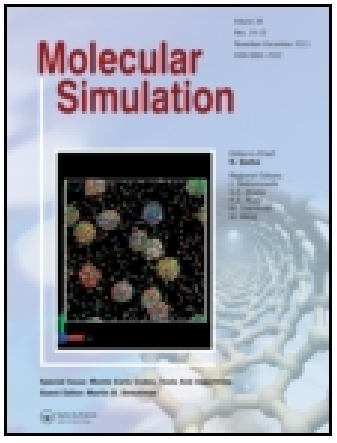

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmos20

Identification of the PcrA DNA helicase reaction pathway by applying advanced targeted molecular dynamic simulations

Hao Wang[a], Wei Hong[bc], Ian C. Paterson[c], Jing Pu[d] & Charles A. Laughton[e]

[a] Department of Medicinal Chemistry, School of Pharmacy, Ningxia Medical University, 1160 Shengli Street, Yinchuan, Ningxia 750004, P.R. China

[b] School of Chemistry and Chemical Engineering, Beifang University of Nationalities, Yinchuan, Ningxia 750021, P.R. China

[c] Dental Research and Training Unit and Oral Cancer Research and Coordinating Centre, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia

[d] Key Laboratory of Fertility Preservation and Maintenance of Ministry of Education, Ningxia Medical University, 1160 Shengli Street, Yinchuan, Ningxia 750004, P.R. China

[e] School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK

Published online: 21 Jan 2014.

To cite this article: Hao Wang, Wei Hong, Ian C. Paterson, Jing Pu & Charles A. Laughton (2014) Identification of the PcrA DNA helicase reaction pathway by applying advanced targeted molecular dynamic simulations, Molecular Simulation, 40:15, 1290-1299, DOI: 10.1080/08927022.2013.875170

To link to this article: http://dx.doi.org/10.1080/08927022.2013.875170

Molecular Simulation, 2014
Vol. 40, No. 15, 1290–1299, http://dx.doi.org/10.1080/08927022.2013.875170

Taylor & Francis
Taylor & Francis Group

Identification of the PcrA DNA helicase reaction pathway by applying advanced targeted molecular dynamic simulations

Hao Wang[a]*, Wei Hong[b,c], Ian C. Paterson[c], Jing Pu[d] and Charles A. Laughton[e]

[a]*Department of Medicinal Chemistry, School of Pharmacy, Ningxia Medical University, 1160 Shengli Street, Yinchuan, Ningxia 750004, P.R. China;* [b]*School of Chemistry and Chemical Engineering, Beifang University of Nationalities, Yinchuan, Ningxia 750021, P.R. China;* [c]*Dental Research and Training Unit and Oral Cancer Research and Coordinating Centre, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia;* [d]*Key Laboratory of Fertility Preservation and Maintenance of Ministry of Education, Ningxia Medical University, 1160 Shengli Street, Yinchuan, Ningxia 750004, P.R. China;* [e]*School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK*

(*Received 27 September 2013; final version received 6 December 2013*)

PcrA DNA helicase uses the free energy of hydrolysis and binding of ATP to unwind double-stranded DNA (ds-DNA). There are two states of PcrA, termed the substrate and product complexes and, through the conformational changes between these two states, PcrA moves along ds-DNA and separates the two strands. In this study, two different methods, namely chain minimisation (CM, less reliable method) and auto targeted molecular dynamic (TMD) simulation (more reliable), were performed to generate two different initial reaction pathways between these two states, and then fixed root mean square distance (RMSD) TMD simulation was performed to optimise these two initial pathways. In general, the two optimised pathways share very similar major conformational changes, but are different in the minor motions. The potential energy profiles of the two improved pathways are generally similar, but the one generated by the improved TMD path is slightly lower. Considering the poor reliability of the initial path generated by CM and insignificant improvements of the auto-TMD path, our study suggests that fixed RMSD TMD simulation can generate reliable reaction pathways, but the different initial paths still have some influence on the detailed conformational analysis.

Keywords: PcrA DNA helicase; targeted molecular dynamic simulation; molecular modelling; molecular dynamics

1. Introduction

DNA helicases use the free energy of hydrolysis and binding of ATP to unwind double-stranded DNA (ds-DNA) into two single strands DNA (ss-DNA). These important enzymes are involved in almost every process that involves nucleic acids.[1,2] The *Bacillus stearothermophilus* PcrA DNA helicase [3] belongs to super family 1 (SF-1) of DNA helicases, displays a 3′ to 5′ polarity in DNA unwinding [4,5] and works as a monomer,[6] which can be described as an inchworm model.[3] The two structures of PcrA which are involved in ds-DNA separation have been solved by X-ray crystallography.[3,6] These are the substrate complex form (with ATP bound) and the product complex form. The reaction mechanism of PcrA unwinding cycle can be divided into the power and relaxation strokes. In the power stroke, the bound ATP in the substrate complex is hydrolysed to drive the conformation of PcrA to the product complex. In the relaxation stroke, one new ATP molecule rebinds to PcrA to drive the conformation back from the product complex to the substrate complex. Current molecular modelling technology can help to further our understanding of the reaction pathway of the PcrA unwinding cycle. In theory, the reaction pathway can be archived if sufficient computational time can be provided for free molecular dynamic (MD) simulations. However, because PcrA contains more than 11,000 atoms and unwinds ds-DNA at around 0.02 s per base pair,[7–9] long simulations on such a large system cannot be afforded with current computational ability.

Currently, the available methods to simulate paths of conformational changes can roughly be divided into two general groups.[10] The first group deals with methods that optimise an existing pathway between two energy minimas. Using these methods, structures on the path are minimised simultaneously thus providing a low-energy path. The starting path can be a collection of linearly interpolated Cartesian structures. An advantage of these methods is that motions involved in the structural reorganisation can be filtered out of the large collection of unproductive thermal motions of an MD simulation, but the manual choice of the intermediate structures can make the reaction path dependent on user intervention.[10]

The second group deals with methods that directly generate pathways. In these kinds of methods, biasing techniques are always applied to drive the trajectory from the initial state to the final state, where both are known from high-resolution structural studies.[11–15] The widely used targeted molecular dynamic (TMD) simulation belongs to this group, in which a trajectory from a

*Corresponding author. Email: paxhw@yahoo.co.uk

known initial structure to a known target structure is generated by the use of a holonomic constraint that reduces the root mean square distance (RMSD) from the target by a preset value at each MD step.[16–20]

In this study, chain minimisation (CM, in the first group, offering relatively poor initial pathway) and TMD simulation (in the second group, offering more reliable initial pathway) were performed to generate initial pathways, and then fixed RMSD TMD simulation was performed to optimise these two initial pathways. The results show that these two optimised pathways are very similar in the major conformational changes, but are different in the minor motions, which cannot be observed from the comparison of the crystal structures.

2. Experimental

2.1 *Chain minimisation*

CM is a method created to optimise the guessed reaction pathway between two conformations by finding low-energy reaction coordinates. The method is shown in Figure 1, and the strategy is listed as follows:

1. A series of snapshots (n snapshots including the two ends) are collected as the initial guessed reaction pathway, for example by linear interpolation between the initial and final states. In Figure 1, the initial and final states are coloured in black and the other reaction coordinates generated by linear interpolation are coloured in light blue.
2. Energy minimisations are performed for all the reaction coordinates except the initial and final states. Small harmonic restraints are applied to ensure that the conformations will not change too much from the original snapshots. The structures generated by these energy minimisations are coloured in red in Figure 1.
3. One average structure (green) is generated from each two adjacent reaction coordinates [those obtained from step 2 (red) and the two end states (black)].
4. Restrained energy minimisations are carried out on the average structures (green) obtained from step 3 (the resulting structures are coloured in yellow). Then all the energy minimised structures from step 2 (red) and this step (yellow) together with the two ends (black) are linked together to form a new reaction pathway. One cycle of CM has been completed.
5. The energy minimised structures (yellow) obtained from step 4 are used as input reaction coordinates to generate new average structures (grey) from each adjacent pair.
6. Restrained energy minimisations are applied to the average structures (grey) obtained from step 5 (the resulting structures are coloured in brown).
7. One average structure (purple) is generated from each two adjacent reaction coordinates [those obtained from step 6 (brown) and the two end states (black)].
8. Restrained energy minimisations are carried out on the average structures (purple) obtained from step 7 (the resulting structures are coloured in orange). Then all the energy minimised structures in step 6 (brown) and this step (orange) together with the two ends (black) are linked together to form a new reaction pathway. Cycle 2 of CM has been completed. Cycle 2 will be repeated until the potential energies of each series of reaction coordinates do not change significantly between cycles. Then the lowest potential energy pathway will be chosen as the predicted reaction pathway, and $2n - 1$ snapshots can be collected from this path.

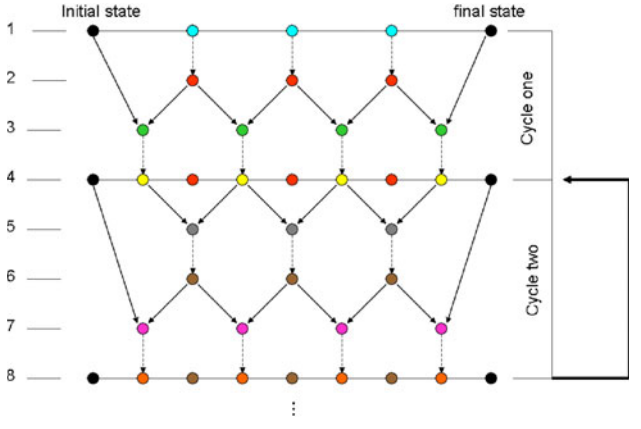

Figure 1. (Colour online) A schematic diagram of CM. The dotted-line arrows represent energy minimisations, and the solid-line arrows represent interpolations. The details are explained in the CM strategy.

The CM process is based on energy minimisations, so it avoids running MD simulations which usually need large amounts of computational resources. However, for the same reason, it may only find the local minimas, and cannot reach the global minimas. The results of CM heavily depend on the initial guessed pathway. This method also only optimises the reaction coordinates, so the possibility of missing transition points along the path cannot be avoided.

2.2 *Targeted molecular dynamics*

TMD can be used to simulate transitions between known end structures.[21–25] By restraining the conformation of the starting structure to match the target slowly (following the RMSD reducing direction), the reaction pathway

between them can be found. There are two ways to perform TMD, called by us as auto-TMD and fixed-TMD. Auto-TMD is combined with a weight change technique to keep decreasing the target RMSD automatically, so that the simulation can be constrained to find a path from the start to the end continuously and the transition points along the path can be found. The restraint force added by auto-TMD in general can be small when the target RMSD is relatively large, since it leaves sufficient freedom for the system to adjust the conformation for low energy states. However, when the target RMSD reduces to a small value, the restraint force usually becomes large, which may lead the system to overcome large energy barriers inaccessible to the unforced system.[26] In fixed-TMD, the target RMSD can be set to a constant value, so that a certain RMSD region around the target can be explored. By comparison to auto-TMD, fixed-TMD can generate a series of windows, and each window contains an equilibrium simulation with lower restraint and longer simulation time to relax the system.

2.3 Principal component analysis

A simulation system may contain more than 10,000 atoms, therefore after a long simulation, a huge amount of data will be generated to describe the coordinates of each atom in every snapshot. It is extremely difficult and unnecessary to analyse all the information from the trajectory, because there are only a few aspects of the data that contribute to the major motion of interests. Principal component analysis (PCA) is a statistical method used to discover correlations between the different components of a data-set, and thereby provides a means for reducing the complexity of the representation of the data.[27] For a detailed discussion of the PCA approach, see Sherer et al. [28].

2.4 Experimental procedures

The original crystal structures of the substrate complex (3PJR) and the product complexes (2PJR) of PcrA were obtained from the Protein Data Bank. In order to simulate the relaxation stroke, 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. All these three structures were explicitly solvated in a truncated octahedral box (at least 10 Å from the complex to avoid periodic artefacts from occurring) of TIP3P model water (25,419 water molecules), and treated carefully under Amber-03 force field to set up reliable starting structures for further simulations as described in our previous publication and supplements.[29] These three structures were then optimised by energy minimisations and MD simulations with our standard equilibration strategy using the Amber 8 package. Initially, the solvent was energy minimised by 50 steps of the steepest descent method and then followed by 10,000 steps of the conjugate gradient method with the default non-bonded cut-off of 8 Å, while the solute was held fixed. Second, the entire system was then energy minimised with the same settings as for the previous step. Third, the solvent was subjected to a short (20 ps) MD simulation at a temperature of 100 K, under constant pressure conditions. For this and all following simulations, MD simulations were performed with explicit solvent models and in the NPT ensemble ($T = 300$ K; $P = 1$ atm). Periodic boundary conditions (PBC) and particle-mesh-Ewald (PME) method [30] were used to model long-range electrostatic effects, while the temperature was coupled to an external bath using a weak coupling algorithm.[31] The cut-off non-bonded interaction was set as 8 Å. The bond interactions involving H-atoms are constrained by using the SHAKE algorithm. Fourth, the solvent temperature was then raised to 300 K over 20 ps. During both this phase and the last, position restraints on every solute atom (force constant 100 kcal/mol/Å^2) maintain it in its energy-minimised conformation. Fifth, over a series of 20 ps constant pressure simulations at 300 K, the restraints on the solute were gradually relaxed (100, 50, 25, 10, 5, 2 and then 1 kcal/mol/Å^2 respectively). Sixth, after all these simulations were done, the entire system was then optimised by a 300 ps unrestrained MD simulation at 300 K.

2.4.1 Generation of initial pathways

The initial CM pathway. The substrate and product complexes were used as the start and the end structures, respectively, and an initial guessed pathway was generated by linear interpolation. The original RMSD between the substrate (or second substrate) and product complexes is 3.19 Å. Five original snapshots, whose RMSDs with respect to the product complex are 0, 0.79, 1.59, 2.38 and 3.19 Å (average increment is about 0.79 Å), were collected along this guessed pathway. CM was applied on each snapshot, and the energy of the system was first minimised using 100 steps of the steepest decent method and then followed by 100,000 steps of the conjugate gradient method or until the root-mean square of the Cartesian elements of the gradient is less than 0.2 kcal/mol/Å. A small harmonic restraint (force constant 0.1 kcal/mol/Å^2) was applied to the system to ensure that the conformation would not change too much from the original snapshots. The same CM strategy was also applied to the relaxation stroke of converting the product to the second substrate complex.

The initial auto-TMD pathway. After the equilibration, auto-TMD was first run for 1 ns to drive the conformation

of the substrate complex (starting structure) to the product complex (target structure). In this simulation, the force constant for TMD was set at 3.2 kcal/mol/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 500,000 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). PBC and PME method [30] are used to model long-range electrostatic effects, while the temperature is coupled to an external bath using a weak coupling algorithm.[31] The cut-off non-bonded interaction was set as 8 Å. The bond interactions involving H-atoms were constrained by using the SHAKE algorithm. 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.

2.4.2 *Optimisation on the initial pathways*

Optimising the CM pathways. As the CM method can only find the local minimal points, an improved version of chain minimisation (ICM) combined with TMD with fixed targeted RMSD (fixed-TMD) was created to optimise the reaction paths. In the power stroke, eight reaction coordinates were collected from the last cycle. RMSDs of these snapshots with respect to the product complex were 3.10, 2.70, 2.30, 1.88, 1.46, 1.01, 0.88 and 0.56 Å. Then for each snapshot, a 2 ns fixed-TMD simulation with a small force constant (0.1 kcal/mol/Å^2) was used to search the conformational space within the RMSD region defined individually by each snapshot at constant temperature and pressure ($T = 300$K; $P = 1$ atm). In each window, the target RMSD was kept constant at the original value and the simulation in each window is an equilibrium simulation. The average structure for each equilibrium window was linked together to represent the reaction path.

The same method was also applied to the seven snapshots in the relaxation stroke, whose RMSDs with respect to the second substrate complex were 3.23, 2.88, 2.49, 2.08, 1.65, 1.21 and 0.76 Å. As the small force constant (0.1 kcal/mol/Å^2) was only able to drive the simulation to reach the lowest RMSD region of 1.902 Å in this case, the fixed-TMD simulations were performed on the last three snapshots with a higher force constant (0.2 kcal/mol/Å^2) to bring the simulation to the lower RMSD regions.

Optimising the TMD pathways. When the auto-TMD method was applied to generate the initial pathway, a large force constant (3.2 kcal/mol/Å^2) was set, and only 1 ns simulation was performed. With such a high-force constant and short simulation time, auto-TMD may drive the trajectory to overcome large energy barriers which usually should be avoided by free simulation. So fixed-TMD was also performed to improve the auto-TMD path (ITMD). The auto-TMD path was used as the initial path and 11 snapshots were collected along it. RMSDs of the snapshots with respect to the product complex were 3.42, 3.07, 2.73, 2.40, 2.05, 1.72, 1.40, 1.08, 0.79, 0.54 and 0.34 Å, in which the first one was the conformation of the substrate complex after equilibrations. For each snapshot, 2 ns TMD simulation with a force constant 0.1 kcal/mol/Å^2 was applied at constant temperature and pressure ($T = 300$ K; $P = 1$ atm). The average structure for each equilibrium window was linked together to represent the reaction path.

In the relaxation stroke, fixed-TMD was repeated on the 11 snapshots, whose RMSDs with respect to the second substrate complex were 3.39, 3.06, 2.72, 2.41, 2.08, 1.76, 1.47, 1.19, 0.96, 0.79 and 0.35 Å, in which the first one was the conformation of the product complex after equilibrations. Similar to the ICM pathway in the relaxation stroke, the higher force constant of 0.2 kcal/mol/Å^2 had to be used on the snapshot of 0.79 Å to bring the simulation to the low RMSD region.

In comparison with the unimproved TMD method, this improved method allows the simulation to run for a longer time in each RMSD region with a smaller force constraint, so that the simulation can return from the peaks of the energy barriers.

3. Results

3.1 Comparison of the two optimised paths in the power stroke

Upon completion of both the ICM and ITMD simulations, the average structure was generated in each window. A two-dimensional (2D) RMSD plot, as shown in Figure 2(a), is used to show the distance between the two paths. The RMSDs for every combination of matching snapshots along the two reaction paths were calculated and put into a matrix, in which the value along the diagonal line shows the RMSD difference between the two snapshots in the same RMSD region with respect to the target. So if the two paths are similar, the RMSD values along the diagonal line should be small. In Figure 2(a), the high RMSD value is highlighted in red and the low value is coloured in blue. In general,

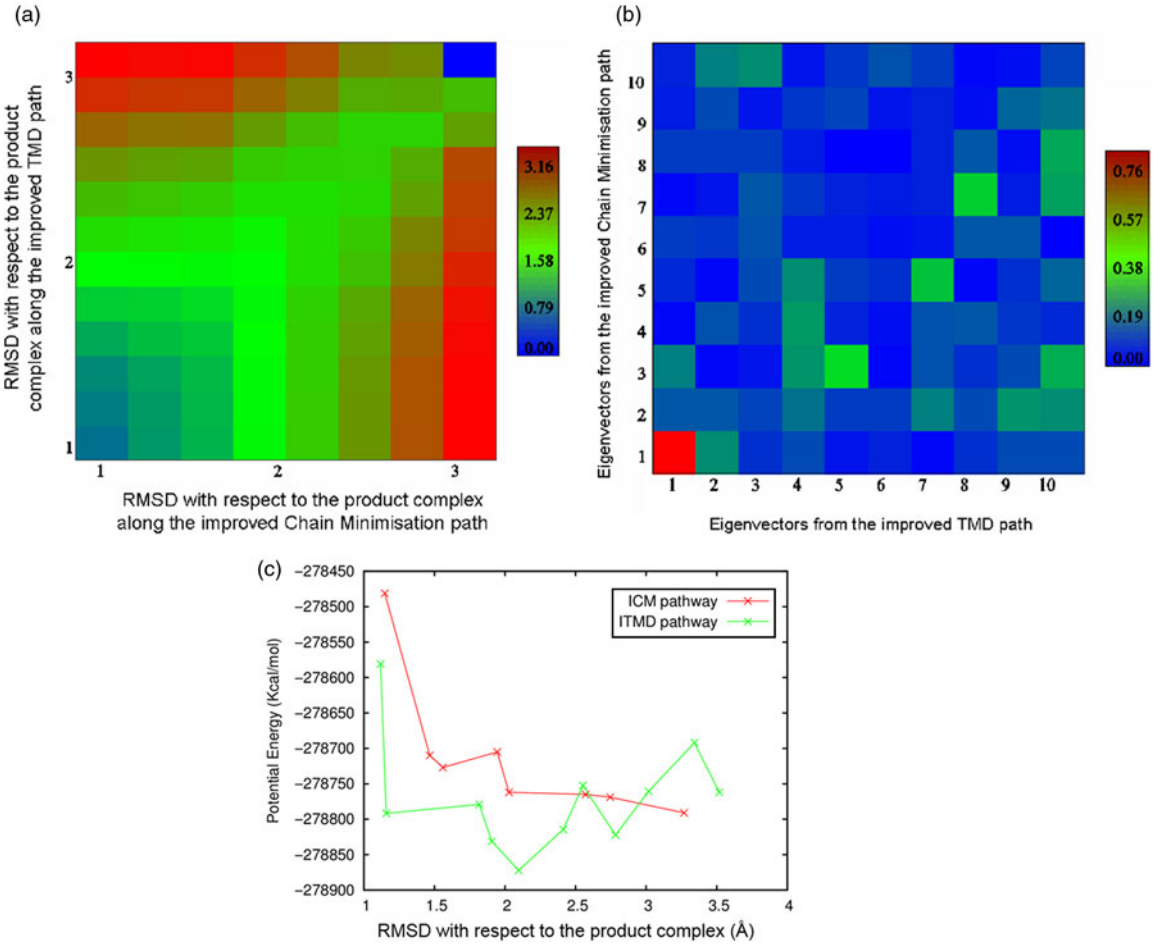

Figure 2. (Colour online) Comparisons between the improved CM and improved TMD paths in the power stroke. (a) The RMSD differences between these two paths. (b) Comparisons of eigenvectors determined from these two pathways. (c) The potential energy comparisons of these two pathways.

Figure 2(a) shows that the two pathways start from the substrate complex (upper right corner), and diverge in the middle (the RMSD between them is around 1.6 Å), and then they converge again when they both reach the product complex region (lower left corner) and at this time the RMSD between them is around 0.69 Å. This result shows that there are some differences between these two pathways, and the largest difference between them is no more than 1.6 Å.

PCA was applied on the backbone of the system (only carbon atoms on the protein and phosphorus atoms on the DNA) to analyse and compare the trajectories generated from the ICM and the ITMD methods. Based on PCA, the quantitative comparisons of the eigenvectors generated from the two paths are accessible through calculating the dot product between each two of the eigenvectors. The high dot product value represents high similarity and a low value represents low similarity. The dot products for every combination of matching components along the two reaction paths are presented in graphical format in Figure 2 (b). The strongest similarity, which is represented by a dot product value of 0.76, is observed from the comparison of both the first eigenvectors along the reaction pathways in the power stroke (Figure 2(b)). Because the first eigenvector shows the most important motion of the system, particularly in this experiment where the first eigenvector carries around 51.3% importance on the ICM path, and 47.7% on the ITMD path, the high dot product value on both the first eigenvectors shows that the major motions of the system along these two pathways are highly similar.

However, in both paths the conformation changes from the same starting structure to the same target, it is perhaps not surprising that the major conformational changes are similar, and it is also necessary to consider the dot product of the second eigenvectors. In the power stroke, the second eigenvector carries 18.5% importance on the ICM path and 16% on the ITMD path. The dot product value between them is 0.13, which is very low.

Two trajectories along these two second eigenvectors were created, and through comparing these two trajectories, the most significant difference of the conformation-

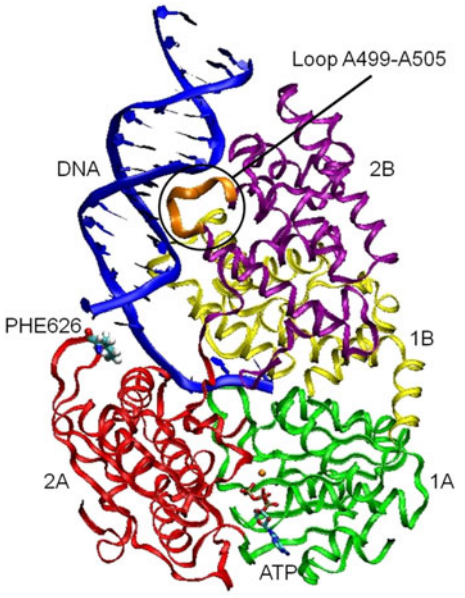

Figure 3. (Colour online) The structure of PcrA DNA helicase. Subdomain 1A is coloured green, subdomain 1B yellow, subdomain 2A red and subdomain 2B purple. The bound DNA is coloured blue. The ATP is coloured by atom type and the magnesium ion is in orange, and they are located in subdomains 1A and 2A. PHE626 (coloured by atom type) is located in subdomain 2A and close to the bound DNA. The loop A499–A505 is coloured in orange and located in subdomain 2B.

al changes is obtained from the motion of subdomain 2A. Along the second eigenvector of the ICM path, subdomain 2A performs a tilting motion with respect to the other subdomains, while along the ITMD path, it performs a rotation motion with respect to the other subdomains. PHE626 is located on the top of subdomain 2A (Figure 3) and plays the role as a door, which allows only one DNA base to pass it each time. Although subdomain 2A performs the tilting motion (along the ICM path) or the rotation motion (along the improved TMD path), the local motions of PHE626 are the same along these two paths. In other words, the different global motions of subdomain 2A generate the same local motion of PHE626, which is believed important for the PcrA unwinding mechanism. Along the second eigenvectors of these two paths, PHE626 moves away from the ss-DNA so that it leaves a space for one base passing it, and then it comes back to the original position to block the next base. A less significant difference is obtained from the motion of subdomain 2B. Along the second eigenvector of the ITMD path, subdomain 2B performs a rotation motion, but along the ICM path, subdomain 2B performs a translation motion. Subdomains 1A and 1B, along both paths, are quite stable.

The potential energies along the two paths were collected for thermodynamic analysis. As the ICM and ITMD generated a series of equilibrium simulations, the average potential energies in each window were collected together and presented as the potential energy curve in Figure 2(c). In general, the trends of the potential energy profiles along these two pathways in the power stroke are similar, although the ICM path is slightly higher. When the conformations get close to the product complex, the potential energy increases dramatically, which is probably because of the poor resolution and missing residues of the product complex. As the crystal structure of the product complex was of a poor resolution (2.9 Å) and many residues/bases were missing, the product complex was rebuilt using the information of the substrate complex as described in our previous paper and Supplementary data.[29]

3.2 *Comparison of the two optimised paths in the relaxation stroke*

The 2D RMSD plot (Figure 4(a)) in the relaxation stroke indicates that the two pathways start from the product complex (upper right corner), and diverge in the middle (the RMSD between them is around 1.6 Å), and then they converge again when they both reach the second substrate complex region (lower left corner) and at this time the RMSD between them is around 0.96 Å.

Based on the PCA study (Figure 4(b)), the highest dot product value, around 0.71, is determined from the comparison of both the first eigenvectors. In this process, the first eigenvector carries 44.6% importance on the ICM path and 52.8% on the ITMD path. The second eigenvector carries 16.5% importance on the ICM path and 14.5% on the ITMD path, and the dot product value between them is only 0.14.

Comparison of the potential energy curves (Figure 4 (c)) of these two pathways shows that the trends of these two profiles are also similar, and again the potential energy curve of the ICM path is slightly higher. When the simulations reach the small RMSD region, the energy curves of both pathways increased, which is because the substrate (or second) complex is in a high-energy state.

The significant difference of the conformational changes is obtained from the motion of subdomain 2B. Along the ICM path, subdomain 2B performs a rotation motion with respect to subdomain 1B, and at the same time the distance between loop A499–A505 on subdomain 2B (Figure 3) and the ds-DNA changes. This distance increases to 15.42 Å (Figure 5(b)) first and then comes back to the original value of 9.92 Å (Figure 5(a)). Along the ITMD path, subdomain 2B performs a twist motion with respect to subdomain 1B, and the distance between loop A499–A505 and the ds-DNA is maintained in a small region from 10.29 Å (Figure 5(c)) to 11.47 Å (Figure 5(d)). The relationship between these motions and the PcrA unwinding mechanism has not been found in this study. For the other subdomains along the ICM path, subdomain 2A performs a tilting motion, subdomain 1A rotates itself

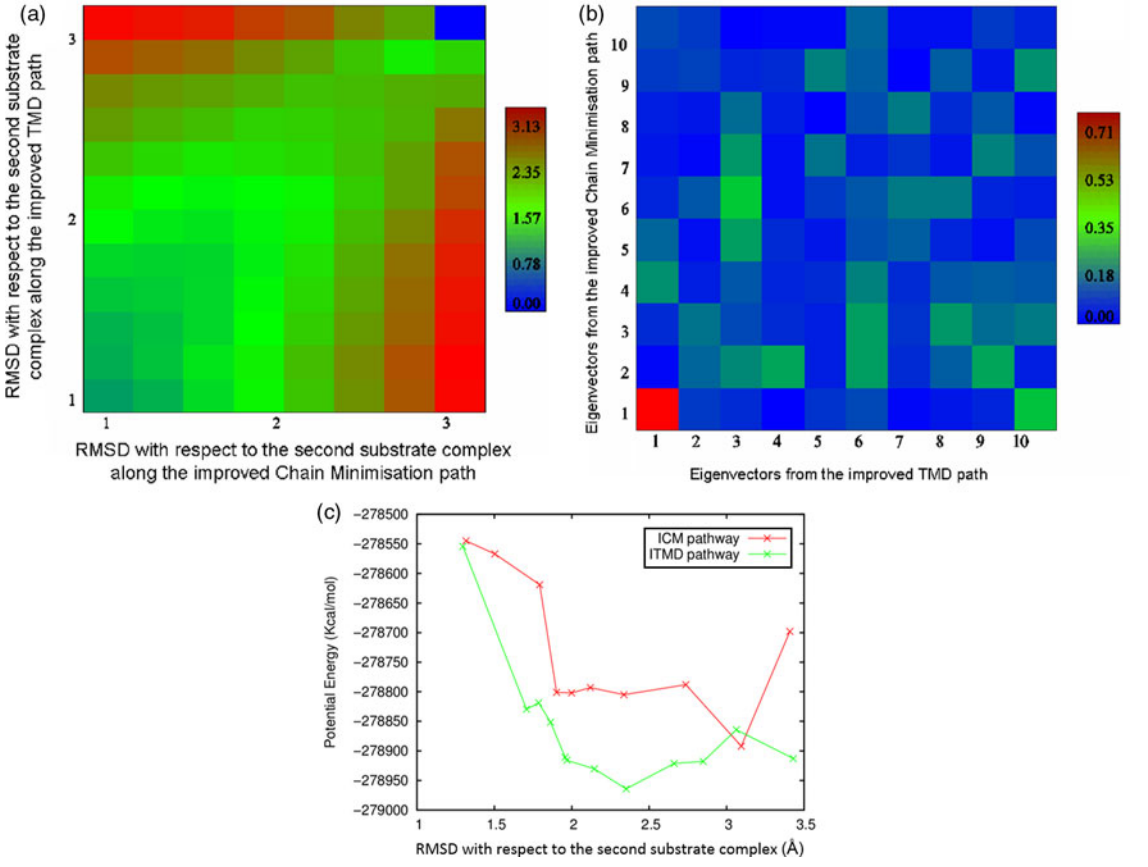

Figure 4. (Colour online) Comparisons between the ICM and improved TMD paths in the relaxation stroke. (a) The RMSD differences between these two paths. (b) Comparisons of eigenvectors determined from these two pathways. (c) The potential energy comparisons of these two pathways.

towards subdomain 2A and subdomain 1B is stable. Along the ITMD path, subdomain 2A performs a rotation motion, subdomain 1A rotates itself towards subdomain 2A and subdomain 1B is stable.

4. Conclusions and discussion

Two different methods (CM and auto-TMD) were applied to generate the initial reaction pathways of PcrA unwinding DNA, and then fixed-TMD was performed to relax and improve these two paths. The ICM and ITMD pathways share very similar major conformational changes (visualised by using PCA and relate to the nature of the first eigenvectors), but along the second eigenvectors there are some differences, which lie in the different motions of the subdomains within PcrA. In the power stroke, the most significant difference is obtained from the motions of subdomain 2A, and this difference will not influence the PcrA unwinding mechanism significantly. In the relaxation stroke, the most significant difference is obtained from the motions of subdomain 2B, although the influence of this difference is not clear. The energy profiles of these two pathways were identified, and in general the two profiles are quite similar, but the energy profile of the ITMD path is slightly lower than that of the ICM path.

In this study, the potential energy profiles for the two pathways are quite similar and help us understand how PcrA utilises the energy of ATP binding and hydrolysis to unwind the ds-DNS. The ATP binding site of PcrA is located in the cleft between subdomain 1A and subdomain 2A, and is in the middle of two important positively charged residues, namely, ARG287 on subdomain 2A and LYS37 on subdomain 1A. As both ARG287 and LYS37 show positive charges and ATP shows negative charges, in the substrate complex, ATP pulls these two residues close to it so that the cleft between subdomains 1A and 2A is closed and installs energy into the structure. After ATP is hydrolysed, ADP leaves the binding site and at the same time the positively charged ARG287 and LYS37 begin to 'see' each other, which drives the ATP binding site to open. The energy caused by the electrostatic repulsions is offset by opening of the binding site, and it is utilised by PcrA to separate the ds-DNA. In the meantime, the conformation of PcrA gradually changes to the product complex, which is a lower energy state. The electrostatic interactions between ARG287 and LYS37 are supported

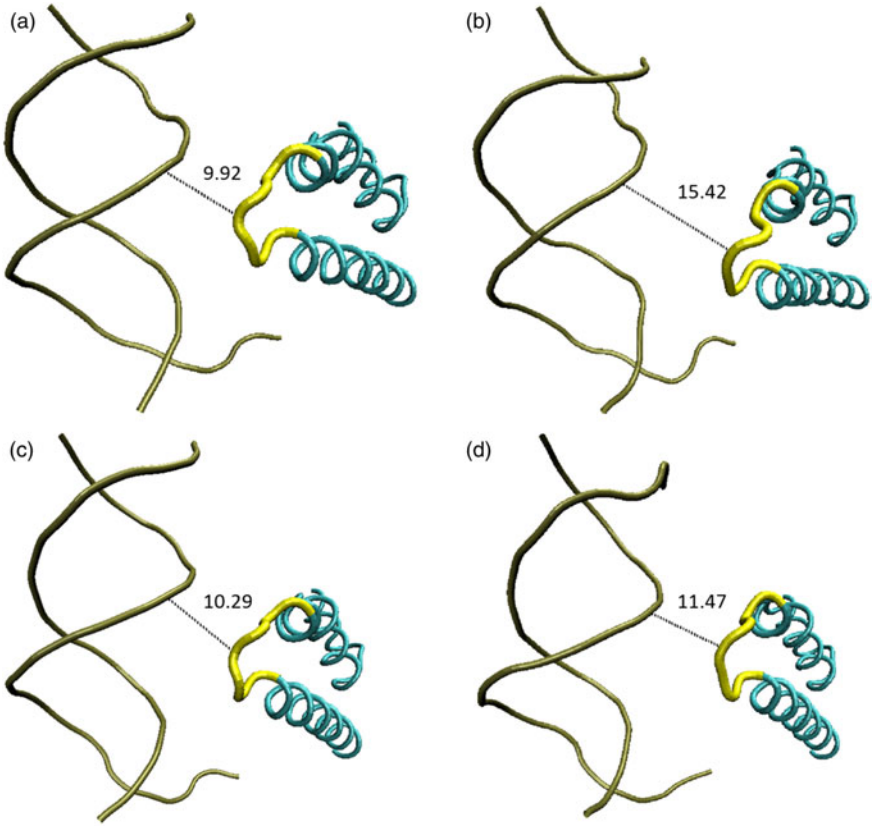

Figure 5. (Colour online) In the relaxation stroke, along the second eigenvector of the ICM path, at the start, the distance between loop A499–A505 (in yellow) on subdomain 2B and the ds-DNA (in tan) is 9.92 Å (a). Because of the rotation motion of subdomain 2B with respect to subdomain 1B, it is increased to 15.42 Å (b) and then comes back to its original value (a). Along the second eigenvector of the improved TMD path, this distance is stable in a small region, which is from 10.29 Å (c) to 11.47 Å (d).

by the comparison of the crystal structures of the substrate and the product complexes. In the substrate complex, the distance between the positive charge centres (CZ on ARG287 and CE on LYS37) is 6.56 Å, while in the product complex, it is 8.54 Å.

We found that the original CM path was very close to the linear interpolation path (data not shown), and the fixed-TMD simulation significantly improved it. However, for the auto-TMD path, such improvement was not so dramatic. In order to test the reliability of the initial auto-TMD pathway with 1 ns for each stroke, longer auto-TMD simulations (10 ns for each stroke) with the same settings were performed. The PCA was performed on all atoms in the system and then the dot product was calculated for each stroke between the long and short simulation path. Very high similarities were identified (for the power stroke, the dot product along the first vectors is 0.7, and along the second vector it is 0.62; for the relaxation stroke, along the first vectors the dot product is 0.71, and along the second vectors it is 0.50), which shows that the initial path generated by short auto-TMD is relatively reliable.

Taken together, our results suggest that if the initial path is not so reliable (e.g. the CM pathway), fixed-TMD has the ability to improve it to become a more 'correct' path. However, if the initial path is more reliable (e.g. the auto-TMD path), fixed-TMD will not change the path significantly. From this angle, the fixed-TMD method is independent of the initial pathways. However, when we compare the two improved paths, the only difference between them is that the different initial paths are used as the inputs for further improvements. This proves that the results of these two methods are somehow dependent on the initial pathways, but the difference only affects the minor submotions of PcrA. These two 'conflicting' conclusions suggest that there are some influences of the initial paths for fixed-TMD to generate reliable reaction paths, but such influences should not be over estimated. In general, fixed-TMD can generate a reliable and low-energy path regardless of the quality of initial paths. Therefore, fixed-TMD can be a powerful tool to study the reaction mechanism of large biomolecules, but the accuracy of detailed conformational changes still needs more careful study.

Although fixed-TMD generates a series of equilibrium windows (belongs to the general category of umbrella sampling method), it cannot be used to calculate the free

energy or potential mean force profile along the reaction path, because when the simulation reaches the low RMSD region with respect to the target, there is an artificial entropy reduction.[32] There are some recently developed methods, such as adaptively biased molecular dynamics (ABMD) [33,34] and steered molecular dynamics (SMD) using the Jarzynski relationship,[35] which can generate relatively reliable free energy profiles along non-equilibrium MD simulations. Both ABMD and SMD require that one or more collective variables be specified for restraint, which can be distances, angles, torsions or even RMSDs of several groups of atoms with respect to reference positions, etc., but they do not restrain the whole structure like TMD. Therefore, they can calculate the free energy by avoiding 'freezing' the system.[36] However, in this study, it is extremely difficult (if not impossible) to describe the conformational changes of PcrA by several groups of variables. In other words, we could not find such variables to confirm that the structure has matched the reference except by using RMSD. Therefore, it is still an important and open question as to how to estimate the artificial entropy caused by restraining the system in a small RMSD region.

Acknowledgements

This work was supported by the University of Nottingham; National Natural Science Foundation of China [grant number 81260481]; Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry; Independent scientific research foundation of Beifang University of Nationalities [grant number 2011ZQY035] and University of Malaya-MOHE High Impact Research [grant number UM.C/625/1/HIR/MOHE/DENT/22].

References

[1] Anand SP, Khan SA. Structure-specific DNA binding and bipolar helicase activities of PcrA. Nucleic Acids Res. 2004;32:3190–3197.
[2] Hall MC, Matson SW. Helicase motifs: the engine that powers DNA unwinding. Mol Microbiol. 1999;34:867–877.
[3] Subramanya HS, Bird LE, Brannigan JA, Wigley DB. Crystal structure of a DExx box DNA helicase. Nature. 1996;384:379–383.
[4] Bird LE, Brannigan JA, Subramanya HS, Wigley DB. Characterisation of *Bacillus stearothermophilus* PcrA helicase: evidence against an active rolling mechanism. Nucleic Acids Res. 1998;26:2686–2693.
[5] Caruthers JM, McKay DB. Helicase structure and mechanism. Curr Opin Struct Biol. 2002;12:123–133.
[6] Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell. 1999;97:75–84.
[7] Dillingham MS, Wigley DB, Webb MR. Demonstration of unidirectional singlestranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. Biochemistry. 2000;39:205–212.
[8] Dillingham MS, Wigley DB, Webb MR. Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. Biochemistry. 2002;41:643–651.
[9] Yu J, Ha T, Schulten K. Structure-based model of the stepping motor of PcrA helicase. Biophys J. 2006;91:2097–2114.
[10] Wroblowski B, Diaz JF, Schlitter J, Engelborghs Y. Modelling pathways of alphachymotrypsin activation and deactivation. Protein Eng. 1997;10:1163–1174.
[11] Tobias D, Brooks CB III. Molecular dynamics with internal coordinate constraints. J Chem Phys. 1988;89:5115–5127.
[12] Harvey S, Gabb H. Conformational transitions using molecular dynamics with minimum biasing. Biopolymers. 1993;33:1167–1172.
[13] Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol. 1999;288:441–459.
[14] Schlitter J, Engels M, Krger P, Jacoby E, Wollmer A. Targeted molecular dynamics simulation of conformational change: application to the T ↔ R transition in insulin. Mol Simul. 1993;10:291–308.
[15] Apostolakis J, Ferrara P, Caflisch A. Calculation of conformational transitions and barriers in solvated systems: application to the alanine dipeptide in water. J Chem Phys. 1999;110:2099–2108.
[16] Kruger P, Verheyden S, Declerck PJ, Engelborghs Y. Extending the capabilities of targeted molecular dynamics: simulation of a large conformational transition in plasminogen activator inhibitor 1. Protein Sci. 2001;10:798–808.
[17] Lee HS, Robinson RC, Joo CH, Lee H, Kim YK, Choe H. Targeted molecular dynamics simulation studies of calcium binding and conformational change in the cterminal half of gelsolin. Biochem Biophys Res Commun. 2006;342:702–709.
[18] Mendieta J, Gago F, Ramirez G. Binding of 5′-GMP to the GluR2 AMPA receptor: insight from targeted molecular dynamics simulations. Biochemistry. 2005;44:14470–14476.
[19] Mendieta J, Martin-Santamaria S, Priego EM, Balzarini J, Camarasa MJ, Perez-Perez MJ, Gago F. Role of histidine-85 in the catalytic mechanism of thymidine phosphorylase as assessed by targeted molecular dynamics simulations and quantum mechanical calculations. Biochemistry. 2004;43:405–414.
[20] Rodriguez-Barrios F, Balzarini J, Gago F. The molecular basis of resilience to the effect of the Lys103Asn mutation in non-nucleoside HIV-1 reverse transcriptase inhibitors studied by targeted molecular dynamics simulations. J Am Chem Soc. 2005;127:7570–7578.
[21] Aci S, Mazier S, Genest D. Conformational pathway for the kissing complex → extended dimer transition of the SL1 stem-loop from genomic HIV-1 RNA as monitored by targeted molecular dynamics techniques. J Mol Biol. 2005;351:520–530.
[22] Compoint M, Picaud F, Ramseyer C, Girardet C. Targeted molecular dynamics of an open-state KcsA channel. J Chem Phys. 2005;122:134707.
[23] Ferrara P, Apostolakis J, Caflisch A. Computer simulations of protein folding by targeted molecular dynamics. Proteins. 2000;39:252–260.
[24] Kamerlin SC, Rucker R, Boresch S. A targeted molecular dynamics study of WPD loop movement in PTP1B. Biophys Res Commun. 2006;345:1161–1166.
[25] Rodriguez-Barrios F, Gago F. Understanding the basis of resistance in the irksome Lys103Asn HIV-1 reverse transcriptase mutant through targeted molecular dynamics simulations. J Am Chem Soc. 2004;126:15386–15387.
[26] van der Vaart A, Karplus M. Simulation of conformational transitions by the restricted perturbation-targeted molecular dynamics method. J Chem Phys. 2005;122:114903-1–114903-6.
[27] Balsera MA, Wriggers W, Oono Y, Schulten K. Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. Phys Rev E. 1996;69:1–13.
[28] Sherer EC, Harris SA, Soliva R, Orozco M, Laughton CA. Molecular dynamics studies of DNA A-tract structure and flexibility. J Am Chem Soc. 1999;121:5981–5991.
[29] Wang H, Laughton CA. The inhibitory effects of vinylphosphonate-linked thymidine dimers on the unidirectional translocation of PcrA helicase along DNA: a molecular modelling study. Phys Chem Chem Phys. 2012;14:12230–12237.
[30] Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. J Chem Phys. 1993;98:10089–10092.

[31] Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81:3684–3690.
[32] Schlitter J. Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. Chem Phys Lett. 1993;215:617–621.
[33] Babin V, Roland C, Sagui C. Adaptively biased molecular dynamics for free energy calculations. J Chem Phys. 2008;128:134101.
[34] Babin V, Karpusenka V, Moradi M, Roland C, Sagui C. Adaptively biased molecular dynamics: an umbrella sampling method with a time-dependent potential. Int J Quant Chem. 2009;109:3666–3678.
[35] Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys. 2003;119:3559.
[36] Moradi M, Babin V, Roland C, Sagui C. Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. Nucleic Acids Res. 2013;41:33–43.