Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics

Supplementary Data

The inhibitory effects of vinylphosphonate-linked thymidine dimers on the unidirectional translocation of PcrA helicase along DNA: A molecular modelling study

Hao Wang and Charles A. Laughton

The System Setup:

The original crystal structures of PcrA, the substrate complex (3PJR) and the product complexes (2PJR), were obtained from the Protein Data Bank. The crystal structures feature a number of differences which must be resolved because the TMD process requires the topologies of all target structures to be identical. The variety of missing residues in the protein components of the crystal structures are not located in the DNA active site (mainly they are the residues mediating the conformation changes from the apo form to the substrate complex, and two residues near the C-terminal, Figure 1), so that if a residue was missing in one of the structures, it was also deleted from all other ones. For the DNA helix component the approach of generating models based on the largest

common subset of residues could not be used, the DNA helix in the substrate complex is complete, but many of the DNA bases in the product complex are missing. To circumvent this, by using the protein residues and DNA bases that are not missing in the product complex as references, a restrained MD simulation was applied to the substrate complex to predict the structure of the missing fragments of DNA in the product complex (Figure 2).

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 by moving the DNA in the original substrate complex one base forward towards the ends of two single strands (Figure 3).

The ds-DNA in the original crystal structure of the substrate complex only contains 10 base pairs. It is too short to show the interaction with the top of the protein [23]. For this reason, a part of the ds-DNA sequence (GCACT) in the crystalographically-determined structure of the substrate complex was copied and manual computer graphics manipulations used to extend the distal end of the ds-DNA in each of these three conformations (Figure 4).

As only the conformation of the protein and DNA is of interest in the research, the bound ATP, magnesium ion and γ phosphate group were removed. After this process each these three complexes contains 11439 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 these three structures were explicitly solvated in a truncated octahedral box of TIP3P model water, and 52 *Na+* ions were added to neutralize the charges of each system. Finally energy minimisations and MD simulations (500 ps each with our standard equilibration strategy [24]) were performed on these three systems to relax the DNA conformations in those parts of the structure which had needed 'remedial' model-building work, but which were remote from the protein cleft of interest. The stabilities of simulations were checked by RMSD plots before the production simulations were begun.

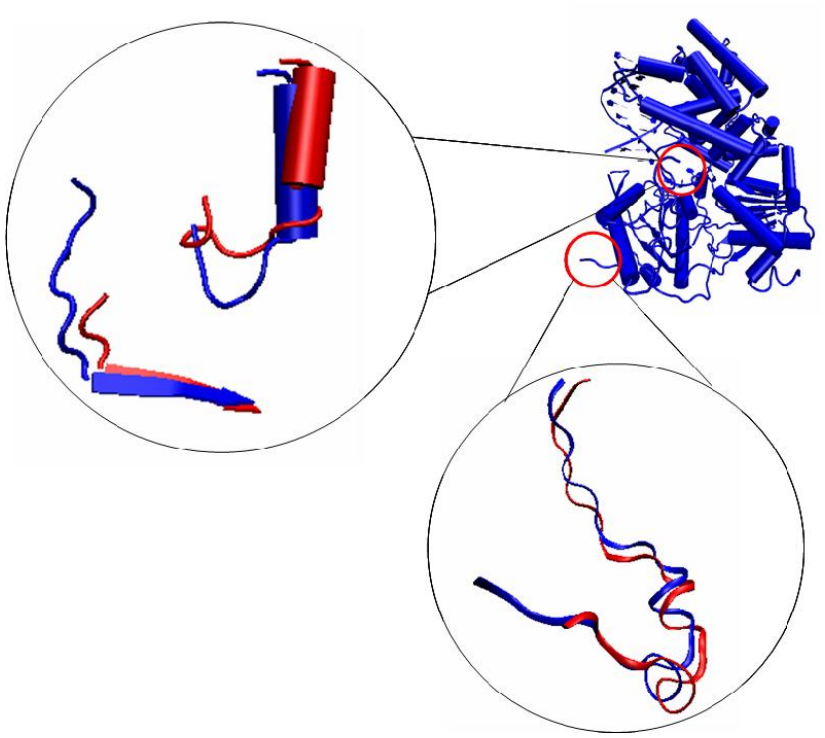


Figure 1: The missing residues in the protein components of PcrA. In PcrA, between the subdomains 2A and 2B, a piece of sequence contains nine residues, which are LDGTEQAAE. This part sequence is believed important for the conformational changes from the apo-form to the substrate complex. However, three successive residues (LDG) are missing in the substrate complex (blue) and six (TEQAAE) are missing in the product complex (red), as shown in the picture in the left circle. At the C terminal, the sequence of the product complex (red) is two residues (RR) shorter than the sequence of the substrate complex (blue).

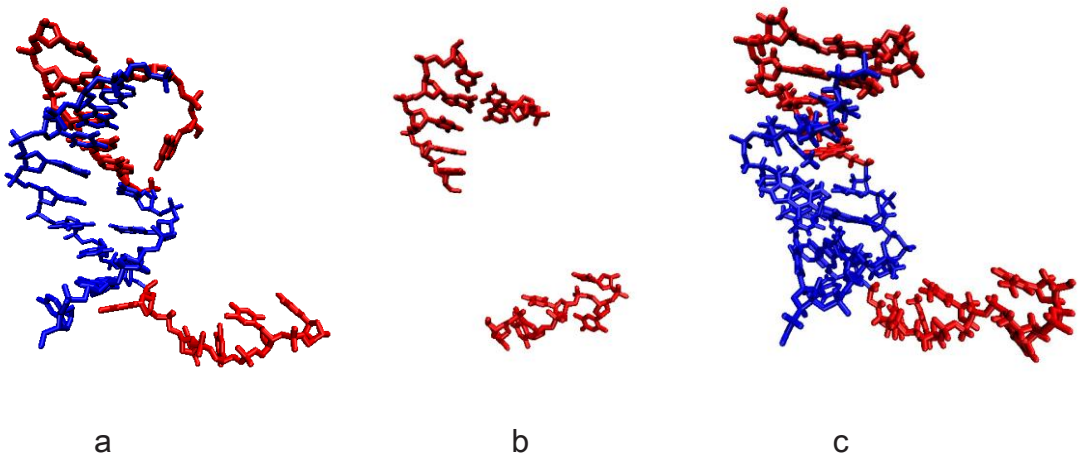


Figure 2: Rebuilding the DNA component of the product complex. (a): The DNA component of the substrate complex. Those bases in red are also contained in the product complex, but the bases in blue are missing. (b): The bases contained in the DNA component of the product complex, which were used as the reference to build the other missing bases. (c) The DNA component of the product complex after rebuilding.

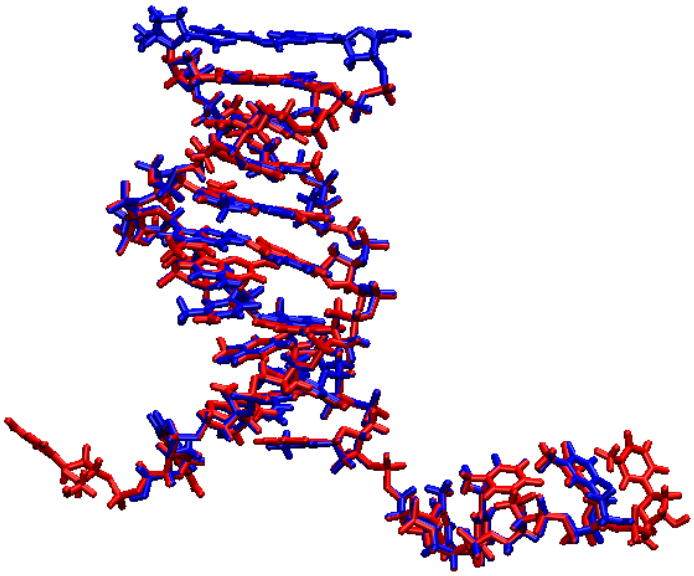


Figure 3: The DNA sequence in blue is the DNA component of the substrate complex. All the bases in it are moved one base forward towards the ends of

two single strands to build the DNA component of the second substrate complex, which is coloured in red.

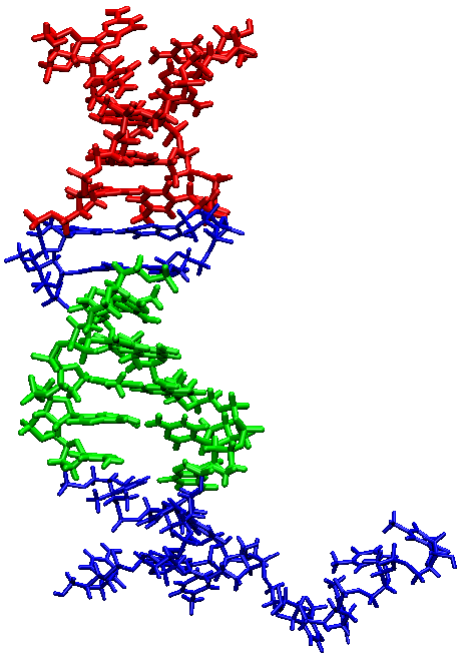


Figure 4: The part (green) of the ds-DNA sequence is copied and extended to the end of the original ds-DNA (blue). The extension part is coloured in red.

Unwinding Cycle 1

The first stage in the transit corresponds to the TMD simulation of the power stroke in which T5 (or VT5) moves from its position in the substrate complex to that in the product complex. As Figure 5 shows, for both the control simulation (*a*, *b* and *c*) and the VT simulation (*d*, *e* and *f*),

T5 and VT5 (ball-and-stick) drop into pocket 2 at the ends of TMD simulations.

As Figure 6 *a* shows, the *β* torsion angle of T5 is restricted by the modification. As Figure 6 *b* shows, for the control simulation, the distance between the N3 atom on T5 and the CZ atom on PHE626 increases quickly to 8.5 Å at 911 ps. While for the VT simulation, the corresponding distance increases at 739 ps. The distance between T5 and HIS587 is quite similar to the corresponding one for the VT simulation (Figure 6 *c*). As Figure 6 *d* shows, for the VT simulation, from 739 the distance between N3 on VT5 and NE1 on TRP259 suddenly decreases from around 15 Å to 7 Å. While, for the control simulation, the corresponding distance decreases to this level after 911 ps.

Visualisation of the trajectory reveals that the paths for T5 and VT5 dropping into pocket 2 are different. For the unmodified system, T5 drops into pocket 3 first after around 530 ps. Then at 911 ps, it flips out of pocket 3 and drops into pocket 2. For the modified system, VT5 moves towards pocket 3 at 378 ps, and then passes over the top of pocket 3 and moves towards pocket 2 at 739 ps. Finally it drops into pocket 2 at 790 ps.

As Figure 6 *e* shows, for the control simulation, during the period when T5 is located in pocket 3 (from around 530 ps to 911 ps), the energy decreases to a low value, and when it flips out of pocket 3 and drops into pocket 2, the energy does not change significantly. As a result, the energy when T5 is located in pocket 2 is almost the same as the one when it is in pocket 3. For the VT simulation, VT5 is located above the top of pocket 3 in the period from 378 ps to 739 ps. In this period, the energy just fluctuates at around -10 Kcal/mol, and after this period, the energy starts to reduce dramatically. This result shows when VT5 is located above pocket 3, it is in a higher energy state, and this high energy is offset only after VT5 drops into pocket 2.

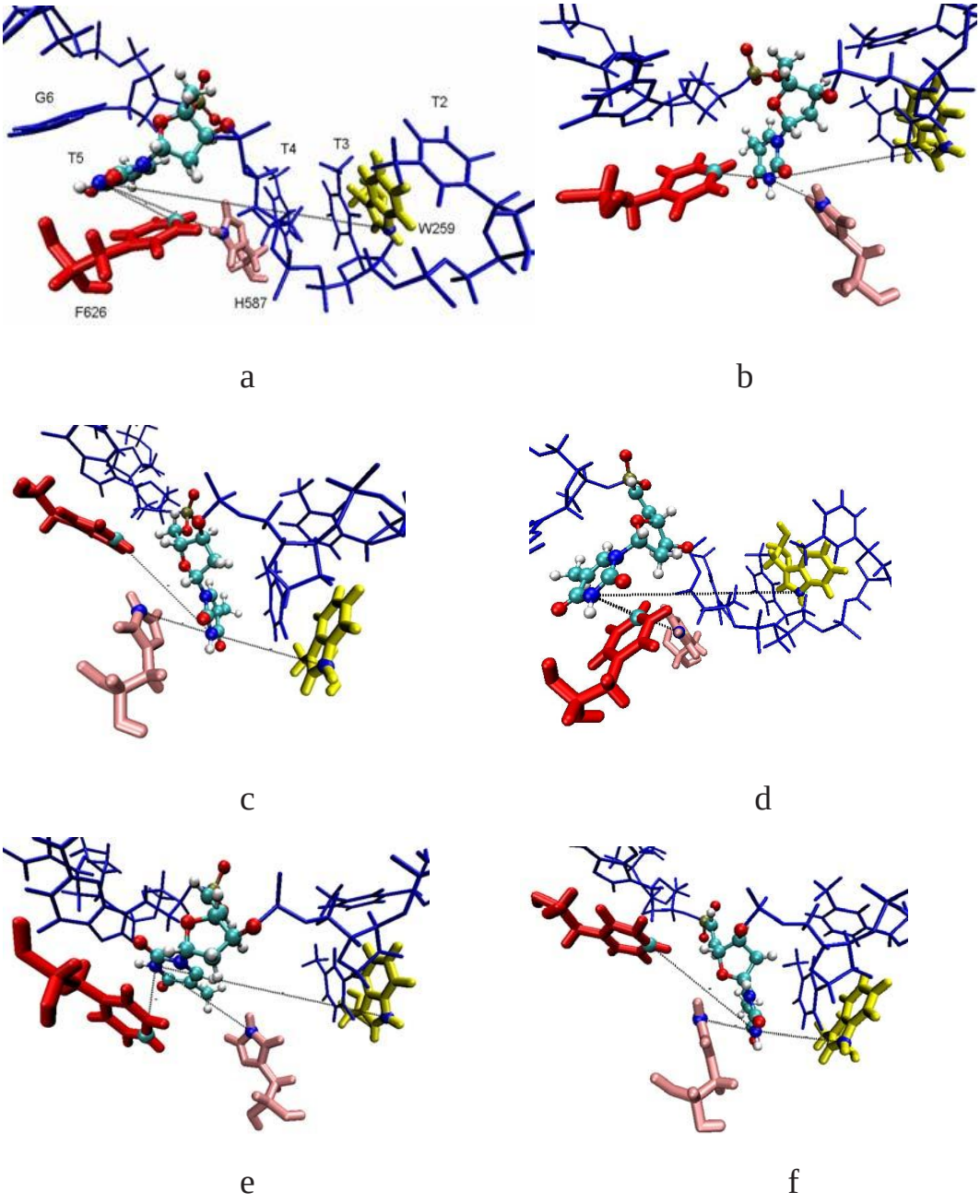
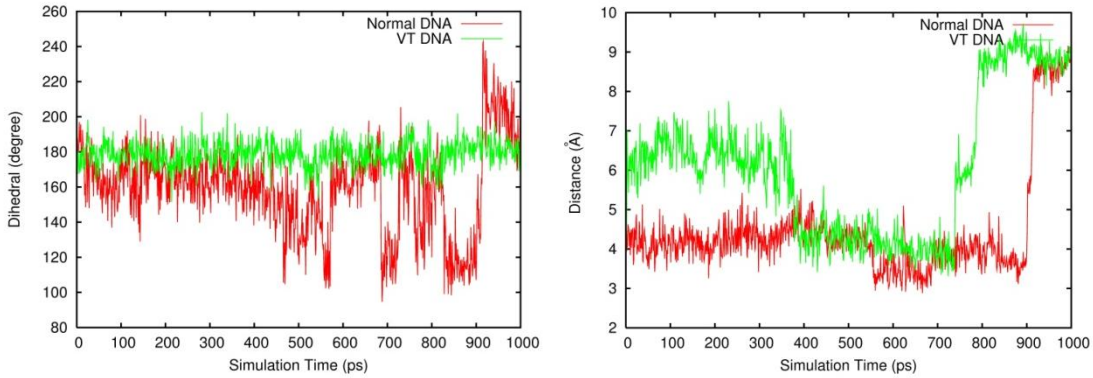
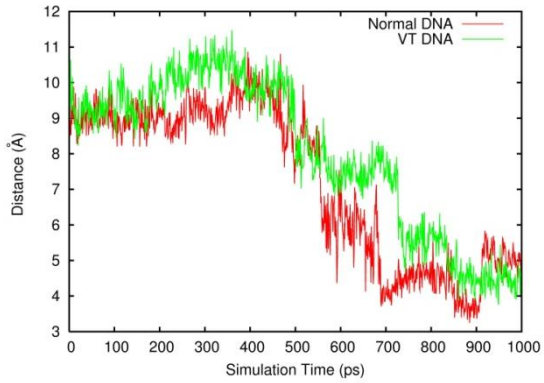


Figure 5: Conformational comparisons between the control and VT simulations of the power stroke when T5 is modified. The ss-DNA is coloured in blue, and T5/VT5 is highlighted in a ball-and-stick representation and coloured by atoms. The dotted lines indicate the distances between N3 on T5/VT5 and the atoms on the relative residues. (a) The starting conformation in the control simulation. (b) The transition conformation in the control simulation. (c) The end conformation in the control simulation. (d) The starting conformation in the VT simulation. (e) The transition conformation in the VT simulation. (f) The end conformation in the VT simulation.

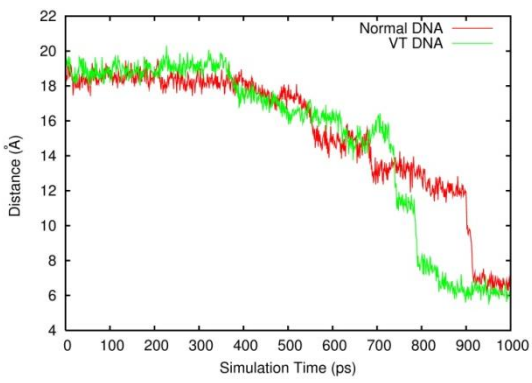


a

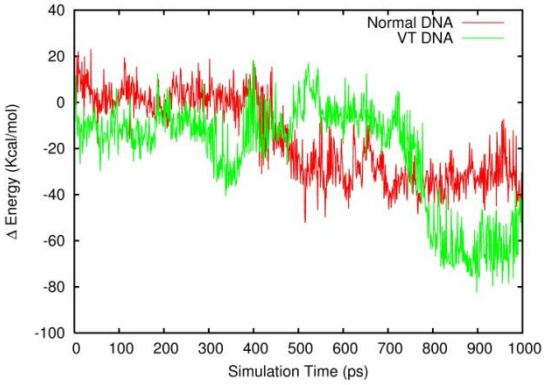
b



c



d



e

Figure 6: Quantitative comparisons between the control and VT simulations of the power stroke when T5 is modified. (a) The comparison of the β torsion angles between the control and VT simulations. (b) The comparison of the distances between the N3 atom on T5/VT5 and the CZ atom on PHE626. (c) The comparison of the distances between the N3 atom on T5/VT5 and the NE2 atom on HIS587. (d) The comparison of the distances between the N3 atom on T5/VT5 and the NE1 atom on TRP259. (e) The comparison of the sum of the intra energy of T5/VT5 and the interaction energies between T5/VT5 and the other groups in the binding site.

Unwinding cycle 3

Unwinding cycle 3 involves a power stroke during which the base of interest (now T3, or VT3) leaves pocket 2. As Figure 7 shows, for both the control simulation (*a* and *b*) and the VT simulation (*c* and *d*), T3 and VT3 flip out of pocket 2, and move close to ARG260 at the ends of the TMD simulations.

As Figure 8 *a* shows, the modification restricts the free rotation of the β torsion angle of VT3. For the distances between the N3 atom on T3 and the NE1 atom on TRP259 (Figure 8 *b)*, and between the N3 atom on T3 and the CZ atom on ARG260 (Figure 8 *c)*, the shapes of the curves in the control and experiment system are almost identical. As Figure 8 *d* shows, the two energy curves of the control and VT simulations are quite similar to each other.

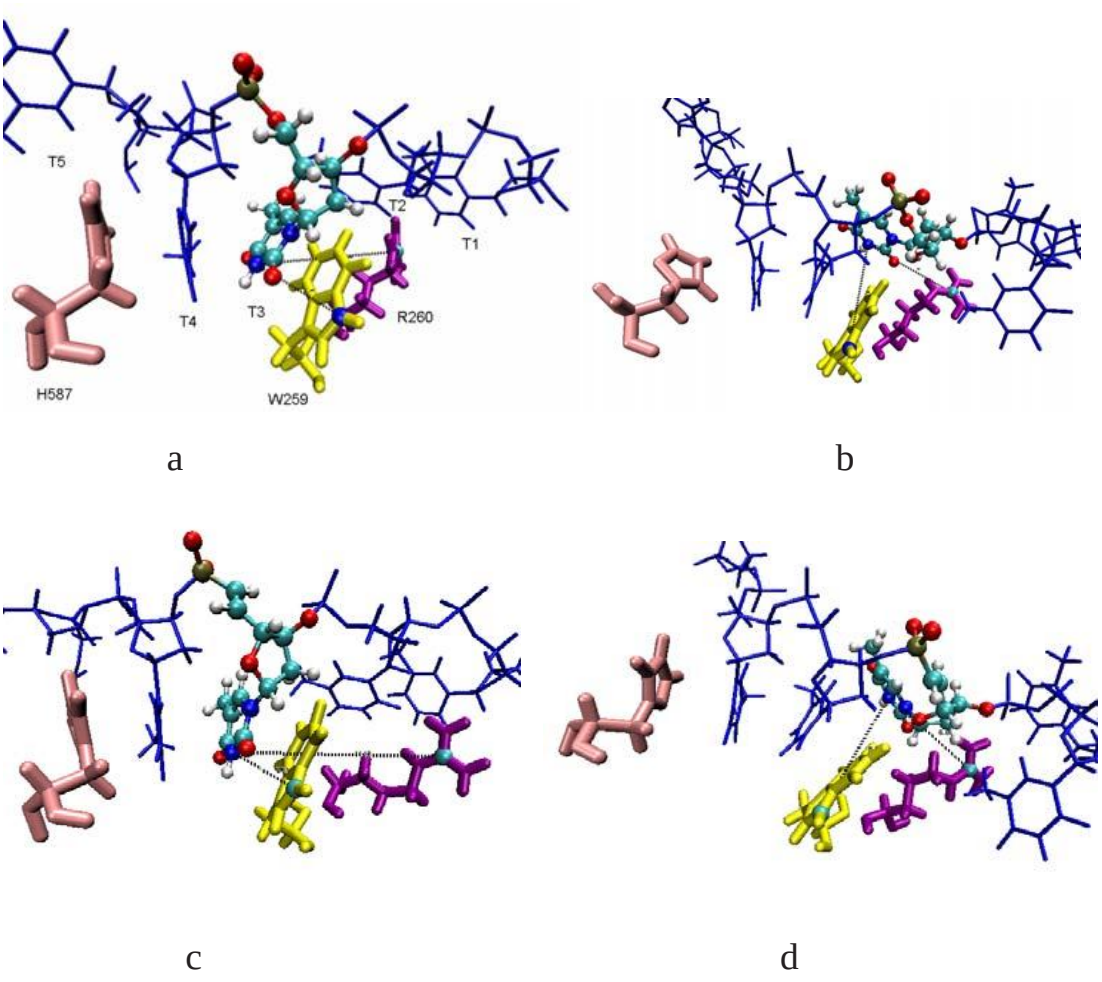


Figure 7: Conformational comparisons between the control and VT simulations of the power stroke when T3 is modified. The ss-DNA is coloured in blue, and T3/VT3 is highlighted in a ball-and-stick representation and coloured by atoms. TRP259 is coloured in yellow and the NE1 atom on it is highlighted. ARG260 is coloured in purple and the CZ atom on it is highlighted. HIS587 is coloured in pink. The dotted lines indicate the distances between the N3 atom on T3/VT3 and the NE1/CZ atom on TRP259/ARG260. (a) The starting conformation in the control simulation. (b) The end conformation in the control simulation. (c) The starting conformation in the VT simulation. (d) The end conformation in the VT simulation.

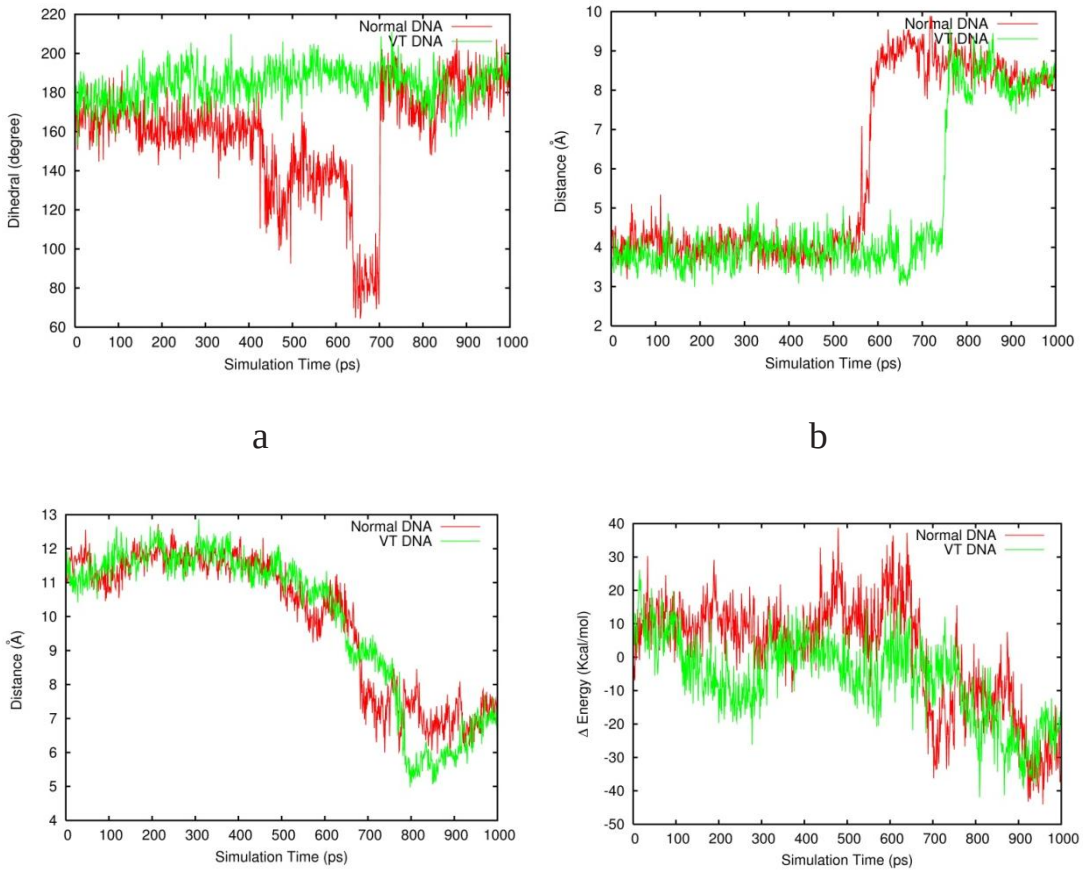


Figure 8: Quantitative comparisons between the control and VT simulations of the power stroke when T3 is modified. (a) The comparison of the β torsion angles between the control and VT simulations. (b) The comparison of the distances between N3 on T3/VT3 and NE1 on TRP259. (c) The comparison of the distances between N3 on T3/VT3 and CZ on ARG260. (d) The comparison of the sum of the intra energy of T3/VT3 and the interaction energies between T3/VT3 and the other groups in the binding site.

Unwinding cycle 4

Presented are two movies (T2-VT2-stag4a and T2-VT2-stag4b), which show the process when T2 and VT2 drop into (during 'power' stroke) and flip out (during 'relaxation' stroke) the pocket 1 respectively. In both movies, the transitions of T2 are on the top and the transitions of VT2 are on the bottom.