Supplementary material

S.1 Methods

The model system we use is an A-family high fidelity polymerase from *Bacillus stearothermophilus*: Bacillus Fragment (BF). The enzyme exhibits a high efficiency (200 base pairs/sec) and processivity (111 nucleotide bases) for accurate (fidelity of 10^{-8}) DNA replication. Crystals of BF are catalytically active (the enzyme can synthesize base pairs in crystal) and this property has been exploited by Beese et. al. to obtain high resolution crystal structures of the enzyme at a number of points along the replication cycle[2,3]. In particular, high resolution crystal structures and kinetic data for correct/ incorrect nucleotide incorporations both with undamaged and oxidatively damaged substrates are readily available for BF.

A-family polymerases are very similar in structure with a highly conserved active. BF is homologous (49 % sequence homology) and also shows extensive structural similarity (root-mean-squared-deviation or RMSD of C_α atoms is 0.64 Å) to the active site of KF and other A-family polymerases[4]. Figure S1 shows the sequence and structural homology for BF with the Klenow fragment and T7 DNA pol I. In Fig. S1, while comparing different protein structures the reported measure Q tells us how identical two structures are. Q is defined as the fraction of similar native contacts between aligned residues in two proteins [5]. A Q value of 1 indicates that the two protein structures are identical while a Q value of 0 indicates that the structures are completely different and no alignment is possible. Often a Q value is calculated for each residue and is denoted as Q_R We use this parameter to compare BF with KF and T7 DNA. Figure S1 shows that the interior regions of the polymerases (especially near the active site) tend to have high Q values. Further even the identities of residues are the same (highly conserved) in all three systems for our choice of the active site fragment as shown. This is characteristic of A-family polymerases and in fact it has been shown using point mutation experiments that the catalytic activity of the polymerase is directly influenced only by a few residues (< 10) at the polymerase active site and that these residues are

conserved across the whole A-family [7]. Thus we expect that our results have implications beyond the model system chosen and would provide insights for the behavior of A family polymerases.

S.1.1 System preparation

We prepared four model systems **G:**d**CTP or G:C** which represents the ternary complex of BF pol with DNA and dCTP opposite the template G in the closed or active state, **G:A**, 8oxo**G:C** and 8oxo**G:A** (see Figure S2) using the insight II modeling software[8] , starting from the crystal structure of a closed ternary BF-DNA-dCTP complex (PDB id: 1LV5[2]). These correspond to cases of correct/incorrect nucleotide incorporation opposite an undamaged/oxidatively damaged **G** template base respectively. The Mn^{2+} ion at the catalytic site in the crystal structure was replaced with a Mg^{2+} ion. Crystallographic waters were discarded. Missing atoms in the crystal structure were added including the terminal primer **A** O3′. For the mispair the incoming dCTP in 1LV5 was replaced with a dATP. For the oxidative damage cases the **G** base in the **G:C** and **G:A** models were modified to 8oxo**G** by adding oxygen and hydrogen atoms at C8 and N7 respectively and by modifying the double bond between C8 and N7 to a single bond. Hydrogen atoms were added to the models using the HBUILD[9] utility in CHARMM with HIS protonation states chosen according recommendations from the WHATIF web interface (http://swift.cmbi.kun.nl/WIWWWI). Protonation states for all other charged groups were chosen according to their pKa values in aqueous solution[10] at a pH of 7.0 (ASP→-1, GLU→-1, LYS→+1, ARG→+1). The models were then solvated using SOLVATE 1.0[11] which also neutralizes the system by placing Na^+ and Cl^- ions at isotonic concentrations (0.154 mol/l), with a Debye-Huckel distribution at 300 K. A total of 98 Na^+ and 66 Cl^- ions were added to neutralize the systems.

The **G:C** system is directly derived from the 1LV5 crystal structure[2]. The other three models **G:A**, 8oxo**G:C** and 8oxo**G:A** were constructed by replacing the **G** with 8oxo**G** and/or replacing the incoming dCTP with a dATP. An important issue to consider during the modeling was the conformation of the

template **G**/8oxo**G** opposite the incoming nucleotide for the latter three systems for which no BF/DNA/dNTP ternary complexes have been crystallized. For correct nucleotide incorporation opposite undamaged DNA substrates, it is known that in the open (inactive) state of the enzyme prior to nucleotide insertion that the template base is in a syn conformation (characterized by a glycosidic torsion angle $\chi = 0$ degrees between the sugar and base groups). Sometime during the nucleotide insertion stage the template base switches to an anti ($\chi = 0$ degrees) conformation which is preserved and observed in post-insertion structures. However the template base can adopt the syn conformation in the event of a mispair and/or in the event of damage[3,12,13]. While there are no crystal structures for a closed ternary complex of BF with an 8oxo**G:**dCTP pair at the active site, a ternary complex of T7 pol I/8oxo**G**DNA/dCTP (prior to catalysis) shows the lesion carrying template base at the active site to be in an anti conformation[13]. However, there are no crystal structures for a BF/DNA/dNTP ternary complex with either **G:A** or 8oxo**G:A** mispair at the active site. Crystal structures of oligonucleotide sequences show that **G** can adopt either a syn or anti conformation opposite an incoming dATP[3] while structures of post-insertion complexes in BF [12], pol β [14], and T7 DNA pol [13] find the template base carrying the lesion in a syn conformation in 8oxo**G:A** systems. We thus carried out simulations for a **G:A** system with the template **G** in a syn and anti conformations and modeled the 8oxo**G:C** and 8oxo**G:A** systems in anti:anti and syn:anti conformations respectively. For the **G:A** system the anti:anti simulations were stable but for the syn:anti simulations we observed an syn→anti template flip with a fast timescale of 800 ps (see Figure S3). This result shows that an anti conformation is reached pre-chemistry and that the base flipping reaction is not rate limiting even in misincorporation reactions for BF. As proposed on basis of indirect kinetic evidence[15] the rate limiting step for mismatch reactions is most-likely the catalysis step owing to a distorted pairing geometry between the template **G** (in an anti conformation) and the incoming dATP. In contrast simulations for 8oxo**G** lesion with an incoming dATP were stable with a syn conformation for the template base which reduces the distortion of the catalytic site thereby significantly enhancing the rate of the misincorporation. We also initiated unconstrained molecular dynamics trajectories for the 8oxo**G:C** and 8oxo**G:A** systems

with the glycosidic torsion restricted to χ=90°, (i.e., close to the presumed transition state between *syn* and *anti* conformations) during the equilibration phase. These simulations showed that the 8oxo**G** template base adopts an *anti* conformation opposite an incoming dCTP (Fig S1 bottom left insert) and a *syn* conformation opposite an incoming dATP (Fig S1 bottom right insert); see also Fig S2. Taken together with the existing structural evidence, these imply that a *syn* conformation of the lesion opposite a dATP and an anti conformation of the lesion opposite dCTP are most-likely to be the stable ground states, thereby validating our model systems.

S.1.2 Forcefield parameterization

The CHARMM27[16] forcefield was used to perform MD simulations. Parameters (partial charges for nonbonded interactions and force constants for bonded interactions) compatible with CHARMM27 for the 8oxo**G** residue were constructed as described by Foloppe et. al.[17]. Partial charges were assigned and refined to reproduce ab-initio 8oxo**G** dipole moments, base-water dimer interaction energies and distances. These values were then used in a genetic algorithm based optimization scheme developed in our lab (Y. Liu, R. Radhakrishnan, unpublished) to construct and refine the CHARMM force field parameters for bond, angle and dihedrals of the 8oxo**G** residue to reproduce ab-initio vibrational frequencies. The new parameters thus obtained, were then used to refine the partial charges further, and the entire procedure was repeated until convergence was reached. The resulting root-mean-squared deviation (RMSD) of $\sigma = 79.78$ cm^{-1} between the newly parameterized CHARMM normal mode frequencies and ab-initio vibrational frequencies is within acceptable limits for small molecules [17]

S.1.3 Simulation protocols

The NAMD simulation package [18,19] with the CHARMM27 force field was used to minimize and equilibrate each model system and for subsequent production runs. The model systems were enclosed in a solvent box (dimensions 111 Å x 91 Å x 95 Å) of 27068 water molecules and periodic boundary

conditions were applied. A 12.0 Å cutoff was applied for non-bonded interactions wherein a switching potential was turned on at 10.0 A. The particle mesh Ewald method[20] was used for the treatment of long range electrostatics. The rigidbonds option (i.e., the rattle algorithm) was used to constrain all bonds involving hydrogen atoms to their values in the CHARMM parameter file. The equilibration protocol for each system was as follows: systems were subjected to two initial rounds of minimization (10000 steps), heating from 0-300K (50000 steps) and NVT equilibration (50000 steps) with 1fs timesteps. The protein and DNA fragment was held fixed in the first round while the second round was unconstrained. Subsequently an NPT equilibration (with a 2 fs timestep) was carried out to obtain the correct density/box size for each system. Finally a 100 ps NVT equilibration run was carried out to arrive at the equilibrated configuration. Following the equilibration 10 ns NVT production runs were carried out. The RMSD of the protein backbone was monitored and data from the last 5ns during which the rmsd was found to be stable (Figure S4) was used for subsequent analysis. The program VMD[21] to visualize and analyze our simulation results as well as to render images of the protein structures.

S.1.4 Principal Component Analysis

Principal component analysis (PCA) [22,23] of MD simulations provides us with a framework to project out independent motions in an MD trajectory and sort them in the order of their dominance (the strongest motions first). This is achieved by diagonalizing the variance-covariance matrix of atomic fluctuations along the trajectory.

PCA solves the eigenvalue equation: $[\sigma - \lambda \mathbf{I}]\xi = 0$ to project out principal components (PC) or independent modes of atomic motion, captured in an MD trajectory and sorts them by their variance (in decreasing order). Here σ is a two dimensional variance-covariance matrix of atomic fluctuations about the trajectory average, with elements $\sigma_{ij}=\langle(x_i-\langle x_i\rangle)(x_j-\langle x_j\rangle)\rangle$ ($i,j =1,...,3N$, N being the total number of atoms with position given by Cartesian coordinates x);. $\xi=(\xi_1,\xi_2,...,\xi_{3N})$ are the 3N independent (uncorrelated) eigenvectors (PC) with eigenvalues $\lambda=(\lambda_I, \lambda_2,...,\lambda_{3N})$ sorted in descending order i.e.

$\lambda_1 > \lambda_2 \ldots, \lambda_{3N-7} > \lambda_{3N-6}$. All global translations/ rotations about the center of mass are removed prior to evaluating σ and the six eigenvalues corresponding to these degrees of freedom are close to zero.

The resulting eigenvectors are the uncoupled principal components (PCs), (modes orthogonal to each other) and the eigenvalues reflect their magnitude (strength) in the trajectory. Since the formalism requires a well-defined average geometry as a reference around which the variance-covariance matrix of atomic fluctuations will be constructed, we chose the average geometry of the ternary complex with bound waters as the reference. The PCA calculation was performed for a small region around the catalytic geometry (denoted as the active site region) which included all heavy atoms of the incoming dNTP, six residues of the DNA template strand (including the template **G**/8oxo**G** of the nascent base pair), four residues of the DNA primer strand (including the terminal **A**), the two Mg^{2+} ions, two highly conserved polymerase aspartate residues D830 and D653 which coordinate the Mg^{2+} ions and bound waters at the catalytic site. The software program CARMA [24] was used to perform PCA on our system. CARMA also enables us to visualize principal modes by projecting out the atomic fluctuations due to the modes along the MD trajectory. The top 10 principal component modes contained most of the atomic fluctuations in the MD trajectory for all systems studied (70% for G:C, 72 % for 8oxoG:C and 80% for 8oxoG:A). See Figures S8-S11 for pairwise correlations between motions of atoms in an extended active site fragment including the polymerase fingers. These results show that the extent of correlations in the active site region is highly context specific.

S.2 Choice of subset (active site region) for PCA

The active site region chosen for PCA includes all residues which participate in the phosphoryl transfer reaction as well as the template strand on which the external force acts. This region represents the site of DNA polymerase interactions and includes the pre-insertion, insertion, post-insertion and DNA duplex binding regions of the BF-DNA-dNTP complex as defined by Johnson and Beese[2]. We have found in our studies on BF[25] that the dominant modes of this subset (for the **G:C** system) show strong correlations

with catalytic site reactive distances and since the these modes are delocalized over the whole fragment (see Movie S6), this implies that significant couplings exist between motions of atoms in the DNA template strand and those at the catalytic site. As a direct consequence of this coupling we hypothesize that external force acting on the DNA template strand would change the primary reactive distance for phosphoryl transfer and thereby affect the rate for catalysis. Including additional polymerase residues should not alter our results as the force is applied to the DNA template strand and should not affect the polymerase directly. Further we expect that the motions of the light hydrogen atoms are well separated and uncorrelated from any global DNA motions and motions at the catalytic site and use only heavy (non-hydrogen) atoms in the active site fragment while performing our PCA. This simply means that the externally applied force does not alter the reactive distance of catalysis through coupling with hydrogen atoms and does not imply that hydrogen atoms are unimportant for catalysis. Indeed hydrogen atoms play a very important role in stabilizing the catalytic site and we fully account for this in our all atom molecular dynamics simulations. The role of the hydrogens is thus implicitly included in the spring constants for the primary reactive distance X calculated from molecular dynamics data. In order to show that increasing the size of the subset and/or the inclusion of hydrogens does not affect our results we have performed calculations with two other subsystems (SS): SS II: Heavy (non-hydrogen) atoms in an extended active site region which also includes the polymerases O and O1-helices (part of the polymerase fingers domain) and several polymerase residues implicated in mismatch sensing and extension which contact the DNA fragment in the duplex region below the nascent base pair ; SS III: All atoms (including hydrogens) in the active site region. Figure S7 (Top) shows the different subsets (**G:C** case) for which calculations were performed and Figure S7 (Bottom) shows (**G:C** case) that the relative phosphoryl transfer rate vs applied force curves calculated for the three different subsets are identical.

One of the limitations in our simulations is that we are able to include only a small DNA fragment, while the DNA sequence in force spectroscopy experiments is many hundreds of base pairs long. However since the coupling of motions in the DNA template strand with the reaction coordinates for catalysis is mediated by the polymerase whose interactions with DNA are limited to the active site

region, we expect that the present simulation provides a reasonable description for the force dependence of phosphoryl transfer given and our results should be directly testable in systems where phosphoryl transfer becomes rate limiting (i.e., for systems with non-cognate DNA/dNTP substrates).

S.3 Displacement of PC by an externally applied force on the template strand

In the zero[th] order approximation an external force applied on the template strand will displace atoms in the active site fragment. Let $\Delta\mathbf{x} = (\Delta x_1, \Delta y_1, \Delta z_1, \Delta x_2, \Delta y_2, \Delta z_2, \ldots, \Delta x_{3N}, \Delta y_{3N}, \Delta z_{3N})$ be the 3N dimensional displacement vector which represents the displacement of the N atoms in the active site fragment due to the applied force $\mathbf{F}$. We can express this displacement vector in term of the 3N normalized PC modes of the active site fragment ξ_m which form a complete basis as $\Delta\mathbf{x} = \Sigma_m a_m \, \xi_m$, with expansion coefficients a_m. The Hamiltonian for the system is given by:

$$H = \frac{1}{2}\sum_{m=1}^{3N} k_m (a_m \boldsymbol{\xi}_m)^2 - \sum_{m=1}^{3N} \mathbf{F} \bullet (a_m \boldsymbol{\xi}_m) \tag{1}$$

At equilibrium we have for $\partial H/\partial a_m = 0$ for each a_m which gives:

$$a_m = \frac{\mathbf{F} \bullet \boldsymbol{\xi}_m}{k_m} = \frac{|\mathbf{F}|\cos\theta_m}{k_m} \tag{2}$$

Here $\mathbf{F}$ is a 3N dimensional vector representing the force on the active site fragment and θ_m is a generalized angle between $\mathbf{F}$ and $\boldsymbol{\xi}_m$ (PCs are normalized):

$$\cos\theta_m = \frac{\mathbf{F} \bullet \boldsymbol{\xi}_m}{|\mathbf{F}|} = \frac{\displaystyle\sum_{\substack{i=1..N \\ j=x,y,z}} F_i^j \xi_{mi}^j}{\sqrt{\displaystyle\sum_{\substack{i=1..N \\ j=x,y,z}} (F_i^j)^2}} \tag{3}$$

Where $F_i^{\ j}$ denotes the component of applied force acting on the ith atom of the active site fragment in the direction j. In the experiments the force is applied to the ends of the template strand and along the helix axis. Thus we define our applied force to be:

$$F_i^{\ j} = F_0(i)n_j \qquad\qquad \forall\ i\ \in\ \{x_T\}$$

$$F_i^{\ j} = 0 \qquad\qquad \forall\ i\ \notin\ \{x_T\} \qquad\qquad (4)$$

Here $F_0(i)$ is the magnitude of force acting on the ith atom ($|\mathbf{F}| = \sqrt{\sum_{i=1}^{N_T}(F_0(i))^2}$) and $\{x_T\}$ is a subset of

N_T atoms from the active site fragment belonging to the DNA template strand. The components n_j (j=x,y,z) belong to a unit vector along the DNA helix axis. In our simulations we assume that $\{x_T\}$ includes only the backbone phosphate (P_α) atoms in the template strand ($N_T=6$) and that the template strand fragment is short enough that the same force $F_0(i) = |\mathbf{F}|/\sqrt{N_T}$ acts on all the phosphate atoms. In our simulations we vary $|F|$ to get the force dependence of the phosphoryl transfer step.

References

1. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. Bmc Bioinformatics 2006;7:-.
2. Johnson SJ, Taylor JS, Beese LS. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. Proc Natl Acad Sci USA 2003;100(7):3895-3900.
3. Johnson SJ, Beese LS. Structures of Mismatch Replication Errors Observed in a DNA Polymerase. Cell 2004;116(6):803-816.
4. Kiefer JR, Mao C, Braman JC, Beese LS. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. Nature 1998;391(6664):304-307.
5. Eastwood MP, Hardin C, Luthey-Schulten Z, Wolynes PG. Evaluating protein structure-prediction schemes using energy landscape theory. Ibm Journal of Research and Development 2001;45(3-4):475-497.
6. Doublie S, Tabor S, Long AM, Richardson CC, Ellenberger T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 angstrom resolution. Nature 1998;391(6664):251-258.
7. Patel PH, Suzuki M, Adman E, Shinkai A, Loeb LA. Prokaryotic DNA polymerase I: Evolution, structure, and "base flipping" mechanism for nucleotide selection. J Mol Biol 2001;308(5):823-837.
8. Insight II molecular modelling software. San Diego: Molecular Simulations Inc.; 2000.

9.   Brünger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. Proteins Struc Func Gen 1988;4:148-156.

10.   Radhakrishnan R, Schlick T. Fidelity discrimination in DNA polymerase beta: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair. . J Am Chem Soc 2005;127:13245-13253.

11.   Grubmuller H, Heymann B, Tavan P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. Science 1996;271:997-999.

12.   Hsu GW, Ober M, Carell T, Beese LS. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. Nature 2004;431:217-221.

13.   Brieba LG, Eichman BF, Kokoska RJ, Doublie S, Kunkel TA, Ellenberger T. Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. Embo J 2004;23(17):3452-3461.

14.   Krahn JM, Beard WA, Miller H, Grollman AP, Wilson SH. Structure of DNA polymerase beta with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. Structure 2003;11(1):121-127.

15.   Joyce CM, Benkovic SJ. DNA Polymerase Fidelity: Kinetics, Structure, and Checkpoints. Biochemistry-Us 2004;43(45):14317-14324.

16.   Jr ADM, Bashford D, Bellott M, Dunbrack RL, Jr, Evanseck MJFJD, Fischer S, Gao J, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, B. Prodhom andW. E. Reiher I, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B 1998;102:3586--3616.

17.   Mackerell AD, Banavali NK. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. J Comput Chem 2000;21(2):105-120.

18.   Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. J Comput Phys 1999;151(1):283-312.

19.   Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26(16):1781-1802.

20.   Cheatham TE, Miller JL, Fox T, Darden TA, Kollman PA. Molecular Dynamics Simulations of Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. J Amer Chem Soc 1995;117:4193-4194.

21.   Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. Journal of Molecular Graphics 1996;14:33--38.

22.   Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins Struct Funct Gen 1993;17(4):412-425.

23.   Amadei A, Linssen ABM, deGroot BL, Aalten DMFv, Berendsen HJC. An Efficient Method for Sampling the Essential Subspace of Proteins. J Biomol Struct Dyn 1996;13(4):615-625.

24.   Glykos NM, Kokkinidis M. Structural polymorphism of a marginally stable 4-alpha-helical bundle. Images of a trapped molten globule? Proteins 2004;56(3):420-425.

25.   Venkatramani R, Radhakrishnan R. The effect of oxidatively damaged DNA on the active site preorganization during nucleotide incorporation in DNA by a high fidelity polymerase from Bacillus stearothermophilus Proteins Struct Funct Bioinf 2007;In press.

Sequence Homology of the polymerase active site used in this study

BF	E696VTPNMRRQAKAVNFGIVYGI SDYGLAQNLN726	D653	H829 D830	Q797	R615
KF	T748VTSEQ RRS AKAI NFGLIYGMSAFGLARQLN778	D705	H881 D882	Q849	R668
T7	- L515P - - TRDNAKTF IYGFLYGAGDEKI GQIVG542	D475	H653 D654	Q615	R429
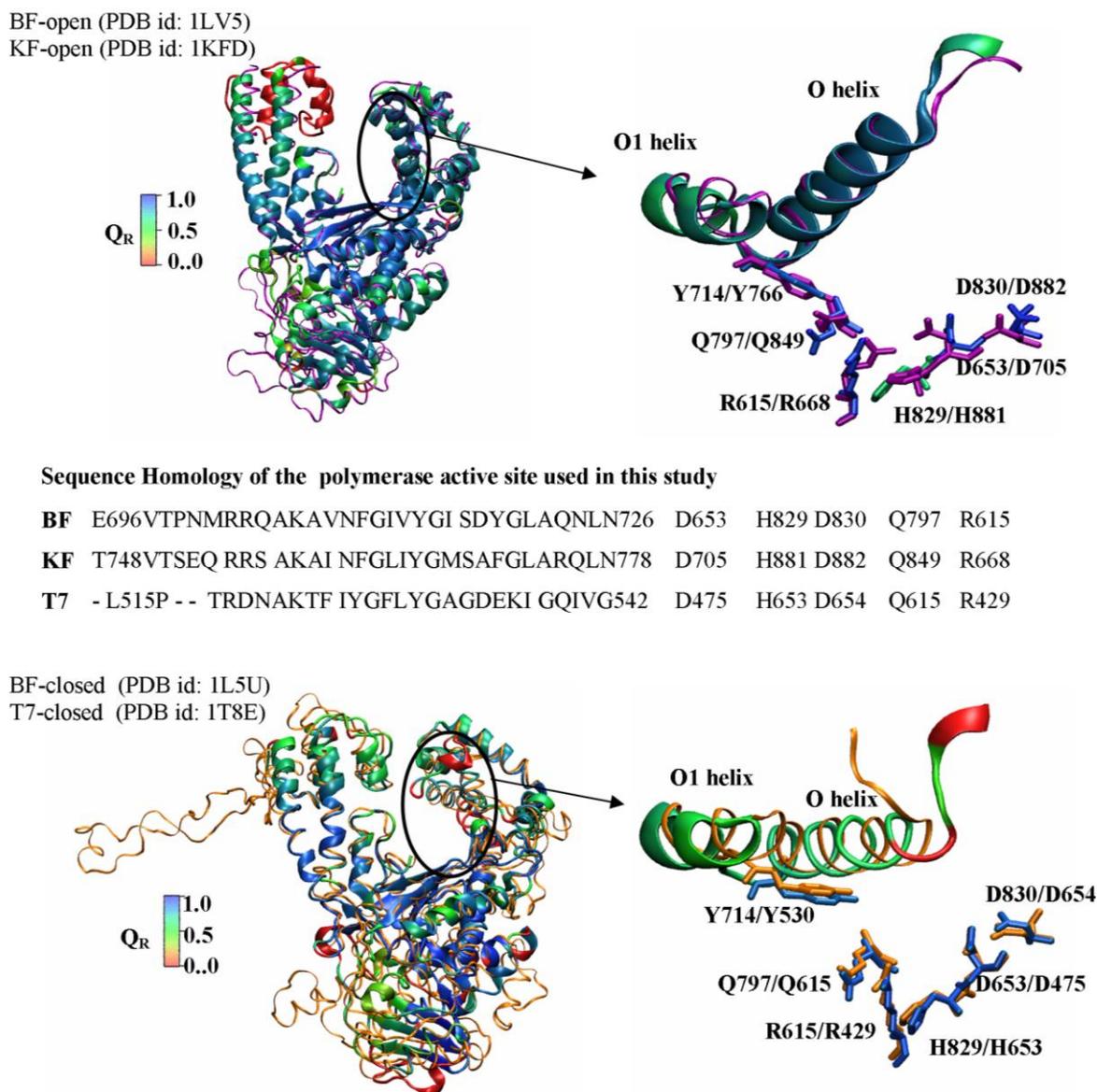
Figure S1: Sequence (center table) and Structural Homology of BF pol I with *E. coli* Klenow and T7 DNA pol. The figures on the left show a superimposition of BF pol structures (color coded according to the measure Q_{res} defined in the text) with KF (top, purple ribbons) and T7 DNA pol (bottom, orange ribbons). Since closed ternary structures for KF is not available we superimpose open binary complexes of BF and KF. The figures on the left show a similar superimposition for the active site fragment of the polymerase used in our simulations. The RMSD for backbone atoms in the active site region for BF and KF is 0.96 Å and for BF and T7 DNA pol is 1.87 Å. Also shown is the sequence homology of the active site fragment. Residue numbers are also given for each structure. We use the MultiSeq[1] plugin in VMD for the alignment of the protein structures.
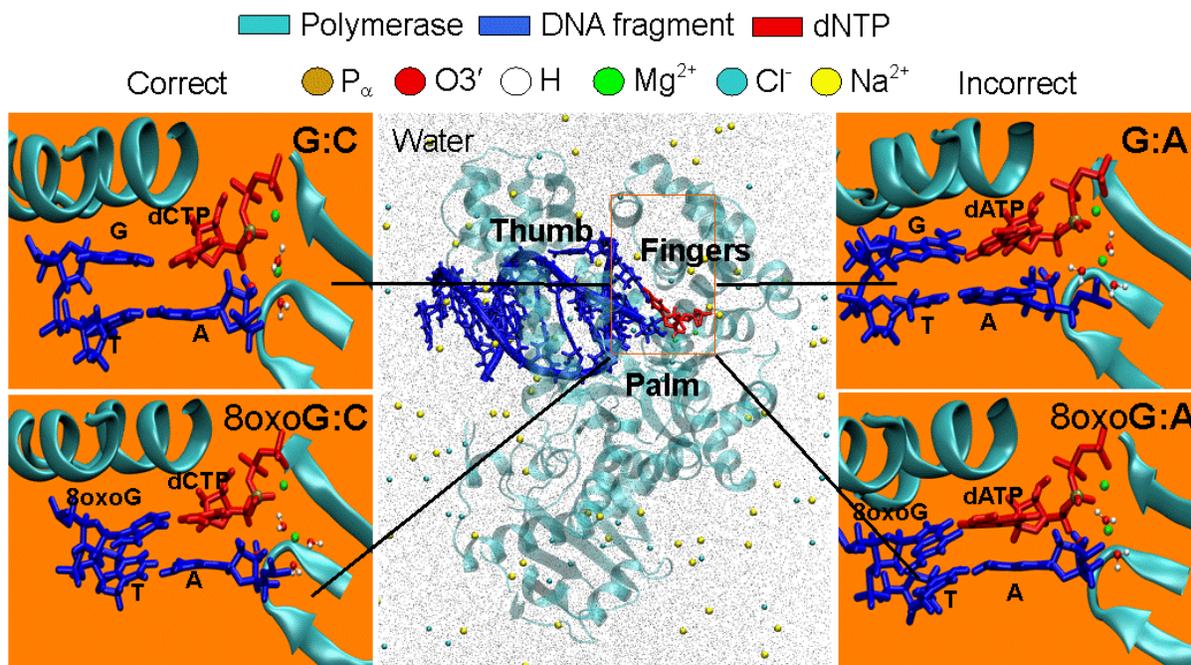
Figure S2: Simulations are carried out on fully solvated and neutralized ternary complexes (center) for the Bacillus fragment (BF). The four insets show the average active site geometry (DNA, incoming dNTP, Mg^{2+}, bound waters, parts of the polymerase fingers and palm domains) from 5 ns classical simulations for the four model systems for correct/incorrect nucleotide incorporation opposite an undamaged/oxidatively damaged **G** template base as indicated.
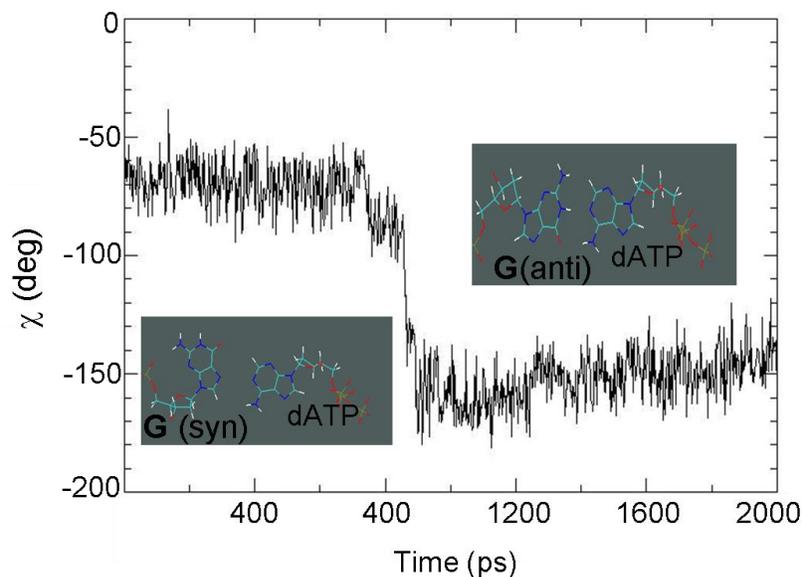
Figure S3: Values of the Glycosidic angle χ during a 2ns production run for the **G**(syn)**:A** case . The templating base **G** starts out in a syn conformation as depicted on the LHS (top left), flipping over to the anti conformation (bottom left).
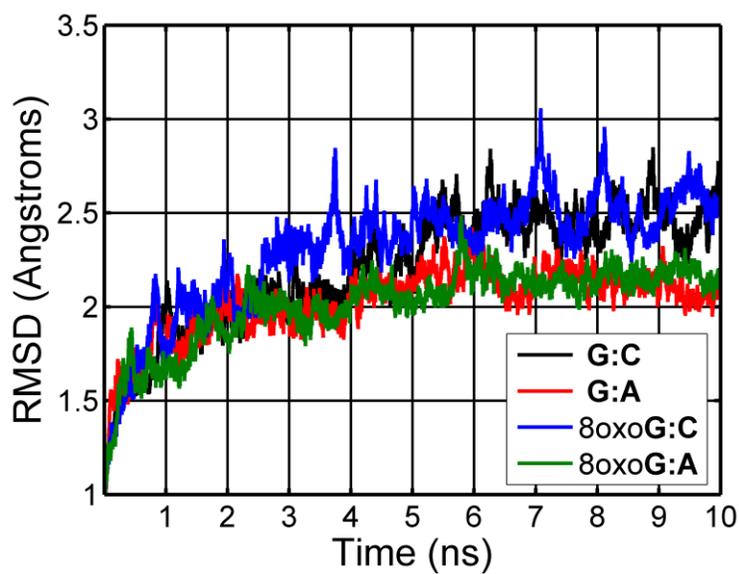
Figure S4: All atom RMSD along a 10 ns, unconstrained, classical NVT production run. The plateau in the last 5 ns is interpreted as the equilibrium phase; data from the last 5 ns was used in further analysis.

G:C (Control)		
Structural parameters **Distances Atom1—Atom2**	**MM** **Å**	**Crystal structure** **Å**
G:C1$'$—dCTP:C1$'$ [(IS)]	10.57 (0.15)	[(a)]10.56 [(b)]10.3 (0.2)
T:C1$'$—**A**:C1$'$ [(PIS)]	11.11 (0.23)	[(a)]10.10 [(b)]10.3 (0.2)
dCTP:P$_\alpha$— **A**:O3$'$	3.51 (0.26)	-
MG2—A:O3$'$	2.66 (0.56)	[(c)] ~2.0[*]
MG2—dCTP:O2$_\alpha$	2.43 (0.35)	[(a)]2.66 [(c)] 2.4
MG2—D831:O1$_\delta$	1.81 (0.04)	[(a)]2.66 [(c)] 2.4
MG2—D653:O2$_\delta$	1.80 (0.04)	[(a)]2.93 [(c)] 2.4
MG2-MG1	3.62 (0.18)	[(a)]3.54[**] [(c)] 3.6
MG2-WATER1:O	1.94 (0.06)	[(c)] 2.6
MG2-WATER2:O	1.93 (0.05)	[(c)] 2.5
D830:O1$_\delta$—D653:O2$_\delta$	2.70 (0.10)	[(a)]3.82
A:H3T—D830:O1$_\delta$	3.02 (0.24)	-
A:H3T—dCTP:O1$_\alpha$	3.20 (0.33)	-

Mispair and Oxidative damage			
Structural parameters **Distances Atom1—Atom2**	**G:A MM** **Å**	**8oxoG:C MM** **Å**	**8oxoG:A MM** **Å**
G:C1$'$—dCTP:C1$'$ [(IS)]	12.04 (0.25)	10.63 (0.16)	10.68 (0.23)
T:C1$'$—**A**:C1$'$ [(PIS)]	10.66 (0.21)	10.80 (0.23)	10.87 (0.24)
dCTP:P$_\alpha$— A:O3$'$	4.74 (0.27)	4.50 (0.42)	3.36 (0.18)
MG2—A:O3$'$	4.72 (0.22)	3.77 (0.23)	2.47 (0.39)
MG2—dNTP:O2$_\alpha$/O1$_\alpha$	2.83 (0.19)	3.63 (0.15)	2.38 (0.35)
MG2—D831:O1$_\delta$	1.83 (0.04)	1.83 (0.04)	1.81 (0.04)
MG2—D653:O2$_\delta$	1.82 (0.04)	1.79 (0.04)	1.81 (0.04)
MG2-MG1	3.80 (0.11)	4.13 (0.11)	3.65 (0.12)
MG2-WATER1:O	2.00 (0.07)	1.94 (0.05)	1.95 (0.06)
MG2-WATER2:O	1.96 (0.06)	1.91 (0.05)	1.94 (0.06)
MG2-WATER3:O	1.95 (0.06)	2.02 (0.08)	-
D830:O1$_\delta$—D653:O2$_\delta$	2.62 (0.08)	2.60 (0.07)	2.70 (0.10)
A:H3T—D830:O1$_\delta$	4.42 (0.24)	2.76 (0.49)	2.94 (0.18)
A:H3T—dCTP:O1$_\alpha$/ O2$_\alpha$	4.77 (0.34)	4.14 (0.51)	3.07 (0.28)

[*]   Modeled
[**]   Mn replaces MG1 in the crystal structure
IS   insertion site (From ref [2]: the site occupied by the incoming nucleotide and its pairing template base n)
PIS   post insertion site (from ref [2] : the n -1st base pair)
(a)   Crystal structure of BF ternary complex (PDB id: 1LV5)
(b)   From Johnson and Beese [3]
(c)   From Doublie et. al. [6]

Table S5: Top: Comparison of structural data for the **G:C** case obtained from classical MM trajectories (last 5 ns) with crystal structure values for BF and T7 DNA pol. Bottom: MM averages of structural data for **G:A**, **8oxoG:C**, and 8oxo**G:A** simulations.

Movie S6: Movie showing the dominant principal component mode (mode 1) for the four simulated model systems (top left – G:C ; top right – G:A ; bottom left – 8oxoG:C; bottom right – 8oxoG:A) at the active site. The region comprises of 7 residues of the DNA template strand (red), four residues of the primer strand (yellow), the incoming nucleotide (orange - dCTP; purple – dATP), and conserved components of the catalytic site (acidic aspartate residues D830 & D653 (iceblue), Mg^{2+} ions and bound water molecules). The catalytic $O3'\text{-}P_\alpha$ distance is also shown (dashed line). The dominant mode in the control G:C system exhibits significant catalytic site motions with large fluctuations in the $O3'\text{-}P_\alpha$ distance which are highly correlated to motions in the DNA template and primer strands. In contrast the remaining three systems (G:A, 8oxoG:C & 8oxoG:A) show a very small perturbation of the catalytic site with large scale motions being localized to the region near the single stranded template overhang. The amplitude of $O3'\text{-}P_\alpha$ distance fluctuations associated with the dominant mode follows G:C >> 8oxoG:A > 8oxoG:C > G:A. A visualization of the top 10 modes for the four model systems showed similar trends.
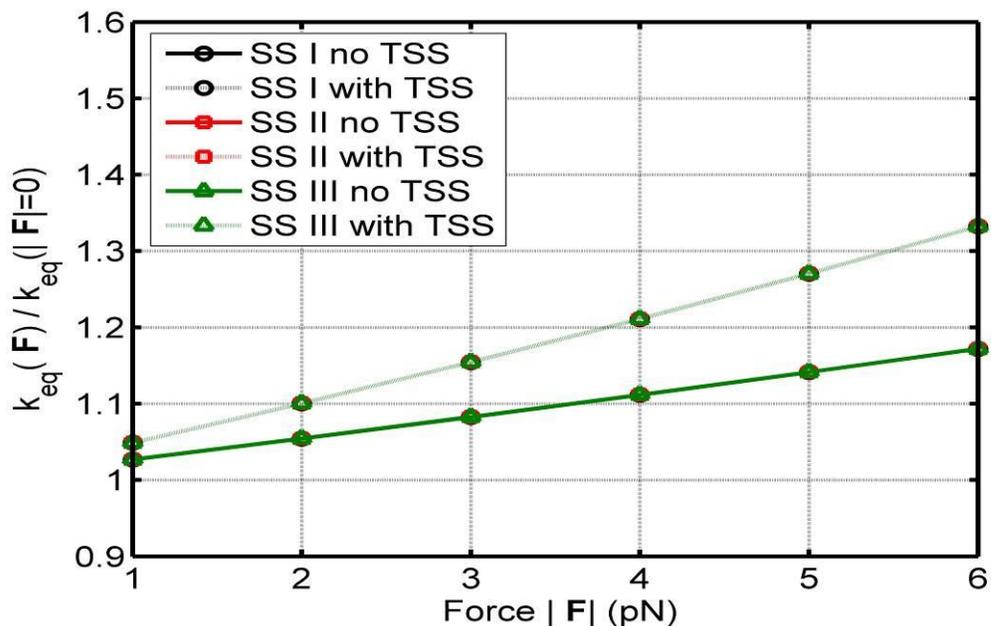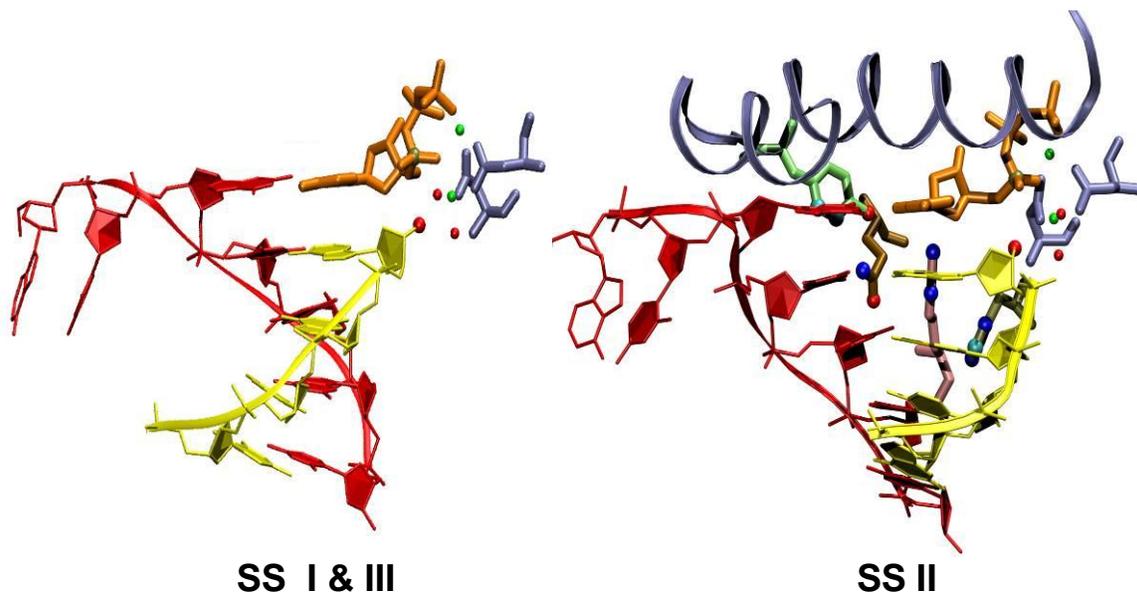
Figure S7: Top: Subsets of atoms used (G:C case) for Principal component analysis. SS I : Active site region (heavy atoms only); SS II : Extended active site (heavy atoms only) ; SS III: Active site region (all atoms). Bottom: Relative Phosphoryl transfer rate vs applied force for the three subsets. The results are identical.
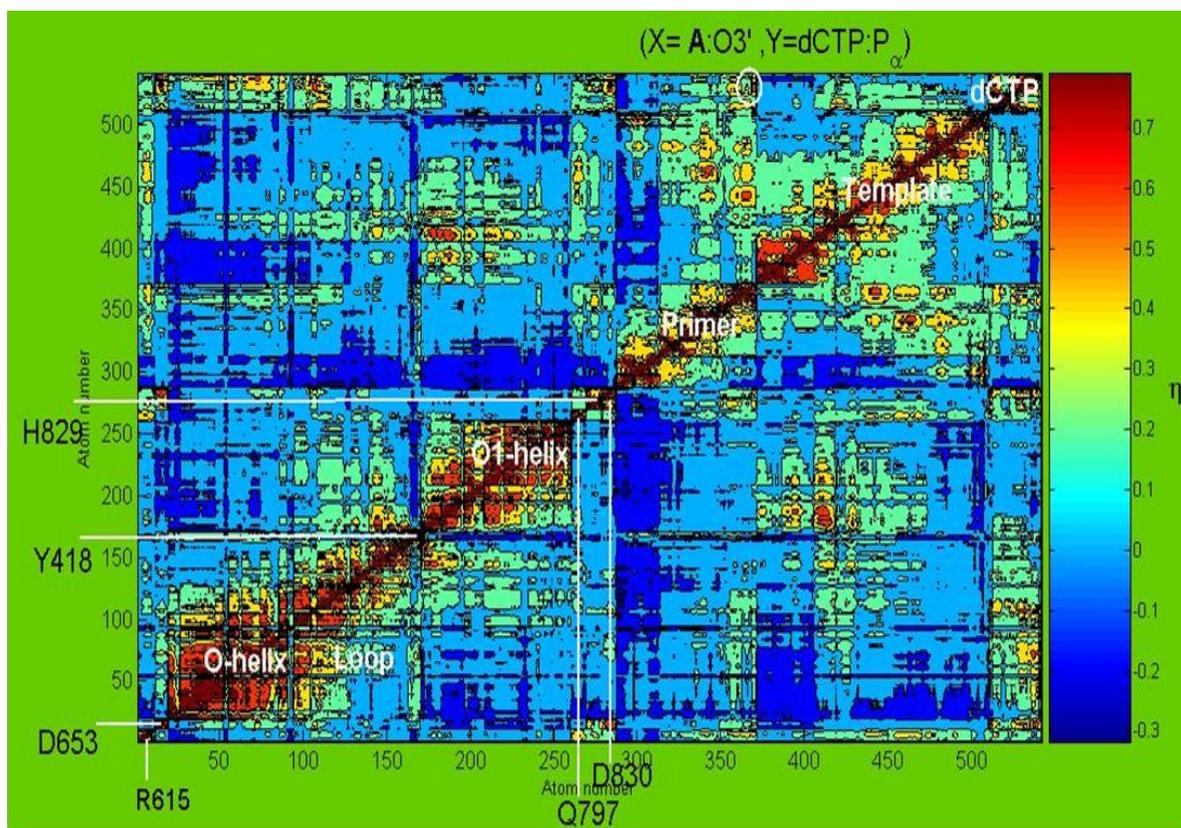
Figure S8: Correlations between vector displacements (**r - <r>**) of atoms in the active site region (Figure S6) for the **G:C** system. Here **r** is a vector drawn from the origin to the atom of interest with average value **<r>.**
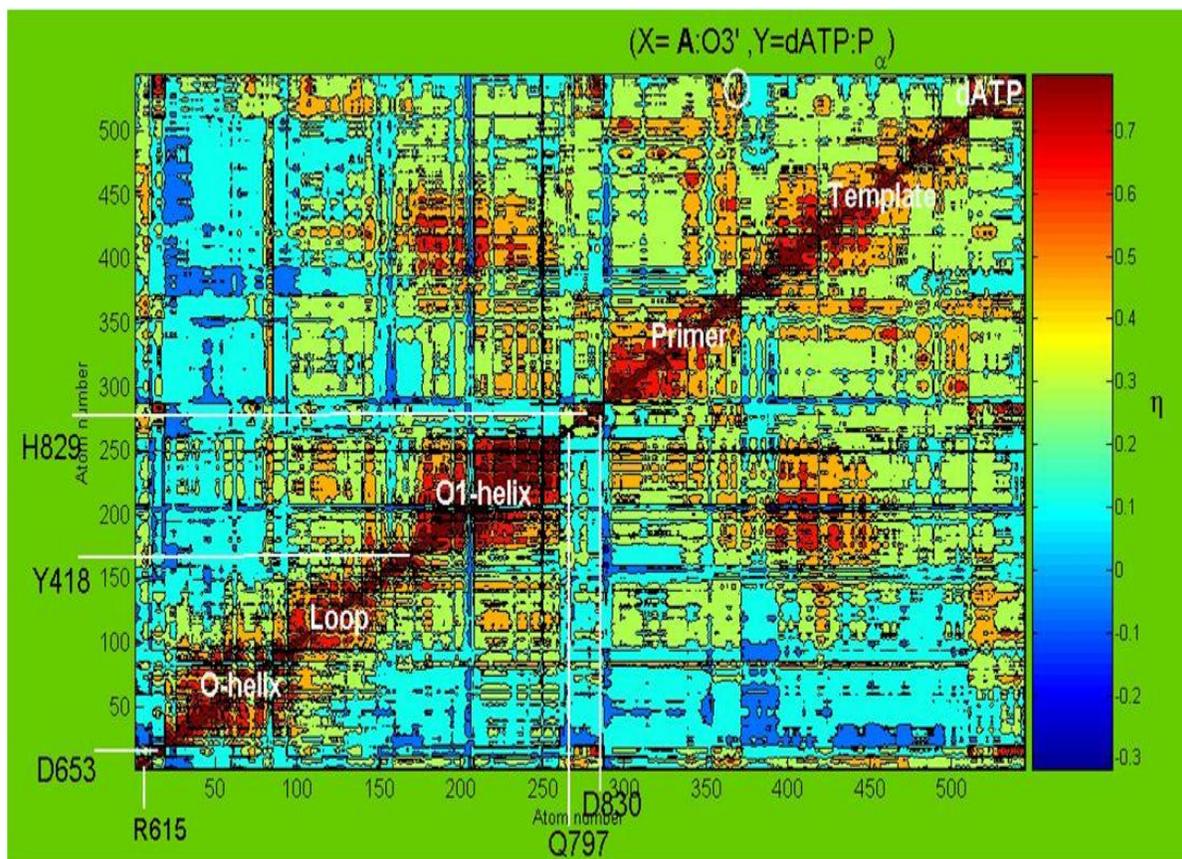
Figure S9: Correlations between vector displacements (**r - <r>**) of atoms in the active site region (Figure S5) for the **G:A** system.
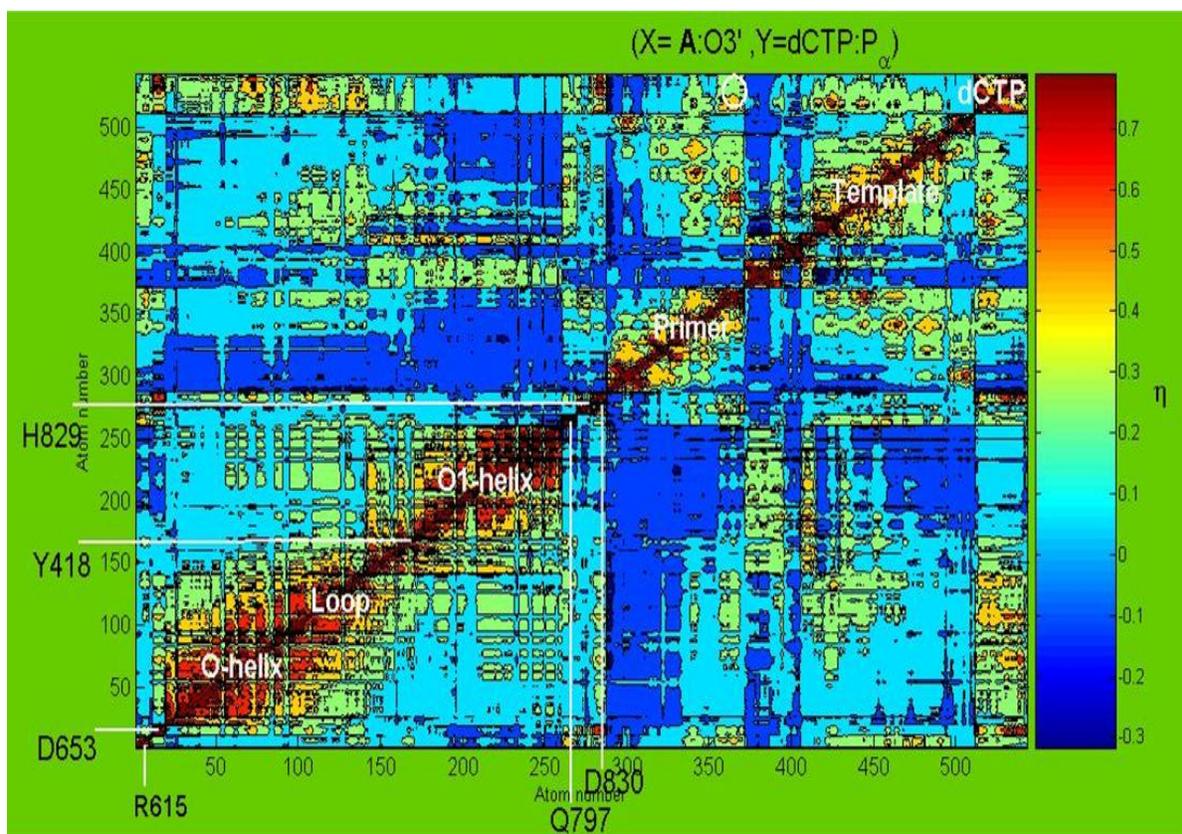
Figure S10: Correlations between vector displacements (**r - <r>**) of atoms in the active site region (Figure S5) for the 8oxo**G:C** system.
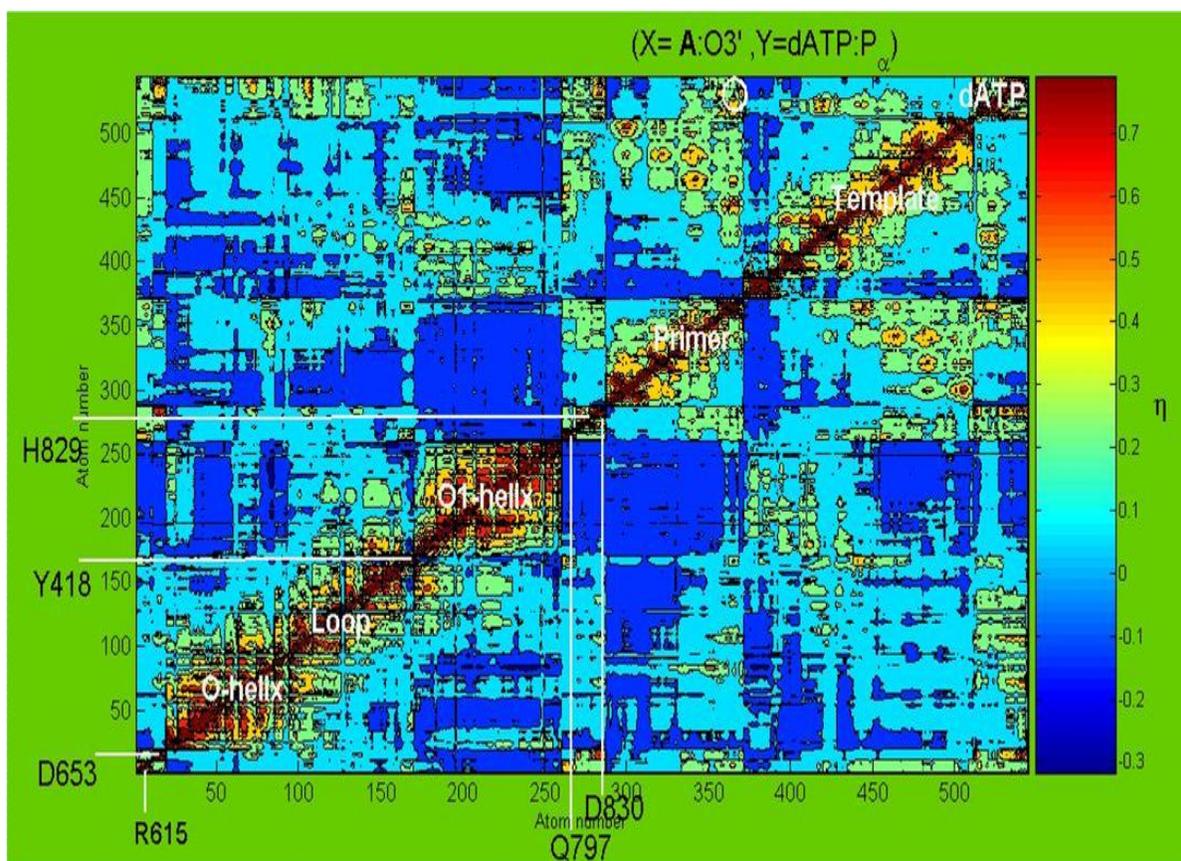
Figure S11: Correlations between vector displacements (**r - <r>**) of atoms in the active site region (Figure S5) for the 8oxo**G:A** system.