

Regular article

Structural and dynamic variations in DNA hexamers containing T-T and F-F single and tandem internal mispairs

Edward C. Sherer, Christopher J. Cramer

Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant St. SE, Minneapolis, MN 55455-0431, USA

Received: 23 June 2003 / Accepted: 14 July 2003 / Published online: 16 February 2004
© Springer-Verlag 2004

Abstract. Molecular dynamics simulations of double-helical DNA oligomers have been performed to investigate differences in the structure, dynamics, and hydration of F-F and T-T mispairs. Hexamers containing F-F pairs were found to be more dynamic, especially in the region of the mispair itself. This dynamic variability derives from greater flexibility of F-F pairs. The T-T mispairs, on the other hand, were found to be comparatively tightly bound as wobble pairs. The major and minor groove edges of the T-T pairs were observed to be solvated at exposed carbonyl positions by at least one water molecule, while F-F pairs lacked solvating waters. Stacking interactions were nearly identical for T-T and F-F pairs, leading to similar average structures, even though F stacking was more dynamically variable. Solvation differences between F-F and T-T therefore support the steric exclusion model for nucleotide incorporation in DNA replication. Large differences in the orientation of minor groove functional groups, in addition to differences in solvation, further rationalize why F bases present during DNA extension events induce stalls. Two novel nucleotides are proposed to further elucidate minor groove interactions of DNA with polymerase molecules.

Electronic Supplementary Material This Material consists of equilibration protocol, plots of center-of-mass stacking, water radial distribution functions, helical parameter dynamics, and dynamics data for a control AT sequence.

Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00214-003-0534-3>.

Keywords: DNA replication – Nucleotide mispair – Unnatural nucleic acid base – Principal component analysis

Introduction

Replication of the genetic information contained in DNA is performed to an exquisitely high level of accuracy. Several recent X-ray crystal structures of polymerases with bound substrates have provided the basis for an atomic level interpretation of the mechanism of DNA replication/translation [1–6]. In particular, these studies have advanced our understanding of how specific nucleotide selection is accomplished. A tight binding pocket surrounding the forming base pair selects correct nucleotides in part on the basis of a Watson-Crick [7] geometry, correctly formed hydrogen bonds, and base stacking interactions. The result is correctly matched G-C and A-T base pairs. Questions in regard to the relative importance of the various factors affecting selection, including whether hydrogen bonding interactions are needed at all, have been the basis of significant scientific debate [8–14]. Moreover, the relative importance of the different factors influencing selection may vary for different replication functions, which include insertion, extension, and editing.

Experiments by Kool and co-workers have been among the most compelling in addressing the effect of correctly formed hydrogen bonds on the fidelity of DNA polymerase activity [15–24]. Taken together, these studies have shown that faithful replication of DNA base pairs can be accomplished even in the absence of hydrogen bonds, as shown using the non-polar shape mimics F (2,4-difluorotoluene) and Z (4-methyl-benzimidazole), designed to mimic T and A, respectively. Not only was it found that the Klenow fragment of DNA polymerase I was able to incorporate F across from A, it

Contribution to the Jacopo Tomasi Honorary Issue

Correspondence to: C. J. Cramer
e-mail: cramer@chem.umn.edu

was also able to incorporate A when F existed in the template strand. These results support the idea that shape complementarity is more important to fidelity than hydrogen bonding is.

These experimental results prompted several theoretical studies demonstrating significantly reduced hydrogen bonding interactions between A-F pairs in comparison to A-T pairs [15, 25–28]. When calculating the difference in gas-phase interaction enthalpies for the two base pairs, A-T is calculated to form with almost 8 kcal mol⁻¹ advantage. However, due to the decreased penalty of desolvating F relative to T in water, this pairing advantage in the gas phase is largely lost in aqueous solution [25]. The unique nature of the hydrophobic effect is manifest insofar as A and F fail to pair in chloroform in contrast to A-T [16, 29].

Since difluorotoluene mimics the shape of thymine so well, much of the biological behavior of F can be closely correlated with that of T. However, F does have several unique characteristics. First, F is incorporated as a pair with another F to a significant extent in DNA replication – about 900-fold more frequently than is the case for T-T pairs [16]. This higher efficiency of F-F formation is nearly identical to the efficiency of A-F formation. The efficiency of F-F formation compared to T-T has been argued to derive in part from different solvation effects: a fully solvated T base does not regain all of its favorable hydrogen bonding interactions when forming a T-T pair, while an F-F pair has no such interactions to lose [13]. In a tight hydrophobic pocket, a T-T pair carrying its associated water molecules would not fit, but an unsolvated F-F pair might.

Difluorotoluene also induces a stall in DNA polymerase replication if two A-F pairs occur consecutively in the template strand [17]. The presence of a single A-F pair, however, does not induce a stall. Finally, when A-F pairs pass through the editing mechanism of the Klenow fragment of DNA polymerase I, the mismatch is edited more rapidly than its sterically identical partner A-T [30]. Since A-F pairs have a greater rate of base pair opening, it has been hypothesized that such a local distortion of base pairing geometry may trigger the editing event [21, 30–32]. Therefore, some relatively subtle differences between T and F appear to have significant effects on the different replication processes.

It has been hypothesized that functional groups in the minor groove of DNA are integral to the extension step of replication [10, 33–36]. Specific amino acid contacts to the bound DNA, in a region of several DNA polymerase co-crystal structures that is termed the “minor-groove region”, (MGR) have led to the suggestion that minor-groove functional groups play a role in DNA translocation [3, 4, 37]. Although significant differences exist between A-T and G-C pairs, the MGR can form hydrogen bonds to either pair in a sequence independent fashion, due to the quasi-symmetrical distribution of hydrogen bond acceptors in the minor-groove-exposed functionality of these base pairs [2, 11].

Specifically, these acceptors are O2 of T paired with N3 of A, and N3 of G paired with O2 of C. Guanine contains an additional N2 hydrogen bond donating amino group. A growing body of work has addressed functional group substitution in both nucleic acid bases and amino acids, with the goal of better characterizing these minor-groove interactions [18, 38–45].

While F represents a specific example of functional group alteration, another example would be the synthesis of analogs of T and C which lack a 2-keto group in the minor groove, and inhibit DNA synthesis altogether [42]. Such mismatches can affect extension rates up to four base pairs upstream from the forming pair [6]. The loss of a single minor groove interaction may influence efficiency by a factor of 300 [18]. This is most clearly demonstrated by the difference in extension efficiencies of the two base pairs Q-F and Z-F, where Q and Z differ in that Q contains a N3 atom in the minor groove [44]. The Q-F pair interacts more favorably with the MGR of polymerases. Unfortunately, no single conclusion can be drawn concerning how any one polymerase will interact with the minor groove of DNA, as several general groupings of polymerase interactions can be made [6, 18, 45]. In part, these groupings depend on the number of hydrogen bonds to the minor groove, and whether certain non-standard pairs are replicated efficiently.

Structural information for certain mispairs, such as A-F and Z-F, exist both from NMR experiments and molecular dynamics (MD) simulations. While the static information for the above pairs indicates that they are only distorted in a relatively minor way from canonical B-form DNA, their dynamic variations were found to be more pronounced. To date, MD has proven very effective for modeling helices containing F nucleotides. In particular, two separate studies have validated parameters developed for F (to be used in conjunction with the force-field of Cornell et al. [46]). In good agreement with experiment [16, 21], MD simulations employing these parameters showed higher base-pair opening rates for A:F [31], and a destabilization of about 5 kcal mol⁻¹ for a DNA duplex after T → F mutation [47].

Modern MD simulations are now widely accepted as not only accurate, but predictive [48–56]. In addition to canonical structures, MD simulations performed for helices containing mispaired or non-standard bases have also met with significant success, as measured by favorable comparison to accurate experimental data [31, 46, 57–64]. Here we carry out simulations to monitor the differences in structure, dynamics, and hydration of T-T versus F-F mispairs in DNA hexamers, with the goal of achieving a better understanding of why F-F pairs are replicated more efficiently. In addition to these single pyrimidine mismatches, two helices containing tandem T-T or F-F pairs are also analyzed. Kool et al. [65] have measured the melting temperature changes induced by T-T and F-F pairs. In a dodecamer helix they found that, relative to T-A, T-T and F-F pairs are 3.4 kcal mol⁻¹ less stable. We use for our simulations four base pairs, including the mispair from the sequence used for

these T_m calculations, and we cap each end with a G-C pair to prevent fraying. As base stacking considerations form a large part of our analysis, we employ here the force-field of Cornell et al. [46], which has been demonstrated to reproduce stacking energetics predicted from high level ab initio calculations [66] with a high degree of accuracy.

Methods

Hexamer sections of DNA were built in canonical B-form having the sequences 5'-CCTTTC/GAATGG, 5'-CCTTTC/GATTGG, 5'-CCFTTC/GAAFGG, and 5'-CCFFTC/GAFFGG (mispairs in bold). These sequences are named **mis-T**, **tandem-T**, **mis-F**, and **tandem-F**, respectively, and the bases are numerically labeled with the 5'-cytosine as C1 and the first thymine/difluorotoluene as residue 3, and so on. An additional control simulation was run with a canonical hexamer having the sequence 5'-CCATTC/GAATGG, which is referred to as **AT**. MD trajectories using the force-field of Cornell et al. [46] were propagated as implemented in the AMBER5 software package [67] for each of the different helices. Force-field parameters for difluorotoluene have been described previously [31, 47]. Each helix contains six residues (379 atoms), and ten sodium ions were added to each in order to neutralize the charge of the polyelectrolyte. Periodic boundary conditions were established for simulation boxes containing approximately 1770 TIP3P water molecules. The structures were equilibrated following a previously described protocol [58, 59, 62, 68], which is listed in detail in the Supporting Information. Following equilibration, each trajectory was propagated until the all-atom root-mean-square deviation (RMSD) was assessed to have converged over 3 ns. Equilibration was further monitored by comparing the average and standard deviation of selected helical parameters over consecutive trajectory blocks (Supporting Information).

During propagation, electrostatic interactions were evaluated without cutoffs using the Particle-Mesh-Ewald (PME) method (1 Å grid with a cubic spline approximation and a direct sum tolerance of 0.000005) [69, 70]. The cut-off for the non-bonded interactions was 9.0 Å, and the 1/4 electrostatics were scaled by a factor of 1.2. Updating of the non-bonded pairlist was performed every 25 fs. Constant pressure (1 bar) and temperature (300 K) were maintained according to the Berendsen [71] coupling algorithm (separate scaling factors for the solvent and solute used; both with a coupling time of 0.2 ps). SHAKE was used to constrain all covalent bond distances (with a tolerance of 0.0005 Å) [72], and a propagation time-step of 2 fs was used. The center-of-mass translational motion was removed from the trajectory every 0.2 ps. Rotations of DNA oligomers within simulation cells did not lead to artifactual interactions with periodic images over the time courses of any of the simulations.

Time-averaged structures over the final 3 ns of each 4 ns trajectory (and over a 6 ns trajectory created from catenation of either the two single-mispair, or two double-mispair trajectories) were calculated and minimized while restraining all heavy atom positions. Helical geometrical parameters [73, 74] for different structures were calculated using Curves [75].

Principal component analysis (PCA) of the helical dynamics was performed as described previously [58, 59, 76]. While PCA has been widely applied in the field of protein dynamics [77–83], until recently [58, 59, 84–86] it has not seen as much application to polynucleotides. PCA, also called essential dynamics, ranks the dominant contributions to a given trajectory as a series of dynamical eigenvectors. The first few such vectors are often observed to account for a very large portion of the total structural variation, such that only one to three degrees of freedom may be needed to describe most of the biologically important dynamics.

Principal component analysis was performed on each individual trajectory, as well as the two catenated trajectories. The positional covariance matrix C was calculated after removal of rotational and

translational motion. Given M snapshots of an N atom system, C is a $3N \times 3N$ matrix with elements

$$C_{ij} = \frac{1}{M} \sum_{k=1}^M (q_{i,k} - \langle q_i \rangle)(q_{j,k} - \langle q_j \rangle) \quad (1)$$

where $q_{i,k}$ is the value for snapshot k of the i th positional coordinate (x , y , or z coordinate for one of the N atoms), and $\langle q_i \rangle$ indicates the average of that coordinate over all snapshots. Diagonalization of C provides the eigenvectors that describe the dynamic motions of the structure, and the associated eigenvalues may be interpreted as weights describing the degree to which each mode contributes to the full dynamics. Qualitative characterization of PCA modes was accomplished by visual inspection of their animation.

The PCA eigenvalue matrix can be employed to estimate the macromolecular configurational entropy. Following the approach of Schlitter [87], Harris et al. have calculated the configurational entropy of certain drug-DNA complexes [88]. The same methodology is used here to estimate the configurational entropy S_{∞} . It should be noted that this estimate is dependent upon the window width of the simulation, but experience has shown that multiple nanosecond simulations approach the converged value (trajectories of less than a few nanoseconds, however, tend to underestimate S_{∞}) [88].

Hydration of the mispairs was investigated in two ways. First, overall helical solvation was calculated by integrating the water density on a 1 Å grid, as described previously [62, 68]. Further analysis of the solvation shells of the mispairs was performed by computing radial distribution functions about their minor- and major-groove functionality using the AMBER analysis package PTRAJ.

Stacking of the internal mispairs either against canonical pairs, or between two consecutive mispairs, was monitored according to the protocol of Nagan et al. [56]. The center of mass (COM) distances between two nucleotide bases were measured over the trajectory, and the analysis performed in the AMBER module CARNAL. Representations of the preferred stacking arrangements were built based on the time-averaged structures.

Results and discussion

General structural information

In the present study, the structure and dynamics of four hexamer duplexes incorporating mispairs were examined. No significant fraying of the terminal base pairs was observed in any of the four MD runs, suggesting that results from larger oligomers would likely be qualitatively similar in the regions of the mispair. Both the all-atom root-mean-square-deviation (RMSD) plots (Fig. 1), and the time-averaged structures (Fig. 2) calculated in reference to these structures for each individual trajectory show evidence of equilibrated, well-behaved simulations. The RMSD plots indicate that larger distortions from the average structure are allowable for the helices containing either one or two F-F base pairs. Differences between either of the single mispairs and their corresponding double mispairs are less pronounced, indicating that the thymine to difluorotoluene perturbation is more structurally destabilizing than adding an additional mispair of the same type. On casual inspection, the average structures do not indicate to any great extent what these structural differences are. Looking down the helix axis, the helices that contain F mispairs are slightly

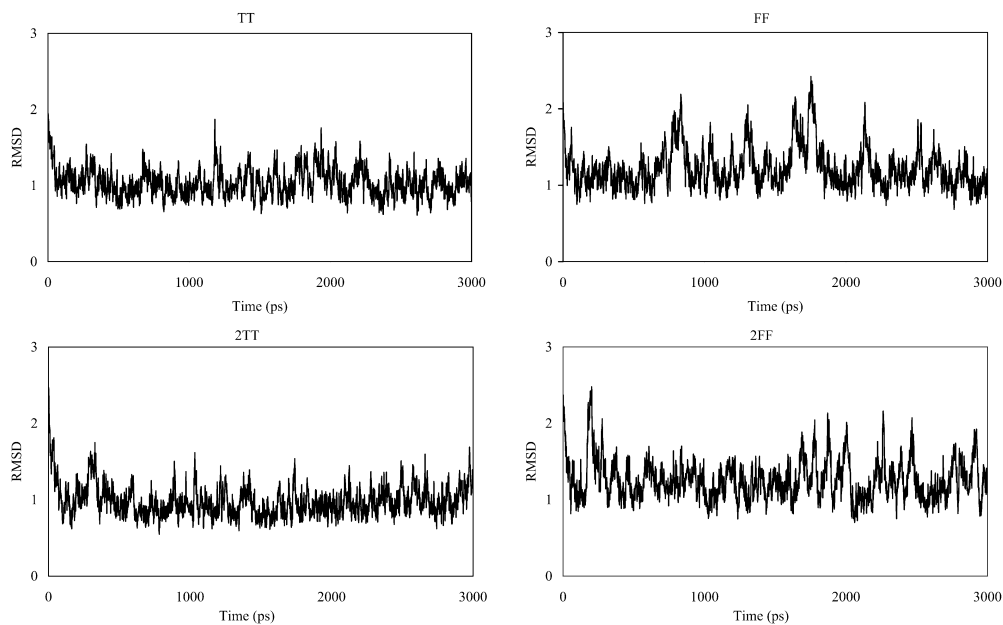


Fig. 1. All-atom RMSD plots (A) calculated over the last 3 ns in reference to the time-averaged structures of **mis-T** (upper left), **tandem-T** (lower left), **mis-F** (upper right), and **tandem-F** (lower right)

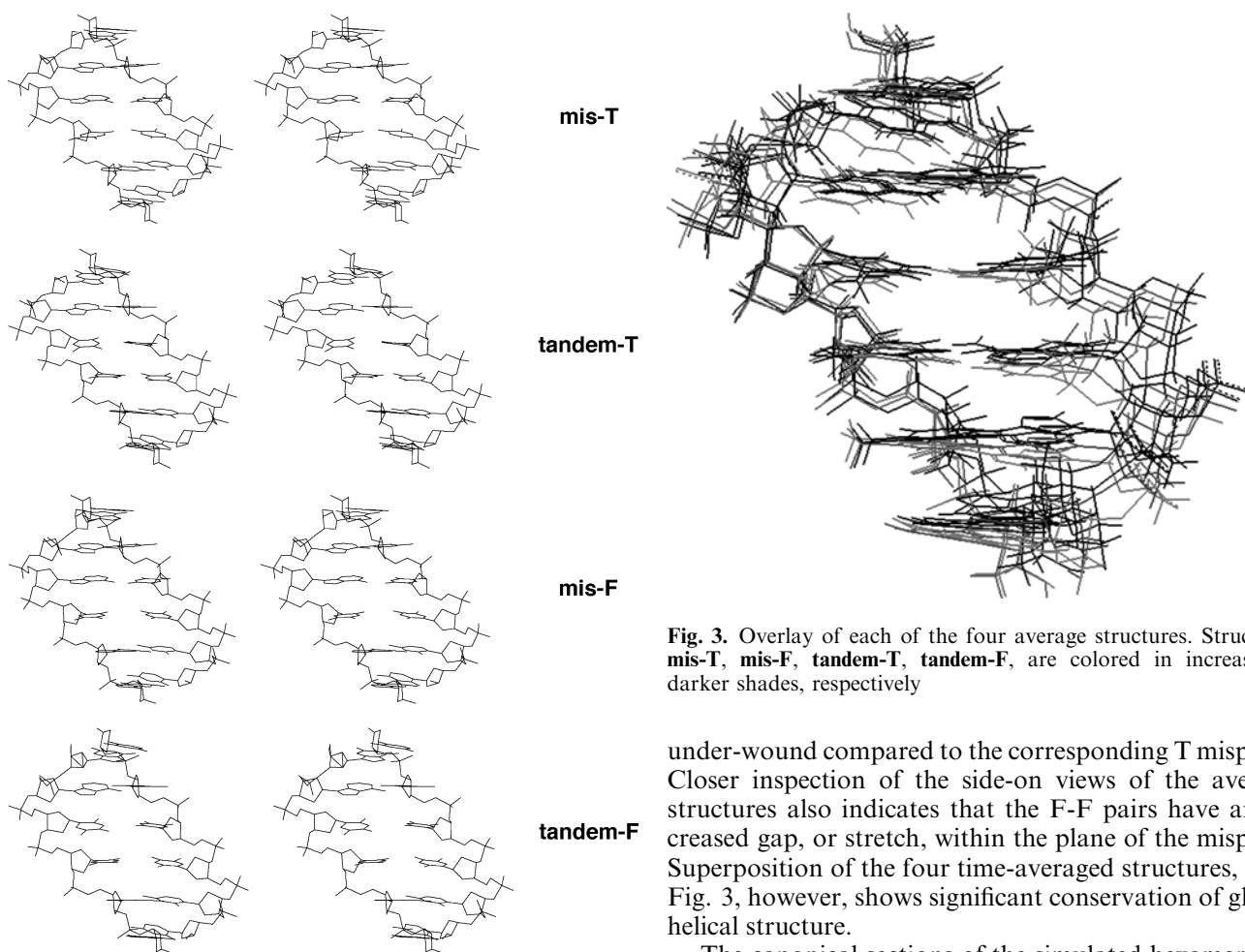


Fig. 2. Time-averaged structures for mismatch trajectories (hydrogen atoms have been removed for clarity)

Fig. 3. Overlay of each of the four average structures. Structures **mis-T**, **mis-F**, **tandem-T**, **tandem-F**, are colored in increasingly darker shades, respectively

under-wound compared to the corresponding T mismatches. Closer inspection of the side-on views of the average structures also indicates that the F-F pairs have an increased gap, or stretch, within the plane of the mismatches. Superposition of the four time-averaged structures, as in Fig. 3, however, shows significant conservation of global helical structure.

The canonical sections of the simulated hexamers are not particularly interesting: they show little deviation from standard B-form DNA structure even when the

behavior of the mispairs is taken into account. We note that incorporations of one A-F or Z-F pair have already been shown to cause little overall structural deformation in surrounding helices, although a higher base pair opening rate for A-F pairs is evident, and a somewhat greater C1'-C1' distance for the experimentally determined Z-F pair has been observed [21, 22, 31, 47].

Concerning the pyrimidine-pyrimidine mismatches, experimental structural information is available for T-T mispairs, but not for F-F mispairs. Several NMR studies have been performed on helices containing single T-T mispairs, and multiple T-T mispairs [89–91]. Not only do these studies show that very little structural distortion is induced by the mismatches, it is also evident that nearest-neighbor base pairs are relatively unaffected by the mispair. Mismatches composed of two thymine nucleotides adopt *anti* glycosidic configurations, stack within the helix, and form wobble base pairs. Since the T-T base pair can form a wobble pair in one of two orientations, which are related by a 180° rotation about the pseudodyadic axis, and for which a simple shearing movement within the plane of the base pair can effect this change, there has been some difficulty in assigning which pairing scheme is dominant. Some support is seen for a single dominant configuration, but the existence of a single set of resonances for the T-T pair might indicate a very rapid exchange between configurations [90]. While the single thymine mismatches were investigated in the context of DNA, a run of four consecutive T-T mispairs has been studied in a section of hexitol nucleic acid (HNA) [91]. The T-T pairs in this duplex adopted both wobble configurations, reduced the phosphate backbone separation between strands, reduced the *x*-displacement, and induced a bend in the central section of the mispairs. In addition, the pairs induced slight over-winding of the helix at the junctions on either end of the four mispairs, and slight under-winding of the base pair steps within the mispair stretch.

This same winding behavior is evident in our simulations (Table 1). Helix twist values for the base pair steps flanking the T-T mispairs of trajectory **tandem-T** of 39.9° and 40.3° indicated significant over-winding, and

the twist value between the T-T pairs has an average under-wound value of 14.5°. Similar to the HNA study, both configurations of the T-T wobble pair were observed during the MD runs. Roughly similar over-winding-around/underwinding-between behavior is seen for **tandem-F**. A particular feature of F mispairing vs T is the greater dynamic behavior of twisting observed for the former, with standard deviations in the F twist values being two to three times those observed for T. We may compare the mismatch values to those from the control sequence **AT**, which has typical B-form twist values of (31.3 ± 4.8)° above the 3:10 step, (29.4 ± 3.9)° between 3:10 and 4:9, and (31.7 ± 4.4)° below the 4:9 step. Block averages over 1 ns sub-blocks of each trajectory do not show significant deviations from the overall average values discussed above (see Supporting Information for **AT** data and block averages).

Mismatch geometries

After averaging the structural information contained in each trajectory, the configuration of each individual mispair was examined (Fig. 4). Each T-T mispair adopted a tight wobble geometry, and it is evident that the two T-T mispairs of **tandem-T** adopt oppositely related configurations. The F-F mispairs are more varied. Not only are the intermolecular distances significantly increased, the F-F pair from **mis-F** does not show any preference for wobbling in one direction or the other. This results in an average structure which takes on an almost Watson-Crick appearance. In the case of **tandem-F**, the two F-F mispairs appear to wobble as in the T-T trajectories, however, this is not due to favorable hydrogen bonding interactions. The apparent wobble geometry of the F-F pairs is not supported when the hydrogen bond lengths are monitored as a function of simulation time (Fig. 5). While the T-T mispairs maintain fairly rigid wobble geometries, the F-F pairs are seen to be highly fluxional in every case. As is seen in similar plots for A-T and G-C pairs, the hydrogen bond plots for the T-T pairs adopt stable values and little deviation is seen from an average value. It has previously

Table 1. Helical parameters for average structures of different hexamers^a

	mis-T T3:T10		mis-F F3:F10		tandem-T T3:T10		tandem-T T4:T9		tandem-F F3:F10		tandem-F F4:F9	
Buckle (°)	-1.6	(9.0)	-4.8	(12.7)	2.6	(10.7)	-6.5	(11.6)	-5.8	(15.3)	-9.8	(15.4)
Opening (°)	9.4	(9.4)	4.8	(18.1)	13.2	(6.9)	13.9	(8.6)	5.9	(17.2)	10.3	(15.9)
PrpTwst (°)	-26.6	(10.7)	-17.9	(13.8)	-15.4	(10.9)	-17.3	(10.8)	-15.4	(14.4)	-18.7	(13.3)
Rise (Å) ^b	3.3	(0.4)	3.6	(0.6)	3.6	(0.4)	3.5	(0.6)	3.9	(0.6)	3.6	(0.8)
	3.7	(0.4)	3.6	(0.6)	3.5	(0.6)	3.6	(0.4)	3.6	(0.8)	3.6	(0.5)
Roll (°) ^b	10.0	(7.5)	9.7	(9.1)	5.0	(7.4)	9.6	(6.9)	5.7	(8.6)	8.6	(8.0)
	0.6	(6.8)	2.1	(8.0)	9.6	(6.9)	-0.3	(6.4)	8.6	(8.0)	1.4	(7.6)
Shear (Å)	-2.2	(0.4)	-0.6	(1.9)	2.3	(0.5)	-2.4	(0.5)	2.0	(1.5)	-1.4	(1.7)
Stagger (Å)	0.5	(0.6)	0.0	(1.1)	0.4	(0.6)	0.7	(0.7)	-0.2	(1.1)	0.1	(1.2)
Twist (°) ^b	21.9	(4.4)	27.5	(12.1)	39.9	(4.4)	14.5	(4.4)	40.1	(7.9)	17.6	(12.3)
	43.7	(4.4)	33.8	(12.1)	14.5	(4.4)	40.3	(4.1)	17.6	(12.3)	35.4	(9.2)

^a Values are averages (standard deviations) for TT or FF pair unless otherwise specified

^b Values are for base steps above/below TT or FF pair

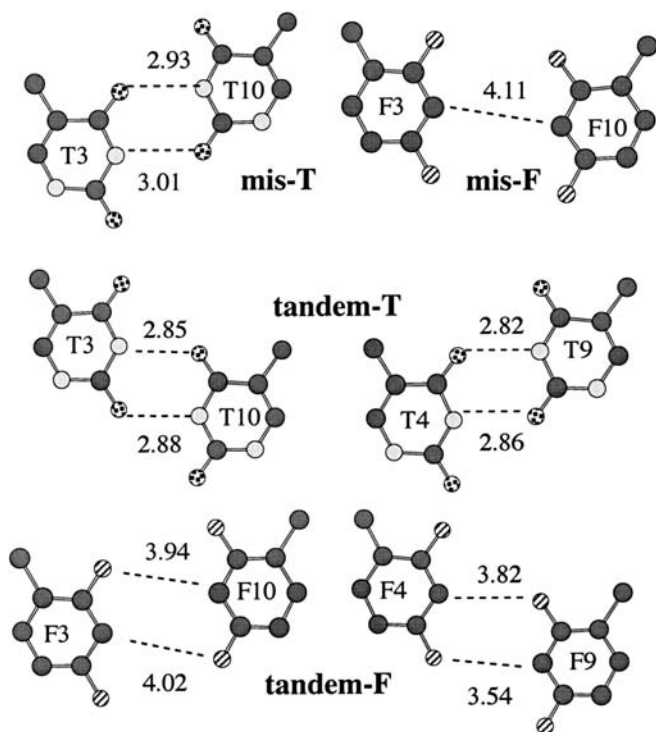


Fig. 4. Mispair geometries taken from average structures. Key heavy-atom bond lengths are reported in Å. The lower edge of each base pair faces the minor groove, and hydrogen atoms have been removed for clarity

been noted that A-F pairs show a high base-pair opening rate, initially identified by an oscillatory plot not unlike those depicted for the F-F pairs in Fig. 5 [31, 32, 47]. It is noteworthy that the T-T pairs do not switch between configurations during the course of trajectories **mis-T** and **tandem-T**. This lends some support to the observation of a single set of NMR resonances for the T-T mispair, indicating a single configuration. However, it would be necessary to extend the time of the current trajectories by orders of magnitude in order to fully test this idea.

Base stacking

The analysis method of Nagan et al. [56], which involves monitoring the center-of-mass (COM) distances between consecutive base pairs both in an intra-strand and inter-strand sense, provides information about the preferred stacking arrangements within a nucleic acid helix. Since the average rise between a perfectly stacked base step can be estimated at roughly 4.0 Å, deviations from this value are indicative of either enhanced or reduced stacking. COM distances were calculated for the six-membered rings of the pyrimidines as well as the six-membered rings of the purines (identical conclusions are reached here if one instead chooses to monitor distances involving the five-membered ring of the purines). Plots were made of the two COM distances for each base over time, and distributions of points within four quadrants of the plot were

taken as indicative of certain stacking interactions (see Supporting Information). As an example, in trajectory **mis-T**, T3 can stack either on C2 or G11. A graph of one COM versus the other shows a distribution of points grouped rather tightly about the 4.5–5.0 Å distance in both cases. This implies that the T3 base does not stack entirely on either of its 5' partners, but instead nearly halfway between the two. This information is also clearly observed in the time-averaged structure. From the COM plots, it is also evident that all of the thymine stacking interactions are associated with tightly grouped “shot patterns” in terms of being narrowly distributed over the possible range of distances. Corresponding *average* F stacking interactions are observed to be nearly identical to those for T, which is an interesting result. A noteworthy difference between the F and T plots, however, is that every F-stacking shot pattern has a much larger distribution in distance space, which indicates larger dynamic variation (see Supporting Information).

A simplified view of the stacking geometries is provided in Fig. 6, where the canonical pairs above and below the mismatches (green) have been color-coded yellow. Starting with **mis-T**, it is evident that little stacking exists between the T-T pair and the pair 5' to T3. This loss of stacking is made up for by the nearly full stacking of the T-T pair on each of the bases 3' to T3. The stacking of G11 between T3 and T10 is reminiscent of a cross-strand purine stack, and examples of this stacking arrangement are seen in other stacking interactions in Fig. 6. Cross-strand stacking was a pivotal motif observed in an internal mispaired RNA loop formed by two G-U wobble pairs, and two C-U pairs [59]. The stacking arrangement adopted for **AT** is found in the Supporting Information. While the stacking interactions flanking the **AT** pairs at positions 3:10 and 4:9 are qualitatively similar to the mismatches stacking interactions, there is a much stronger stacking interaction formed between the 3:10 and 4:9 step in **AT**, as would be expected for this canonical sequence.

While the 5' arrangement for the F-F pair of **mis-F** is identical to that seen in **mis-T**, the 3' interactions are somewhat reduced, possibly owing to the increased intermolecular F-F distance. The tighter T-T wobble pair holds T3 more directly over its 3' neighbor. These observations hold true for base steps 5' and 3' to the double mismatches for both **tandem-T** and **tandem-F**. Similarly interesting is the distribution of pyrimidine bases when looking down on the base step corresponding to a mispair-mismatch step. In both the T-T and F-F steps of **tandem-T** and **tandem-F**, no full overlap of pyrimidine rings is evident. These results fully agree with those found in the NMR structure of the HNA containing four consecutive T-T mismatches [91]. The helix twist values encompassing the four T-T pairs calculated for the HNA structure were 40.6, 26.7, 11.4, 26.8, and 40.6°. Tandem U-U base pairs adopt a similar nucleotide base stacking arrangement as that observed in Fig. 6 [92, 93].

As was the case with two consecutive T-T mismatches, when one T-T mismatch is embedded in a helix, the structural distortion is manifested in

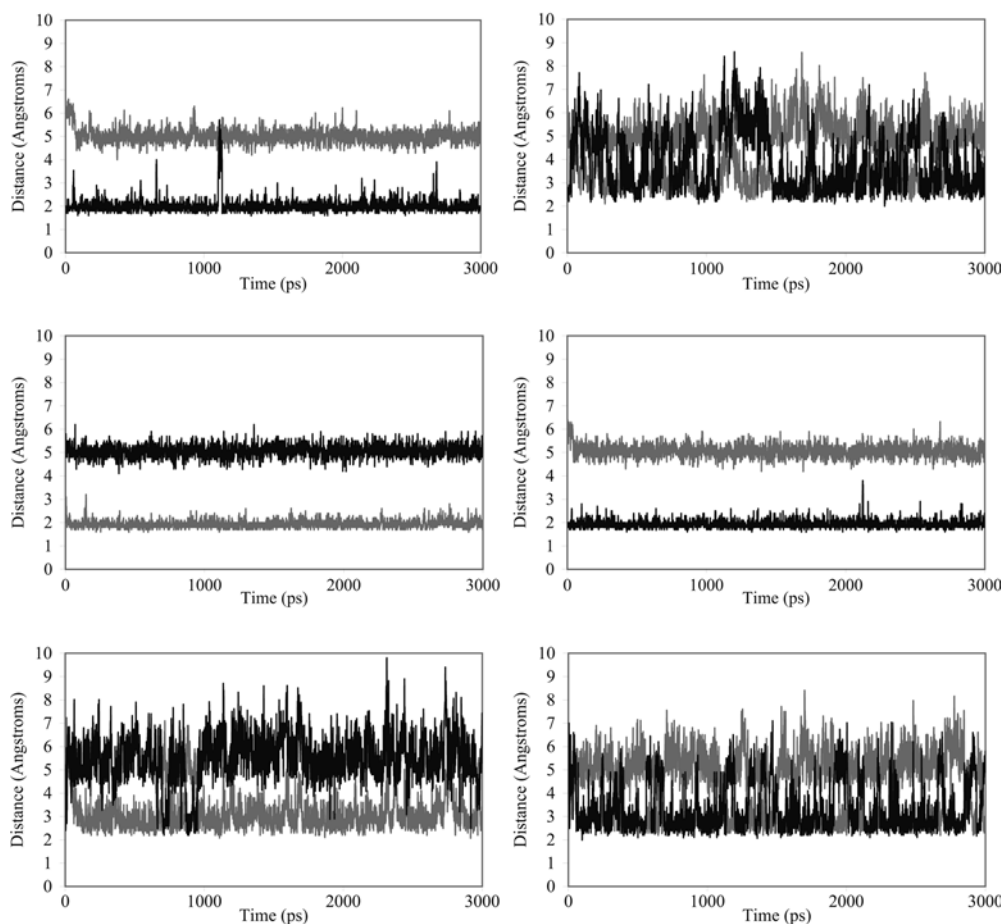


Fig. 5. Hydrogen bond lengths (over the last 3 ns) for each mispair as a function of time. Trajectory **mis-T** (upper left) bonds are T3:O4–T10:H3 (black) and T3:O2–T10:H3 (gray). Trajectory **mis-F** (upper right) bonds are analogous to **mis-T**, but using F4 and F2 atoms, respectively. Trajectory **tandem-T** (middle) bonds are equivalent to **mis-T**, T3 and T10 (left), T4 and T9 (right). Trajectory **tandem-F** (bottom) bonds are analogous to **tandem-T**, but using F4 and F2 atoms

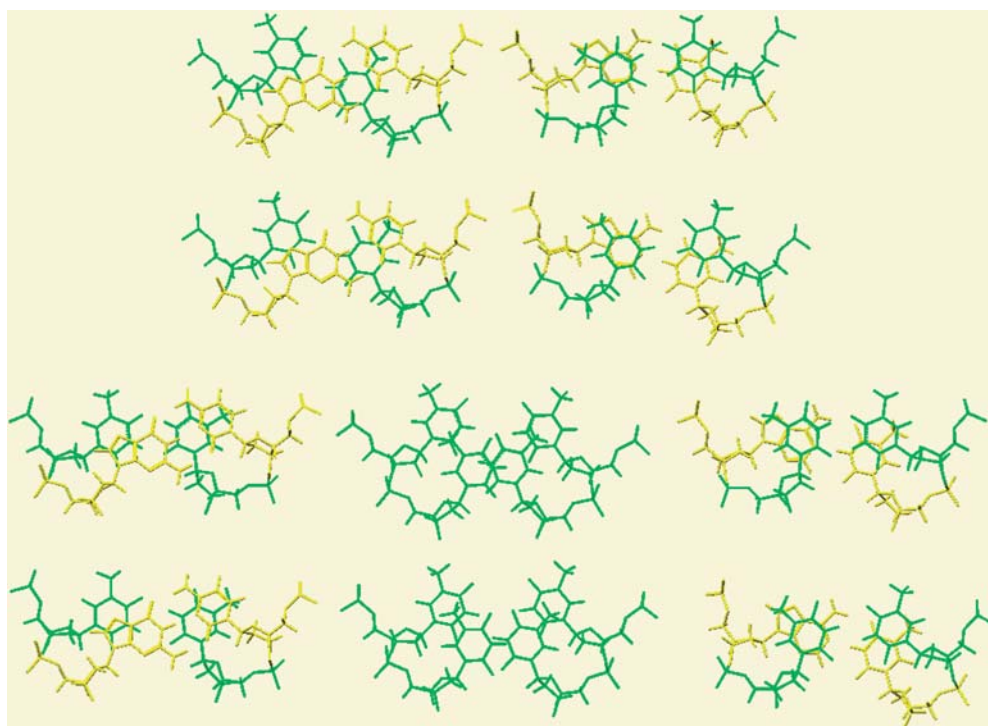


Fig. 6. Base stacking arrangement of mispairs (green) and their flanking partners (yellow). The 5' step to the left, the 3' to the right. The ordering of the figure from top to bottom is **mis-T**, **mis-F**, **tandem-T**, **tandem-F**. The 3:10 base mispair is always depicted on the far left-hand side

over-winding and under-winding of the helix on either side of the mispair. The over-winding of the helix takes place at the 3' step in relation to T3 for **mis-T**, according to Table 1. In the single F-F simulation, the deviations from standard twist values were less pronounced. In studies of G-U wobble pairing, such base step twisting distortions are also evident. When considering consecutive G-U wobble pairs, it is possible to form three motifs U-G/G-U, G-U/U-G, and U-G/U-G, where the first base pair is 5' to the second listed in each grouping [94, 95]. Whether certain base steps above or below the mispairs are under-wound or over-wound is dependent upon the sequence of the tandem pairs [94]. The qualitative amount of base stacking for each tandem G-U motif, as determined from plots like that shown in Fig. 6, correlates well with the thermodynamic stabilities of each motif. Several uracil bases in each of the motifs exist in an unstacked configuration, similar to what is seen with the T-T to T-T step of **tandem-T**. Since different ordering of G-U wobble pairs, which are unable to adopt two degenerate configurations (as with T-T), causes switching of whether a base step adjacent to the mispairs is under-wound or over-wound, it is interesting to speculate what would happen with the switching of T-T wobble configurations. In the case of **mis-T**, one T-T configuration was sampled, and the over-winding was 3' to the T3 base. If the T-T configuration was formed with a hydrogen bond to O2 of T3 instead of O4, over-winding of the helix step would instead occur 5' to T3. In any case, an alteration of the twist angle is induced in order to maximize the stacking interactions for one or more base steps about the mispairs.

While the F-F mispairs, unlike T-T pairs, lack favorable intermolecular hydrogen bonding, the increased propensity for F to stack in comparison to T prevents the F bases from sliding into the grooves without penalty. Stacking of F as a dangling residue on a DNA duplex has an increased favorable free energy contribution of 1.5 kcal mol⁻¹ compared to T [96]. Decreased stacking favorability of T is attenuated, however, by a large gain in interaction energy between the bases in the T-T wobble pair. A theoretical estimate of the difference in the gas-phase interaction enthalpy for the two mispairs is -9.4 (experimental value: -9.0 ± 1.0 [97, 98]) versus +2.2 kcal mol⁻¹ for T-T and F-F, respectively [99]. These energies are calculated at the *mPWPW91/MIDI!* density functional level, which has a mean unsigned error of 0.2 kcal mol⁻¹ relative to available experimental gas phase thermochemical data for base pairing.

Hexamer dynamics

While the static structural information is interesting, it is clear that the mispairs exhibit substantial dynamic range. Not only do the COM plots appear more widely dispersed for any step involving F, nearly every geometrical parameter measured and reported in Table 1

has a higher standard deviation when F has been incorporated. Base pair opening can be directly correlated with the strength of the interaction energy between two bases. For T-T, the average base opening angle was (9.4 ± 9.4)°, which is a much smaller standard deviation than (4.8 ± 18.1)° for F-F. This same trend is seen for the double mispairs. An increased variability in twist values is also pronounced for **mis-F** and **tandem-F**. Shear and stagger parameters provide support for a more stable wobble pairing arrangement for all of the T-T pairs, and higher mobility of the F-F pairs.

Calculated average structures from trajectories **mis-T** and **mis-F** have a rather low RMSD of only 0.46 Å (Table 2). Slightly higher variation was found between the double mispair average structures. Since the trajectory pairs **mis-T/mis-F** and **tandem-T/tandem-F** contain the same number of atoms, catenated trajectories can be created, and a composite average structure can be calculated. The RMSD between each individual average, and the average of the two catenated trajectories was very low, again supporting the argument that little structural distortion outside the mispairs is taking place. Table 2 contains the average RMSD calculated over each 3 ns trajectory in reference to either the initial starting structure, or the time averaged structure. While the standard deviations are nearly conserved in each case, the average dynamic variability in any F-containing helix is slightly higher. Dynamic variability is driven by specific motions, and the method of principal component analysis allows these motions to be identified.

Principal component analysis

Principal modes of motion in each of the catenated trajectories were calculated and analyzed; results are summarized in Table 3. As with other nucleic-acid simulations, the dominant modes of motion are helix twisting, bending, and breathing modes. In each case, nearly 50% of the overall motion of the trajectories is

Table 2. Heavy-atom RMSDs (Å) for mispair trajectories relative to different standards^a

Trajectory	mis-F	tandem-F	mis-T/mis-F ^b	tandem-T/tandem-F ^b
mis-T	0.46	–	0.23	–
mis-F	–	–	0.23	–
tandem-T	–	0.61	–	0.30
tandem-F	–	–	–	0.31
Trajectory	RMSD vs. Initial Structure	RMSD vs. Average Structure		
mis-T	2.3 (0.3) ^c	1.0 (0.2)	–	–
mis-F	2.4 (0.4)	1.2 (0.3)	–	–
tandem-T	2.8 (0.4)	1.0 (0.2)	–	–
tandem-F	2.6 (0.4)	1.3 (0.3)	–	–

^a From minimized average structures over final 3 ns

^b From catenated trajectories

^c Reported as average (standard deviation)

Table 3. Principal component analysis for catenated trajectories **mis-T/mis-F** and **tandem-T/tandem-F**

Trajectory		Analysis			Description
mis-T/mis-F		Coefficient ^a			
Mode	Percent	mis-T	mis-F		
1	21.0	-57.6 (8.4)	-60.8 (11.7)	helix twist	
2	13.7	-24.3 (6.4)	-29.6 (9.2)	helix bending	
3	10.9	-26.1 (7.5)	-24.8 (7.3)	minor groove breathing	
Subtotal	45.6	-	-	-	
tandem-T/tandem-F		Coefficient ^a		Description	
Mode	Percent	tandem-T	tandem-F		
1	22.1	-75.4 (8.4)	-67.7 (11.4)	helix twist	
2	14.4	26.6 (6.1)	31.7 (8.1)	minor groove breathing	
3	11.8	9.2 (5.8)	13.1 (9.0)	helix bending	
Subtotal	48.3	-	-	-	

^a Reported as average (standard deviation)

reproduced by the first three modes. Each mode of motion, or eigenvector, may be considered to be a coordinate in a reduced dimensionality space in which specific trajectory snapshots may be assigned coordinate positions, the coordinates being the coefficients of each mode required to best reproduce the snapshot geometry.

Plots of these coefficients are well-behaved, and indicate the trajectories themselves to be well equilibrated (no drift observed). Coefficient histograms for each component are Gaussian in shape, again indicating a well equilibrated trajectory (Fig. 7). Each catenated trajectory of 6 ns is composed of two halves, each of 3 ns. Therefore, a histogram of the first 3000 coefficient values from the combined trajectory **mis-T/mis-F** provides dynamic information for trajectory **mis-T**. Significant overlap in each histogram indicates that the

different structures have essentially identical dynamics. In both cases, however, a widening of the distribution is seen when F replaces T. This is quantified in Table 3, where the standard deviation of all the coefficients is higher for the F containing portion of the catenated trajectories.

Principal component 1 (PC 1) – the most dominant motion of the helix – is helix twist in all cases. In both the single and double mispair catenated trajectories, low twist was characterized by more negative coefficients, and vice versa. A single F-F mispair induced a smaller average helical twist compared to T-T, but the double T-T mispair had a lower average twist as compared to two F-F pairs. While the second and third components for the two catenated trajectories were minor groove breathing and helix bending, the ordering was not the

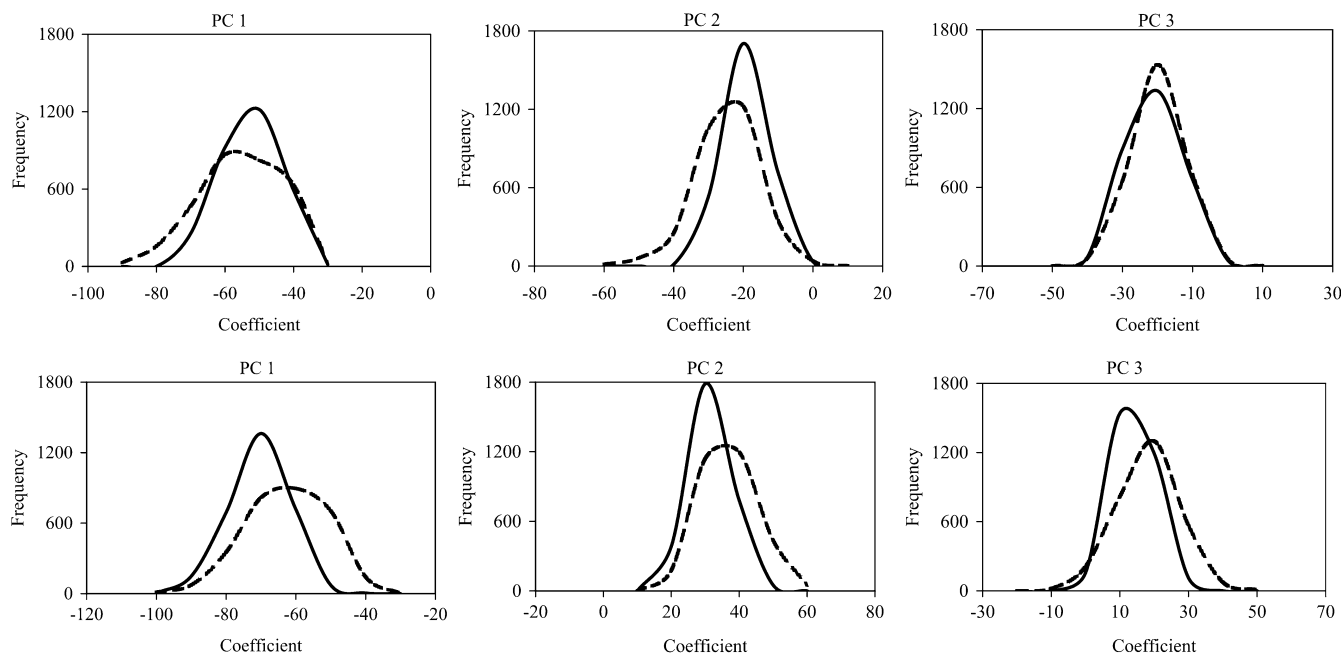


Fig. 7. Histograms of the principal component coefficients for the catenated trajectories **mis-T/mis-F** (top) and **tandem-T/tandem-F** (bottom). Helices containing F-F pairs are drawn as dotted curves, T-T as solid curves

same. Helical deformations characterizing the limits of each mode can be observed in Figures 8 and 9.

Helix bending was the second most dominant motion for **mis-T/mis-F**. A slight compression of the minor groove in the region of the mispair causes the ends of the helix to reversibly bend into the major groove. When the coefficient takes on less negative values, the mispair exists in a canonical wobble configuration. At the more negative extent, the major groove is wide, and the mispair is translated into the groove to some extent. This helical conformation is characterized by a more Watson-Crick mispair geometry, much like the average F-F base pair seen in Fig. 4 for **mis-F**. The distribution of coefficients for **mis-T** is indicative of a tighter wobble pair for T-T compared to F-F.

Minor groove breathing dominates PC 2 for **tandem-T/tandem-F**. Larger coefficients correlate with a wider minor groove. A narrowing of the minor groove brings T4 into closer proximity to T10, which represents the formation of a cross-strand pyrimidine stacking arrangement. While the T bases are not fully stacked, Fig. 6 does show more stacking of these two bases over each other as compared with the corresponding F

nucleotides. As would be expected, the distribution of coefficients for this mode shifts to a lower value for **tandem-T**. If structures composing trajectory **tandem-T** are to have narrower minor grooves, the C1'-C1' distances for the T-T pairs must be shorter than those for the F-F pairs. The data presented in Fig. 10 show this trend. Not only are the C1'-C1' distances shorter for T-T than a representative canonical pair from the simulation, they are much more stable than the F-F distances plotted. The groove width as determined by the C1'-C1' distance of either of the F-F pairs is highly dynamic.

Component three for **mis-T/mis-F** was minor groove breathing. Less negative coefficients of this mode induced the helix to have a higher twist value, a slightly more compact appearance, and again contained the Watson-Crick type F-F arrangement. When the eigenvector transitioned to more negative values, the mispair began to wobble to a greater extent. This movement was in concert with a buckling of the mispair and the bases on either side of it. As is evident in Fig. 8, this ends up bending the C6-G7 base pair towards the minor groove. The minor groove of this end of the helix adopts wider conformations, as well. Since the average coefficient of

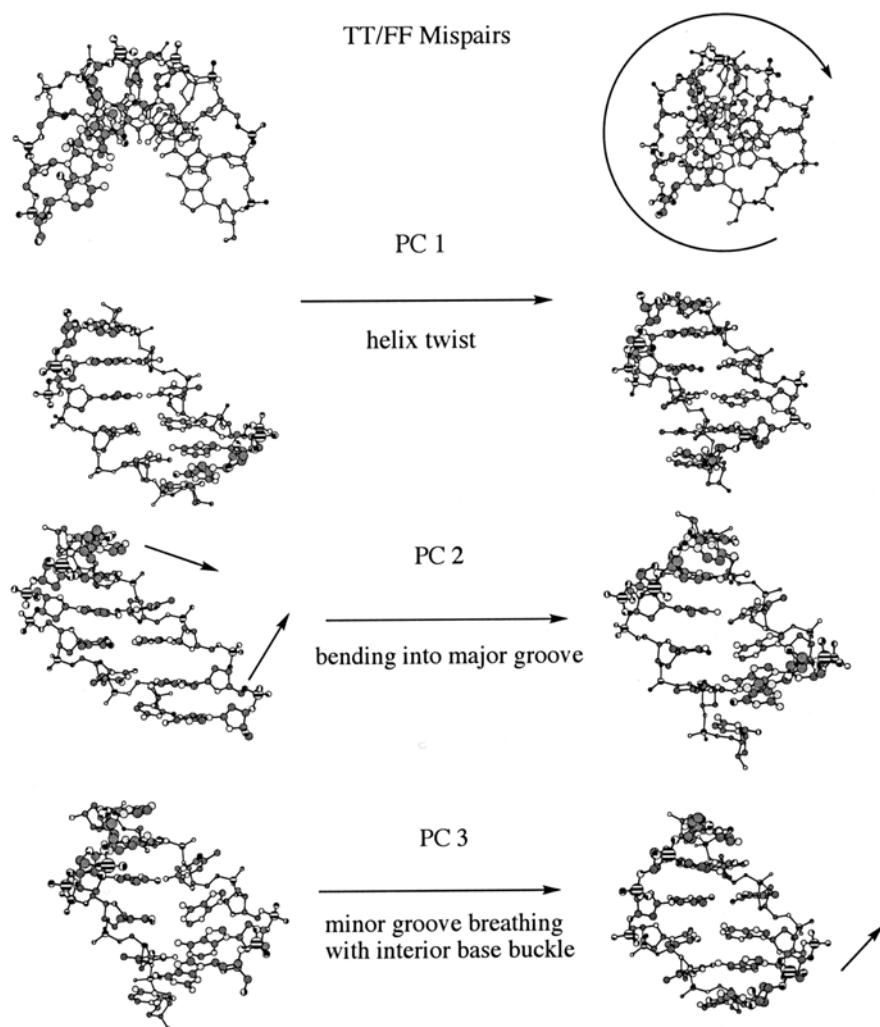


Fig. 8. Structural deformations found at the limit of each principal component for **mis-T/mis-F**

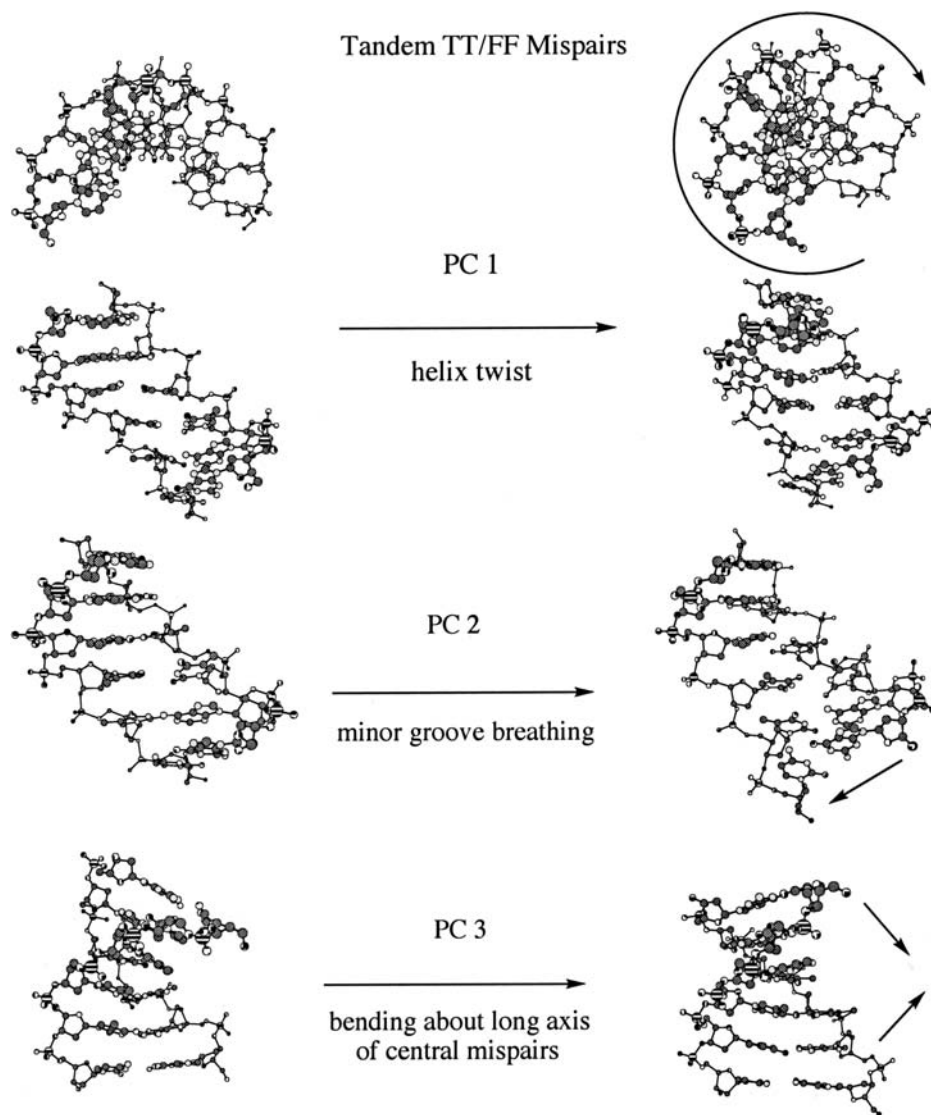


Fig. 9. Structural deformations found at the limit of each principal component for **tandem-T/tandem-F**

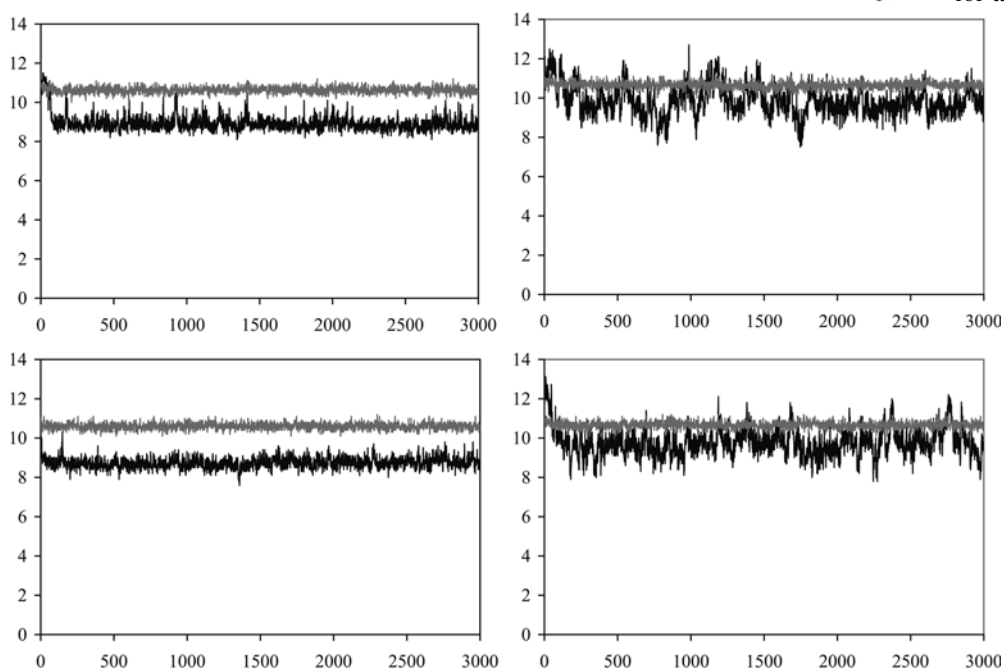


Fig. 10. Plots of C1'-C1' distances (Å) for the mispairs (over the last 3 ns) vs. time. The upper graphs are for **mis-T** (left) and **mis-F** (right). The bottom curves represent averages over both mispairs of **tandem-T** (left) and **tandem-F** (right). Black lines in each plot correspond to the mispair distances, while the C1'-C1' distance of a canonical G-C pair has been plotted in gray for comparison

mis-T is slightly more negative than that for **mis-F**, the T-T pair is again observed to wobble to a larger extent.

Bending of the helix axis about the middle of the two mispairs was the characteristic motion of PC 3 for **tandem-T/tandem-F**. A more compressed minor groove and opened major groove was observed when the coefficient adopted more negative values. This behavior was seen for **tandem-T** to a larger extent and correlates well with the shorter C1' distances for T-T pairs.

Trajectories containing one mispair did not contain the same number of purines and pyrimidines as the double mispair simulations. For this reason, the effects of adding an additional mispaired T-T or F-F pair to an already existing pair can only be inferred from the results of the physically meaningful catenated trajectories. The variance in the adopted coefficients of PC 2 and PC 3 for **mis-F** were decreased in comparison with **mis-T**. It would follow from this result that the existence of two T-T mispairs is stabilizing to the helix in relation to only one. In part, this structural stability arises from an increased stacking interaction in the mispaired region induced by the cross-strand stack of thymines. A measure of the structural variability of a molecule is obtained by calculating the configurational entropy. The final 3 ns for trajectories **mis-T-tandem-F** had configurational entropies of 4209, 4136, 4374, and 4454 cal mol⁻¹ K⁻¹, respectively. It is evident that a slight

structural stabilization exists for **tandem-T** over **mis-T**. Both simulations containing F-F pairs are more dynamic. Given the greater dynamic freedom of any helix containing F-F pairs, it is interesting to examine why the two hexamers containing T-T pairs may be more constrained.

Hydration analysis

The hydration of nucleic-acid bases and nucleic-acid structures is of fundamental interest, and much work has been done focusing on specific hydration of individual nucleotide functional groups, as well as the identification of motifs such as spines of hydration [100–105]. The overall distribution of water molecules around a macromolecule is somewhat difficult to measure. Hydration analysis usually involves calculating water residence lifetimes, or calculating water occupancies on a given grid built around the molecule. Although rigid solutes do not move associated water molecules into different grid locations during a water analysis, structural waters bound to highly mobile helical components can be moved from grid location to grid location, effectively washing out some high density information. Nevertheless, hydration plots here are still of interest (Fig. 11). Hydration of the helices has been superimposed onto

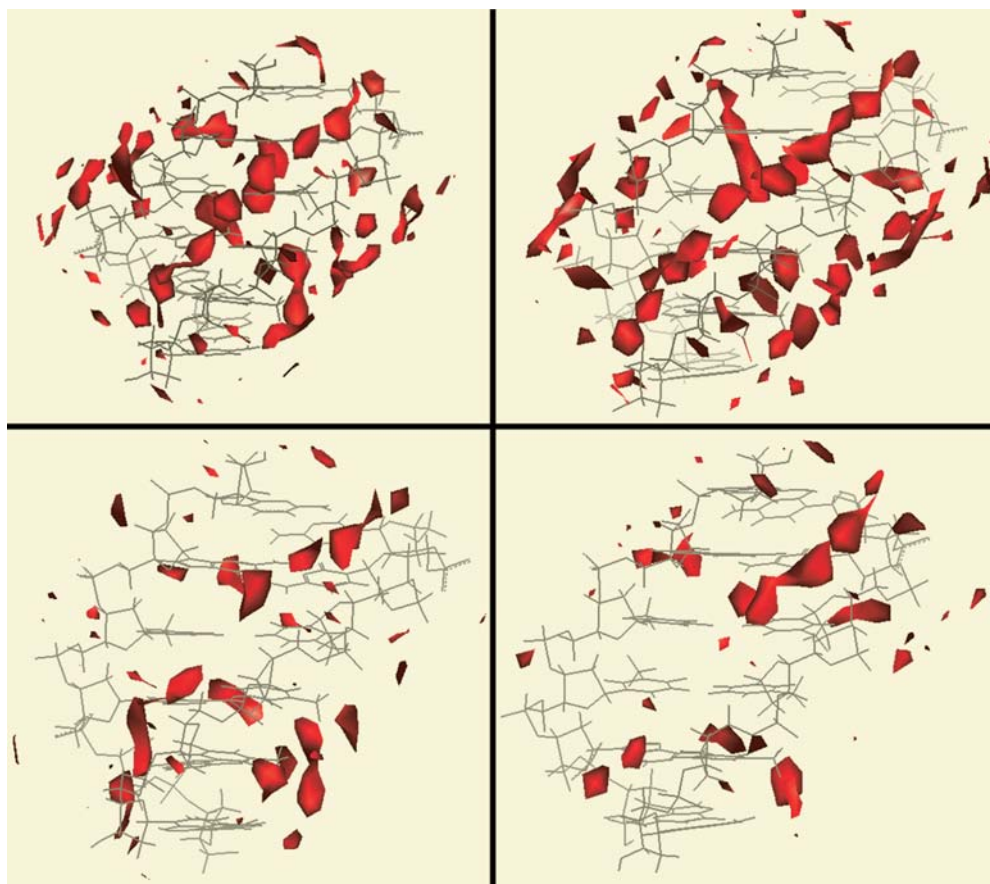


Fig. 11. Water density contoured at a level twice that of bulk water for **mis-T** (upper left), **tandem-T** (upper right), **mis-F** (lower left), and **tandem-F** (lower right)

each of the four time-averaged structures with a contour level corresponding to roughly twice the density of bulk water. A substantial number of regions of high-density water are observed for structures **mis-T** and **tandem-T**. Though the greater structural rigidity of **tandem-T** allows for a denser hydration plot, further analysis argues for similar hydration characteristics (see below).

The hydration of helices containing F-F pairs is much less pronounced, not only in the immediate area of the mispair(s), but over the entire helix. What is most clearly apparent is a distinct lack of hydration of F, which is manifest in gaps in the major and minor groove spines of hydration where waters are clearly present for the T-containing helices.

The strong solvation of T-T mispairs is not unexpected. Wobble pairing of T-T causes T carbonyl groups to be thrust into the grooves of the helix where

the exposed oxygen atoms may act as hydrogen bond acceptors. Water molecules were observed to have densities of nearly seven times the normal density of bulk water near the T-T and double T-T mispairs.

Another method of tracking hydration differences is to calculate radial distribution functions (r.d.f.s) about relevant functionality. This type of plot depicts solvation shells and permits determination of the total number of water molecules in each solvation shell. Such r.d.f.s calculated for O and F atoms in T-T and F-F pairs are provided in Fig. 12. As expected, the solvation of F is essentially featureless, and bumps in the plot are most likely residual contributions from water networks built up around more hydrophilic portions of the grooves. Both the major and minor grooves of the T-T pairs are solvated. In each of the three T-T mispairs, the major groove O4 not involved in the wobble hydrogen bonding motif is solvated to the greatest extent. Next best

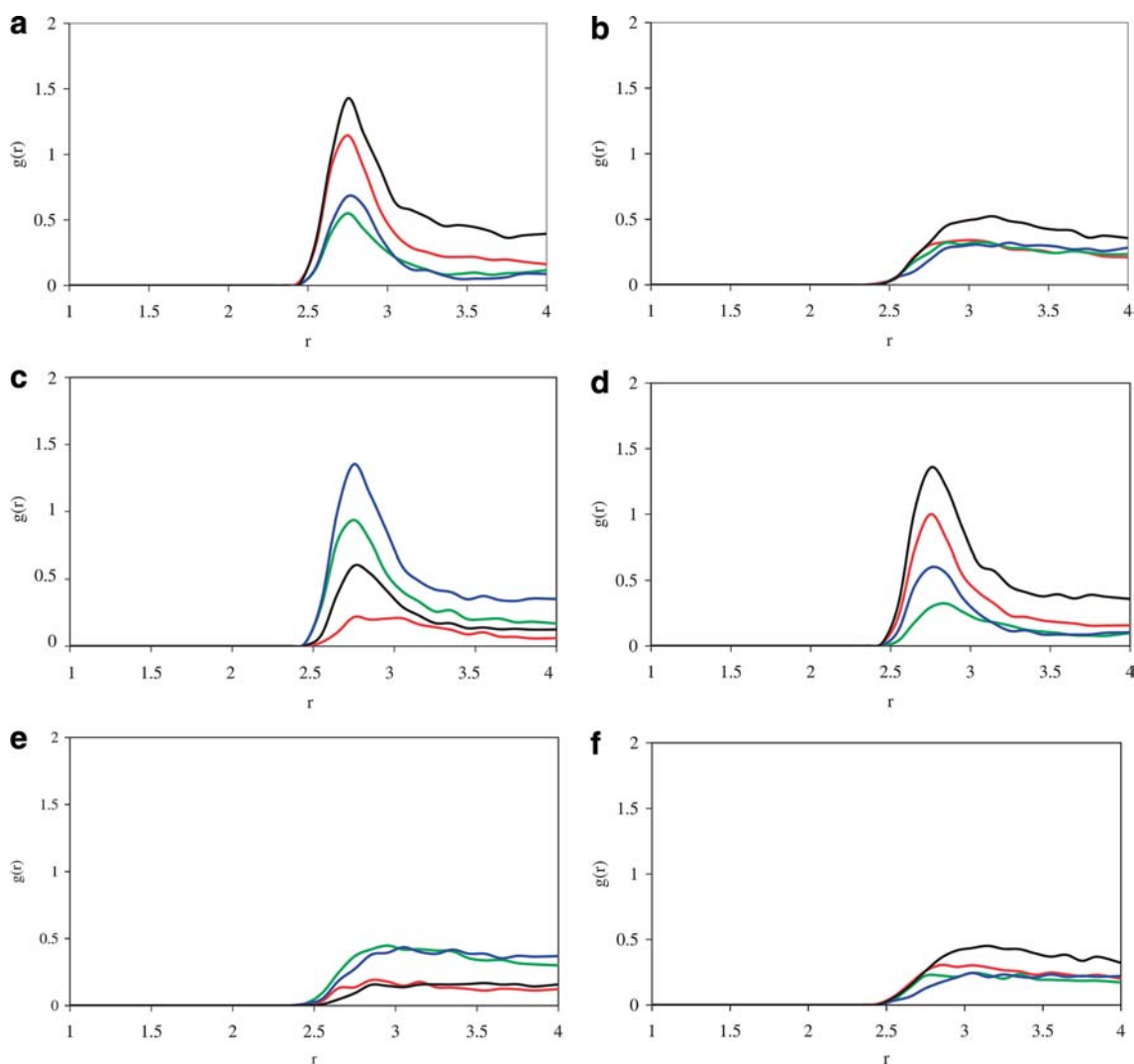


Fig. 12. Radial distribution functions for water O atoms about minor groove and major groove exposed carbonyl oxygens (T) or fluorines (F) of the mispaired bases. **a:** **mis-T**; **b:** **mis-F**; **c:** **tandem-T 3:10**; **d:** **tandem-T 4:9**; **e:** **tandem-F 3:10**; **f:** **tandem-F 4:9**. Curves are colored red (Y3/4:O2/F2), green (Y9/10:O2/F2), blue (Y3/4:O4/F4), and black (Y9/10:O4/F4), where Y is either pyrimidine

solvated is the minor groove O2 that also is not involved in T-T hydrogen bonding, then the major groove O4 that is a hydrogen-bond acceptor in the wobble pair, and finally the least solvated carbonyl oxygen is the minor groove O2 that is directly involved in the wobble interaction. Irrespective of whether the carbonyl group is involved in the wobble pairing or not, the larger size of the major groove allows more waters to bind as compared to the minor groove.

[As a technical point, we note that the r.d.f. plots of Fig. 12 do not show an asymptotic approach of $g(r)$ to 1.0 over the distance plotted. Because the helix itself screens solvent molecules to a significant extent, bulk water densities are not found around the carbonyl groups until distances of about 10–15 Å; r.d.f. plots out to 15.0 Å are provided in the Supporting Information.]

Integration of the number of water molecules in the first solvation shell (Table 4) indicates T-T mispair carbonyl groups to be solvated in the minor and major groove by 0.83 and 0.65 waters, respectively. These values decrease when considering carbonyl oxygens already hydrogen bonded to partner pyrimidines. No F base was ever solvated by more than 0.22 waters in the first solvation shell.

Implications for DNA replication

Owing in part to the high steric similarity of the two mispairs, relatively little large-scale variation in the overall dynamical behavior of the T-T and F-F trajectories is evident. This is manifest in the low RMSDs between average structures calculated over the 3 ns trajectories. However, closer inspection does reveal some important variations.

Replication begins with the correct pairing of nucleic acid bases in a tight hydrophobic pocket which helps to align the 3' hydroxyl group of the primer base with the phosphate group of the incoming dNTP for successful phosphodiester bond formation [5, 8, 43, 106]. For correct alignment in the transition state, the nascent base pair should have a C1'-C1' distance of approximately 10.5 Å [107]. This is precisely what is observed for canonical pairs present in the hexamers, but from Fig. 10, it is evident that the mispairs adopt different

geometries. The T-T C1'-C1' distance during the simulation averaged 8.9 ± 0.4 Å. When two T-T mispairs were tandem to each other, this C1'-C1' distance was further decreased to 8.7 ± 0.4 Å. Single and double F-F C1'-C1' distances adopted not only longer distances, but more dynamic ranges of 9.9 ± 0.8 Å and 9.7 ± 0.7 Å, respectively. This suggests that F-F incorporation should have a higher efficiency than T-T, because of this pair's preference for adopting an intermolecular separation close to that for Watson-Crick base pairs.

In addition, the smaller C1'-C1' distance calculated for the pyrimidine mispairs reduces minor groove widths, and multiple mispairs magnify this narrowing. The information contained in the principal components analysis supports a narrower minor groove for helices containing T-T in comparison to F-F.

According to Kool, in most instances incorporation of incorrect wobble pairs has a low efficiency due to an inability for these pairs to fit a steric requirement enforced by the polymerase [13]. In the case of F-F, the hydrophobicity of F allows two F bases to be placed into the binding pocket without any requirement to remove bound water molecules. For T-T mispairs, on the other hand, the situation is different. While T-T pairs are able to compensate for desolvation by forming a wobble pair that includes two hydrogen bonds, waters bound to the exposed hydrophilic functional groups must be accounted for. Kool asserts that these bound waters inhibit the fitting of T-T pairs into the pocket due to the steric bulk of the solvated pair. The results of these simulations show that, on average, the T-T wobble pairs do have at least one water molecule on each face of the pair, and accounting for second solvent shells only increases steric bulk.

One last geometrical issue associated with incorporation of these two pyrimidine mispairs is that T-T wobble pairs are very stable, while F-F pairs are able to adopt a very wide range of structures, as evidenced by the plots of hydrogen-bond lengths sampled during the simulations. A tight pocket that selects a Watson-Crick-like geometry will favor the more deformable F-F pair, and disfavor the tighter T-T wobble pair. Any deviation from Watson-Crick geometry affects the ability of the polymerase to form a closed DNA complex facilitating nucleotide incorporation, as is the case for dNTP incorporation [108]. Assuming that either of these mispairs forms, it becomes necessary to consider whether the pair will move through the MGR, responsible for extension of the duplex, during replication. The large body of work covering extension stresses the importance of the functional groups in the minor groove in this process. In the T-T pair, it might be hypothesized that only one carbonyl O2 atom would exist in the groove to accept a hydrogen bond, and that the location of the O2 atom would be different than would be seen in an A-T pair. Wobbling of the other pyrimidine might remove the additional O2 from the minor groove. This represents loss of one additional interaction from a correctly paired A base, a loss possibly significant enough to stall

Table 4. Number of water molecules in the first solvation shell about exposed T-T and F-F functional groups^a

Trajectory	Base Pair	O2 ^b	O2 ^b	O4 ^b	O4 ^b
mis-T	T3:T10	0.73	0.33	0.40	0.85
	tandem-T	T3:T10	0.11	0.62	0.82
mis-F	T4:T9	0.60	0.16	0.36	0.83
	F3:F10	0.18	0.15	0.11	0.18
tandem-F	F3:F10	0.10	0.22	0.16	0.04
	F4:F9	0.15	0.12	0.07	0.14

^a From integration of r.d.f. peaks for T oxygen atoms (O2 or O4) and F atoms (F2 or F4)

^b O2 and O4 columns are ordered with lower numbered pyrimidine first

the process. The hydration analysis, however, does show that hydrogen bonds can be made to each of the four carbonyl oxygens of the T-T wobble pairs. Depicted in Fig. 13 are the average structures of an A-T and T-T pair from **mis-T**. The arrows point to functional groups able to accept hydrogen bonds from the MGR. The T-T wobble pair has a significant similarity to the canonical minor-groove contact orientation: the O2 atoms of the T:T pair overlap with the T O2 and A N3 of the A-T pair. Any stall induced by a T-T pair may be less pronounced because of this overlap.

Inability to hydrogen bond to F2 atoms in the F-F pair, supported by extensive quantum mechanical studies, molecular dynamics studies, and the analysis presented here (bond lengths, hydration), argues for the unlikelihood of F forming any favorable interaction with the MGR. Therefore, the stall observed when two A-F pairs are replicated in succession is similarly likely to occur when two F-F pairs are tandem to each other, and indeed even a single F-F pair might be expected to induce stalling.

One other factor which not only influences binding of the incoming dNTP but the overall stability of any double helix is base stacking. While the T-T and F-F mispairs, and tandem mispairs, adopt strikingly similar stacking arrangements within the helix, the F stacks are somewhat more dynamic. This structural freedom is imparted to the global helix structure in that the configurational entropy for **mis-F** and **tandem-F** is higher than **mis-T** and **tandem-T**. Even though F-F has no hydrogen bonding as a driving force for maintenance of a wobble configuration, as does T-T, it appears that on average, the favorable stacking of the nonpolar base F artificially induces a wobble configuration. When T-T or F-F mispairs were tested for thermodynamic stability in the center of a DNA helix, both pairs were observed to lower the melting temperature to a similar extent relative to A-T occupying the same position. Even though F stacks better as a dangling base than T, when F pairs within a duplex, stacking advantages of F versus T are less pronounced [30]. The simulations presented here suggest that the hydrogen bonding interaction that is lost in going from T-T to F-F may be balanced in part

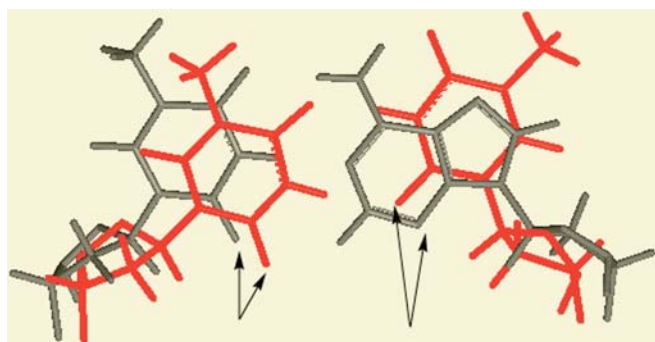


Fig. 13. Overlay of T-T (red) wobble pair and canonical A-T (black) Watson-Crick pair from time-averaged structure of **mis-T** showing relative positioning of minor groove contacts

by increased duplex stability associated with greater conformational flexibility (and so greater configurational entropy) associated with F-F pairs. Along with recently published molecular dynamics studies of DNA-polymerase complexes, the simulations presented here offer an interesting insight into the subtle differences between canonical and analog base interactions throughout the replication process [31, 109, 110].

One goal of computational chemistry is to help to direct the focus of future studies. In this spirit, having explored the structural and dynamic aspects of T-T and F-F and their implications for the possible creation of such base pairs in the polymerase active site, we consider modifications of these mispairs that might be more amenable to extension. We note that it has already been shown that modification of the non-polar base Z to Q, by addition of an N3 atom, decreases the loss in extension efficiency of a Z-F pair. Therefore, the two bases proposed in Fig. 14 have been designed to contain O2 hydrogen bonding groups at either the 2 or 4 position. Minor groove interactions should be regained when A pairs with 4-fluoro-5-methyl-2-pyrimidinone (F-O2), showing an increase in extension efficiency over A-F pairs. The base 6-fluoro-3-methyl-4-pyrimidinone (F-O4), when paired with A, should have a higher insertion fidelity as compared to F due to better hydrogen bonding. Both of the proposed bases would necessarily have varied base stacking interactions. At the *mPWPW91/MIDI!* level, the interaction enthalpies [99] (corrected for BSSE) for the fully optimized A-T A-F, A-F-O2, and A-F-O4 base pairs are calculated to be -11.3 , -0.1 , 0.7 , and -7.1 kcal mol $^{-1}$, respectively. There is clearly an increase in intermolecular attraction between A and F when the F4 atom is replaced with O4. Substitution at O2 does not alter the hydrogen bonding characteristics, as the difference between -0.1 and 0.7 kcal mol $^{-1}$ is insignificant. It would be expected that increased minor groove interactions to F-O2 would be readily apparent, either from further MD simulations, or extension experiments.

Supporting information available

Warm-up protocol, COM plots of base stacking, radial distribution functions to 15.0 Å, time plots of key helical

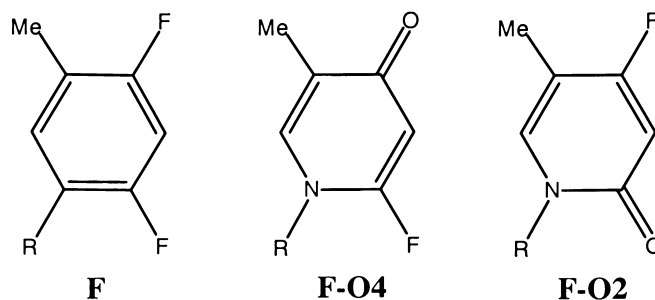


Fig. 14. Structures of F, F-O4, and F-O2

parameters, block averaged helical parameters, data for control trajectory **AT**.

Acknowledgements. We thank Sarah Harris, Charlie Laughton, Maria Nagan, and George Shields for stimulating discussions.

References

1. Doublet S, Tabor S, Long AM, Richardson CC, Ellenberger T (1998) *Nature* 391:251–258
2. Beard WA, Wilson SH (1998) *Chem Biol* 5:R7–R13
3. Kiefer JR, Mao C, Braman JC, Beese LS (1998) *Nature* 391:304–307
4. Eom SH, Wang J, Steitz TA (1996) *Nature* 382:278–281
5. Steitz TA (1999) *J Biol Chem* 274:17395–17398
6. Kunkel TA, Wilson SH (1998) *Nat Struct Biol* 5:95–99
7. Watson J, Crick HC (1953) *Nature* 171:737–738
8. Kuchta RD, Mizrahi V, Benkovic PA, Johnson KA, Benkovic SJ (1987) *Biochemistry* 26:8410–8417
9. Sloane DL, Goodman MF (1988) *Nucleic Acids Res* 16:6465–6475
10. Echols H, Goodman MF (1991) *Annu Rev Biochem* 60:477–511
11. Diederichsen U (1998) *Angew Chem Int Ed Engl* 37:1655–1657
12. Goodman MF (1997) *Proc Natl Acad Sci USA* 94:10493–10495
13. Kool E (1998) *Biopolymers* 48:3–17
14. Kool ET (2001) *Annu Rev Biophys Biomolec Struct* 30:1–22
15. Barsky D, Kool ET, Colvin ME (1999) *J Biomol Struct Dyn* 16:1119–1134
16. Moran S, Ren RX-F, Kool ET (1997) *Proc Natl Acad Sci USA* 94:10506–10511
17. Liu D, Moran S, Kool ET (1997) *Chem Biol* 4:919–926
18. Morales JC, Kool ET (2000) *J Am Chem Soc* 122:1001–1007
19. Morales JC, Kool ET (1998) *Nat Struct Biol* 5:950–954
20. Guckian KM, Kool ET (1997) *Angew Chem Int Ed Engl* 36:2825–2828
21. Guckian KM, Krugh TR, Kool ET (1998) *Nat Struct Biol* 5:954–959
22. Guckian KM, Krugh TR, Kool ET (2000) *J Am Chem Soc* 122:6841–6847
23. Guckian K M, Morales J C, Kool ET (1998) *J Org Chem* 63:9652–9656
24. Matray TJ, Kool ET (1998) *J Am Chem Soc* 120:6191–6192
25. Sherer EC, Bono SJ, Shields GC (2001) *J Phys Chem B* 105:8445–8451
26. Wang X, Houk KN (1998) *Chem Commun* 2631–2632
27. Meyer M, Suhnel J (1997) *J Biomol Struct Dyn* 15:619–624
28. Florian J, Goodman BF, Warshel A (2000) *J Phys Chem B* 104:10092–10099
29. Schmidt KS, Sigel RK, Filippov DV, van der Marel GA, Lippert B, Reedijk J (2000) *New J Chem* 24:195–197
30. Morales JC, Kool ET (2000) *Biochemistry* 39:2626–2632
31. Cubero E, Sherer EC, Luque FJ, Orozco M, Laughton CA (1999) *J Am Chem Soc* 121:8653–8654
32. Wattis JA, Harris SA, Grindon CR, Laughton CA (2001) *Phys Rev E* 63:061903
33. Patel PH, Loeb LA (2001) *Nat Struct Biol* 8:656–659
34. Johnson KA (1993) *Annu Rev Biochem* 62:685–713
35. Seeman NC, Rosenberg JM, Rich A (1976) *Proc Natl Acad Sci USA* 73:804–808
36. Kennard O (1988) In: Sarma RH, Sarma MH (eds) *Structure and expression. Adenine*, NY, Vol 2 pp 1–25
37. Pelletier H, Sawaya MR, Kumar A, Wilson SH, Kraut J (1994) *Science* 264:1891–1903
38. Osheroff WP, Beard WA, Wilson SH, Kunkel TA (1999) *J Biol Chem* 274:20749–20752
39. Spratt TE (2001) *Biochemistry* 40:2647–2652
40. Summerer D, Marx A (2002) *J Am Chem Soc* 124:910–911
41. Horlacher J, Hottiger M, Podust VN, Hübscher U (1995) *Proc Natl Acad Sci USA* 92:6329–6333
42. Guo M-J, Hildbrand S, Leumann CJ, McLaughlin LW, Waring MJ (1998) *Nucleic Acids Res* 26:1863–1869
43. Kool ET (2000) *Cold Spring Harbor Symp Quant Biol* 65:93–102
44. Morales JC, Koll ET (1999) *J Am Chem Soc* 121:2323–2324
45. Morales JC, Kool ET (2000) *Biochemistry* 39:12979–12988
46. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz K, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) *J Am Chem Soc* 117:5179–5197
47. Cubero E, Laughton CA, Luque FJ, Orozco M (2000) *J Am Chem Soc* 122:6891–6899
48. Cheatham III TE, Miller JL, Fox T, Darden TA, Kollman PA (1995) *J Am Chem Soc* 117:4193–4194
49. Cheatham III TE, Kollman PA (1996) *J Mol Biol* 259:434–444
50. Cheatham III TE, Kollman PA (1997) *J Am Chem Soc* 119:4805–4825
51. Cheatham III TE, Crowley MF, Fox T, Kollman PA (1997) *Proc Natl Acad Sci USA* 94:9626–9630
52. Cheatham III TE, Kollman PA (2000) *Annu Rev Phys Chem* 51:435–471
53. Norberg J, Nilsson L (2002) *Acc Chem Res* 35:465–472
54. Beveridge DL, McConnell KJ (2000) *Curr Opin Struct Biol* 10:182–196
55. Giudice E, Lavery R (2002) *Acc Chem Res* 35:350–357
56. Auffinger P, Westhof E (1998) *Curr Opin Struct Biol* 8:227–236
57. Nagan MC, Kerimo SS, Musier-Forsyth K, Cramer CJ (1999) *J Am Chem Soc* 121:7310–7317
58. Sherer EC, Harris SA, Soliva R, Orozco M, Laughton CA (1999) *J Am Chem Soc* 121:5981–5991
59. Sherer EC, Cramer CJ (2002) *J Phys Chem B* 106:5075–5085
60. Schneider C, Brandl M, Suhnel J (2001) *J Mol Biol* 305:659–667
61. Beuning PJ, Nagan MC, Cramer CJ, Musier-Forsyth K, Gelpi J-L, Bashford D (2002) *RNA* 8:659–670
62. Shields GC, Laughton CA, Orozco M (1998) *J Am Chem Soc* 120:5895–5904
63. Soliva R, Sherer E, Luque FJ, Laughton CA, Orozco M (2000) *J Am Chem Soc* 122:5997–6008
64. Nagan MC, Beuning P, Musier-Forsyth K, Cramer CJ (2000) *Nucleic Acids Res* 28:2527–2534
65. Kool ET, Morales JC, Guckian KM (2000) *Angew Chem Int Ed Engl* 39:990–1009
66. Hobza P, Kabelac M, Sponer J, Mejzlik P, Vondrasek J (1997) *J Comput Chem* 18:1136–1150
67. Case DA, Pearlman DA, Caldwell JW, Cheatham III TE, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Ferguson DM, Radmer RJ, Seibel GL, Singh UC, PK W, Kollman PA (1997) *AMBER 5*. University of California, San Francisco, CA
68. Shields GC, Laughton CA, Orozco M (1997) *J Am Chem Soc* 119:7463–7469
69. Darden TA, York D, Pedersen L (1993) *J Chem Phys* 98:10089–10092
70. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) *J Chem Phys* 103:8577–8593
71. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) *J Chem Phys* 81:3684–3690
72. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) *J Comput Phys* 23:327–341
73. Lavery R, Sklenar H (1989) *J Biomol Struct Dyn* 6:655–667
74. Lavery R, Sklenar H (1988) *J Biomol Struct Dyn* 6:63–91
75. Lavery R, Sklenar H (1998) *Curves*, 53rd edn. Institut de Biologie Physico-Chimique, Paris
76. Wlodek ST, Clark TW, Scott LR, McCammon JA (1997) *J Am Chem Soc* 119:9512–9522
77. Caves LSD, Evansck JD, Karplus M (1998) *Protein Sci* 7:649–666
78. de Groot BL, van Aalten DMF, Amadei A, Berendsen HJC (1996) *Biophys J* 71:1707–1713
79. Balsara MA, Wriggers W, Oono Y, Shulten K (1996) *J Phys Chem* 100:2567–2572

80. van Aalten DMF, de Groot BL, Findlay JBC, Berendsen HJC, Amadei A (1997) *J Comput Chem* 18:169–181
81. Kitao A, Hayward S, Go N (1998) *Proteins* 33:496–517
82. de Groot BL, Daura X, Mark AE, Grubmüller H (2001) *J Mol Biol* 309:299–313
83. García AE, Blumenfeld R, Hummer G, Krumhansl JA (1997) *Physica D* 107:225–239
84. Bostock-Smith C, Harris SA, Laughton CA, Searle MS (2001) *Nucleic Acids Res* 29:693–702
85. Tsui V, Case DA (2000) *J Am Chem Soc* 122:2489–2498
86. Zacharias M (2000) *Biopolymers* 54:547–560
87. Schlitter J (1993) *Chem Phys Lett* 215:617–621
88. Harris SA, Gavathiotis E, Searle MS, Orozco M, Laughton CA (2001) *J Am Chem Soc* 123:12658–12663
89. Gervais V, Cognet JAH, Le Bret M, Sowers LC, Fazakerley GV (1995) *Eur J Biochem* 228:279–290
90. Kouchakdjian M, Li BFL, Swann PF, Patel DJ (1988) *J Mol Biol* 202:139–155
91. Lescrinier E, Esnouf R M, Schraml J, Busson R, Herdewijn P (2000) *Helv Chim Acta* 83:1291–1310
92. Baeyens K, De Bondt H, Holbrook S *Nature Struct Biol* (1995) 2:56–62
93. Lietzke SE, Barnes CL, Berglund JA, Kundrot CE (1996) *Structure* 4:917–930
94. Deng J, Sundaralingam M (2000) *Nucleic Acids Res* 28:4376–4381
95. McDowell JA, He L, Chen X, Turner DH (1997) *Biochemistry* 36:8030–8038
96. Guckian KM, Schweitzer BA, Ren RX-F, Sheils CJ, Paris PL, Tahmassebi DC, Kool ET (1996) *J Am Chem Soc* 118:8182–8183
97. Yanson IK, Teplitsky AB, Sukhodub LF (1979) *Biopolymers* 18:1149–1170
98. Sukhodub LF, Yanson IK (1976) *Nature* 264:245–247
99. Sherer EC, York DM, Cramer CJ (2003) *J Comput Chem* 24:57–67
100. Auffinger P, Westhof E (2001) *J Mol Biol* 305:1057–1072
101. Auffinger P, Westhof E (2001) *Biopolymers* 56:266–274
102. Auffinger P, Westhof E (2000) *J Mol Biol* 300:1113–1131
103. Auffinger P, Westhof E (1997) *J Mol Biol* 269:326–341
104. Schneider B, Patel K, Berman HM (1998) *Biophys J* 75:2170
105. Makarov V, Pettitt BM (2002) *Acc Chem Res* 35:376–384
106. Showalter AK, Tsai M-D (2002) *Biochemistry* 41:10571–10576
107. Saenger W (1984) *Principles of nucleic acid structure*. Springer, Berlin Heidelberg, New York
108. Dzantiev L, Alekseyev YO, Morales JC, Kool ET, Roman LJ (2001) *Biochemistry* 40:3215–3221
109. Yang L, Beard WA, Wilson SH, Roux B, Broyde S, Schlick T (2002) *J Mol Biol* 321:459–478
110. Florian J, Goodman MF, Warshel A (2002) *J Phys Chem B* 106:5739–5753