocket 2. The importance of some of these interactions has been suggested previously, because mutations of subdomain 2B (LYS456A and LYS419) and subdomain 1B (LYS137A) significantly reduced helicase activity without affecting the translocation of ss-DNA or ss-DNA binding [13].

The analysis of the interactions between the protein and ss-DNA reveals that ARG365 forms H-bonds with the 5′ end of the 5′-3′ chain of the DNA to prevent it from moving back. ARG359 forms H-bonds with the two bases in pocket 2 (T4 and T5), and ASN361 forms H-bonds with the backbones of

T3 and T4. These two residues are involved in the motions of T3 flipping out of pocket 2 and T4 moving forward.

In general, based on the study of the interactions between the protein and the DNA (both ds-DNA and ss-DNA), we observed that most of the H-bonds (except ARG359) are formed between the protein residues and the backbone of the DNA. This result indicates that the translocation of PcrA is independent of the DNA sequence. The TMD simulation provides information which cannot be observed from the crystal structures, especially the breakage and formation of the H-bonds during the middle of the reaction. For example, as described above, in the crystal structures of the substrate complex and the product complex, ARG142 has no H-bond interaction with the DNA. However, from the TMD simulation, strong H-bond interactions have been found between it and the ss-DNA backbone during the middle of the reaction.

We observed that two residues of subdomain 2B (LEU540 and SER542) and two of subdomain 1A (HIS93 and ASN66) form 4 H-bonds with the 3′ end of the ss-DNA in the product complex and only two (between subdomain 1A and ss-DNA) in the second substrate complex. It suggests a mechanism that in the power stroke, subdomains 2B and 1A hold the ss-DNA tightly and pull it forward according to the global conformational changes of PcrA described in Fig. 2. In the relaxation stroke, the interaction between subdomains 2B with ss-DNA are released to allow the ss-DNA to move across the surface of subdomain 1A. This mechanism is supported by previous MD simulation study, by using steered molecular dynamics simulations, that ss-DNA can be pulled past subdomain 2A more easily than past 1A in the substrate complex, whereas the situation is reversed in the product complex, so that the unidirectional translocation of ss-DNA can be ensured [31, 32].

It was generally thought that the ss-DNA translocation is mainly provided by subdomains 1A and 2A, and ds-DNA destabilization is provided by subdomains 1B and 2B. However, based on the interactions between ARG142, LEU540 and SER542 with ss-DNA, we suggest that subdomain 1B is also involved in the ss-DNA translocation. When we consider that the position of ARG142 is close to the separation fork, it suggests that ARG142 could be a very important link to couple the processes of ss-DNA translocation and ds-DNA destabilization; mutating this residue would be helpful to further understand the PcrA unwinding mechanism.

The simulations indicate that the ds-DNA is separated upon the ATP rebinding, rather than the hydrolysis, which suggests that the two strokes in the mechanism have two different major roles. Firstly, in the power stroke (ATP hydrolysis), most of the translocations of the bases from one pocket to the next occur. In the relaxation stroke (ATP binding), most of the 'work' is being done to 'melt' the DNA at the separation fork. Therefore, we propose a mechanism whereby the translocation of the ss-DNA is powered by ATP hydrolysis and the separation of the ds-DNA is powered by ATP binding.

**Acknowledgments** This work was supported by the University of Nottingham; National Natural Science Foundation of China [81260481]; Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry [2013-693]; Chunhui Program, Ministry of Education of the People's Republic of China [Z2011050]; Scientific Research Project for Ningxia Colleges and Universities, Department of Education of Ningxia [2011JY004] and University of Malaya-MOHE High Impact Research grant (UM.C/625/1/HIR/MOHE/DENT/22).

## References

1. Anand SP, Khan SA (2004) Nucleic Acids Res 32(10):3190
2. Hall MC, Matson SW (1999) Mol Microbiol 34:867
3. van Brabant AJ, Stan R, Ellis NA (2000) Annu Rev Genomics Hum Genet 1:409
4. Ellis NA (1997) Curr Opin Genet Dev 7(3):354
5. Mackintosh SG, Raney KD (2006) Nucleic Acids Res 34(15):4154
6. Subramanya HS, Bird LE, Brannigan JA, Wigley DB (1996) Nature 384:379
7. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB (1999) Cell 97(1):75
8. Bird LE, Brannigan JA, Subramanya HS, Wigley DB (1998) Nucleic Acids Res 26(11):2686
9. Caruthers JM, McKay DB (2002) Curr Opin Struct Biol 12(1):123
10. Dillingham MS, Wigley DB, Webb MR (2000) Biochemistry 39(1):205
11. Dillingham MS, Wigley DB, Webb MR (2002) Biochemistry 41(2):643
12. Cox K, Watson T, Soultanas P, Hirst JD (2003) Proteins 52(2):254
13. Soultanas P, Dillingham MS, Wiley P, Webb MR, Wigley DB (2000) EMBO J 19(14):3799
14. Soultanas P, Dillingham MS, Velankar SS, Wigley DB (1999) J Mol Biol 290(1):137
15. Withers IM, Mazanetz MP, Wang H, Fischer PM, Laughton CA (2008) J Chem Inf Model 48(7):1448. doi:10.1021/ci7004725, URL http://www.ncbi.nlm.nih.gov/pubmed/18553961
16. Dillingham MS, Soultanas P, Wiley P, Webb MR, Wigley DB (2001) Proc Natl Acad Sci U S A 98(15):8381
17. Bertram RD, Hayes CJ, Soultanas P (2002) Biochemistry 41(24):7725
18. Abbas S, Bertram RD, Hayes CJ (2001) Org Lett 3:3365
19. Wang H, Laughton CA (2012) Phys Chem Chem Phys (PCCP) 14(35):12230. doi:10.1039/c2cp41193h, URL http://www.ncbi.nlm.nih.gov/pubmed/22864246
20. Aci S, Mazier S, Genest D (2005) J Mol Biol 351(3):520
21. Compoint M, Picaud F, Ramseyer C, Girardet C (2005) J Chem Phys 122(13):134707
22. Ferrara P, Apostolakis J, Caflisch A (2000) Proteins 39(3):252
23. Kamerlin SC, Rucker R, Boresch S (2006) Biochem Biophys Res Commun 345(3):1161
24. Rodriguez-Barrios F, Gago F (2004) J Am Chem Soc 126(47):15386
25. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) AMBER 8
26. Schlitter J, Engels M, Kruger P (1994) J Mol Graph 12(2):84
27. van der Vaart A, Karplus M (2005) J Chem Phys 122(11):114903
28. Shields GC, Laughton CA, Orozco M (1997) J Am Chem Soc 119(32):7463. doi:10.1021/ja970601z, URL http://pubs.acs.org/doi/abs/10.1021/ja970601z
29. Darden T, York D, Pedersen L (1993) J Chem Phys 98(12):10089
30. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) J Chem Phys 81(8):3684
31. Yu J, Ha T, Schulten K (2006) Biophys J 91:2097
32. Yu J, Ha T, Schulten K (2007) Biophys J 93:3783