# Molecular Dynamics Study of Oligonucleotides Containing Difluorotoluene

**Elena Cubero,[†] Charles A. Laughton,[‡] F. Javier Luque,[§] and Modesto Orozco*[,†]**

*Contribution from the Departament de Bioquímica, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, Cancer Research Laboratories, School of Pharmaceutical Sciences, University of Nottingham, NG7 2RD, UK, and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda Diagonal sn, Barcelona 08028, Spain*

*Received January 10, 2000*

**Abstract:** Extended molecular dynamics (MD) and thermodynamic integration (MD-TI) calculations have been used to determine the structural and energetic changes in DNA that accompany the replacement of thymine (T) by the nonnatural isostere difluorotoluene (F). In a duplex DNA oligonucleotide, it is found that the T→F mutation leads to only small changes in the average structure, but to important alterations in flexibility, hydration, and recognition properties. The T→F mutation in the Watson−Crick or Hoogsteen position of a pyrimidine·purine·pyrimidine type DNA triplex does not lead to dramatic changes in the general structure of the triplex, but again, detailed analysis shows some alterations in flexibility, hydration, and recognition properties. MD-TI calculations on the T→F mutation in duplex DNA reproduce the experimentally determined free energy differences with good accuracy, and detailed analyses of the trajectories have enabled us to rationalize these. Finally, MD-TI simulations have been used to predict the changes in stability of a triplex due to a T→F mutation in either the Watson−Crick or Hoogsteen-binding pyrimidine strands. We predict that in either case the mutation will reduce stability, being most unfavorable in the Watson−Crick strand.

## Introduction

DNA replication by DNA polymerase is a high-fidelity process with an error rate of the order of one mismatch per $10^4$ to $10^5$ bases.[1,2] It might be thought that this fidelity has its origins in the specificity of Watson−Crick hydrogen bonding, but it has been argued that this is not the case. The free energy difference between matched and mismatched terminal base pairs in solution is estimated[3] to range between 0.2 and 0.4 kcal $mol^{-1}$. This difference is too low to account for the observed error rate, suggesting that shape-complementarity must also play an important role in ensuring the fidelity of replication.[4] However, other workers[5] have argued that within the environment of the polymerase, the free energy difference between matched and mismatched base pairs may be much greater due to reduced solvation and so the hydrogen-bonding argument is sufficient.

These competing hypotheses have been tested using the thymine mimic difluorotoluene (F, Figure 1). F was designed as a nonpolar homologue of thymine,[6] which lacks H-bonding capability, even in apolar solvents such as chloroform.[7] Despite its nonpolar nature, it was found that DNA polymerase I would incorporate F across from A, and A across from F, in a precise fashion.[7,8] Despite this specificity, thermal denaturation studies[9] show that replacing T by F destabilized DNA duplexes by 3.0−3.6 kcal $mol^{-1}$. The significance of these results was challenged by Evans and Seddon,[10] who argued on the basis of quantum mechanical calculations that F was a significantly polar molecule and could form nonclassical hydrogen bonds. Recently, this conclusion has been disputed by Wang and Houk,[11] whose quantum mechanical and molecular mechanical calculations support the view that F does not hydrogen bond.

The structure of an AF-containing DNA dodecamer has been recently determined by NMR.[12] The structure refinement involved numerous short (approximate 25 ps) molecular dynamics (MD) simulations with NMR-derived distance restraints. No unusual behavior of the dodecamer was observed during the MD simulations, and the refined structure showed standard B-type characteristics. On the theoretical side, very recent MD simulations[13] have confirmed that the T→F mutation does not induce major changes in the average structure of a duplex, but revealed important changes in its flexibility. It was possible to detect, within a 10 ns trajectory, several "breathing" movements of the AF base pair, whereas normal AT and GC base pairs only breath on the microsecond time scale.

In contrast to the wealth of information on the effect of T→F mutations on the structure and stability of duplex DNA, there

[†] Facultat de Química, Universitat de Barcelona.
[‡] University of Nottingham.
[§] Facultat de Farmàcia, Universitat de Barcelona.

(1) Kornberg, A.; Baker, T. A. *DNA Replication*, 2nd ed.; W. H. Freeman: New York.
(2) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83.
(3) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252.
(4) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.
(5) Johnson, K. A. *Annu. Rev. Biochem*. **1993**, *62*, 685.
(6) Guckian, K. M.; Kool, E. T. *Angew. Chem., Int. Ed*. **1997**, *36*, 2825.
(7) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506.
(8) Liu, D.Y.; Moran, S.; Kool, E. T. *Chem. Biol*. **1997**, *4*, 919.
(9) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc*. **1997**, *119*, 2056.
(10) Evans, T. A.; Seddon, K. R. *J. Chem. Soc., Chem. Commun*. **1997**, 2023.
(11) Wang, X.; Houk, K. N. J. *Chem. Soc., Chem. Commun*. **1998**, 2631.
(12) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nature Struct. Biol*. **1998**, *5*, 954.
(13) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653.

10.1021/ja000117n CCC: $19.00 
Published on Web 07/08/2000

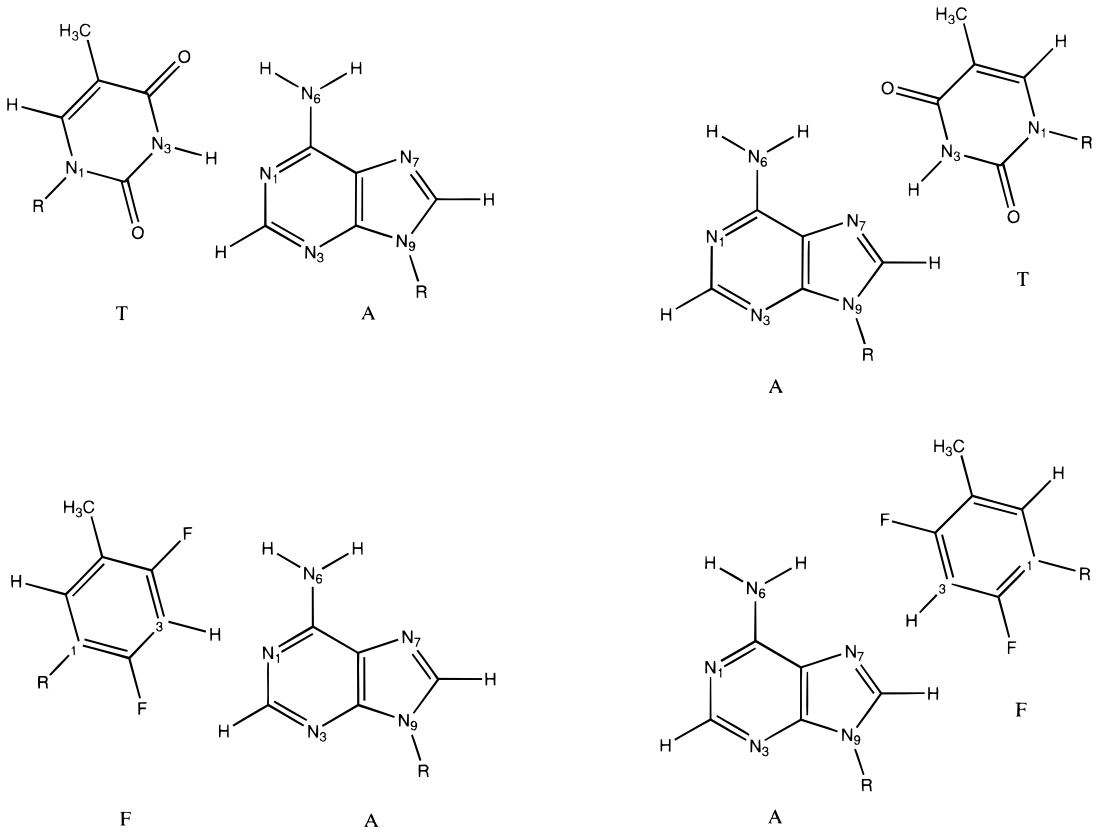


**Figure 1.** Chemical structure of 1,3-difluorotoluene (F) and thymine bound to adenine in the Watson−Crick and Hoogsteen orientations.

**Scheme 1.** DNA Duplex and Triplex Sequences Used in This Study

```
5'-AAGAAAGAAAAG-3'        5'-AAGAAAGAAAAG-3'
3'-TTCTTTTCTTTTC-5'       3'-TTCTTTFCTTTTC-5'
        (1)                       (2)

5'-AAGAAAGAAAAG-3'   5'-AAGAAAGAAAAG-3'   5'-AAGAAAGAAAAG-3'
3'-TTCTTTTCTTTTC-5'  3'-TTCTTTFCTTTTC-5'  3'-TTCTTTTCTTTTC-5'
5'-TTCTTTTCTTTTC-3'  5'-TTCTTTTCTTTTC-3'  5'-TTCTTTFCTTTTC-3'
        (3)                  (4)                  (5)
```

have been, to our knowledge, no similar studies looking at triplex DNA. In this paper we present an extension of our previous MD simulations of F-containing oligonucleotides. In addition to a 10 ns trajectory of a duplex containing a T→F mutation (and the corresponding 1.5 ns of trajectory for a control duplex containing no mutation), we have produced 3 ns trajectories for DNA triplexes containing the T→F mutation in both the Watson−Crick (d(T−A·F) motif) and the Hoogsteen (d(F−A·T) motif) positions, and a 1.5 ns control trajectory for the parent unmutated triplex. In addition, a series of MD-TI simulations have been performed to study in more detail the influence of the T→F mutation on the stability of both duplex and triplex DNA. MD-TI simulations are expected to be very useful to determine the stability of nucleic acid structures containing unnatural bases pairs,[14,15] but due to technical problems MD-TI calculations on these systems are scarce.

Calculations provide a detailed view on the changes in structure, flexibility, reactivity, and stability induced in duplex and triplex DNA by the T→F mutation. The implications of the results in the design of new antigene and antisense strategies are discussed.

## Methods

**MD Simulations.** All molecular dynamics simulations were performed using the AMBER 5.0 suite of programs[16] in association with the AMBER 95 force field. Missing AMBER parameters for difluorotoluene were those previously determined in our previous studies.[13] All simulations were performed at constant temperature (300 K) and pressure (1 atm). Long-range electrostatic interactions were handled using the particle mesh Ewald (PME)[17] method. SHAKE[18] was used to constrain all bonds, allowing the use of a 2 fs time step.

The DNA sequences studied are shown in Scheme 1. The duplex structures **1** and **2** were constructed in standard B-type conformation.[19] Triplexes **3**, **4**, and **5** were generated using our equilibrated structure

(14) Soliva, R.; Luque, F. J.; Orozco, M. *Nucleic Acid. Res.* **1999**, *27*, 2248.

(15) Hernández, B.; Soliva, R.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**. Submitted for publication.

(16) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W.S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 5, University of California: San Francisco, 1997.

(17) Cheetham. T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.

(18) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(19) Arnott, S.; Hukins, W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.

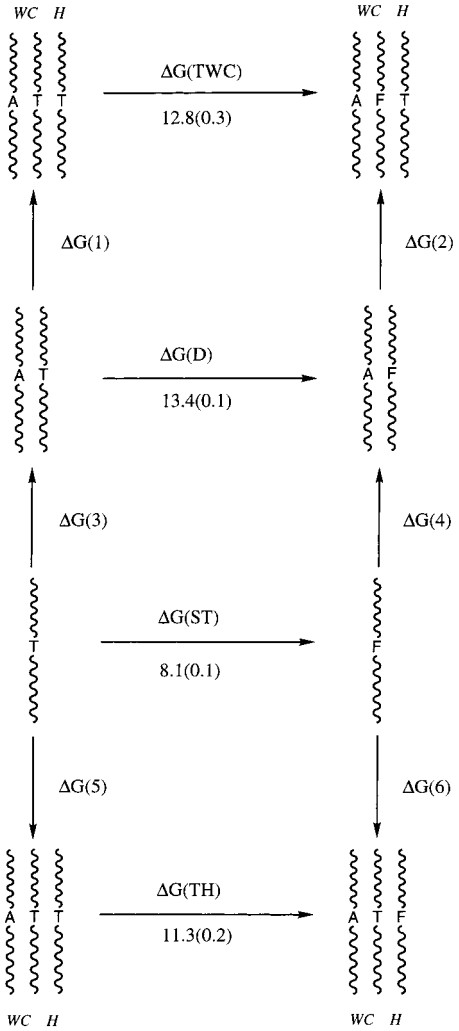


**Figure 2.** Thermodynamic cycles used in the MD/TI calculations. Average results obtained from the simulations are displayed with their standard errors (all in kcal/mol).

of the d(A·T−T) triplex.[20] Periodic boxes were constructed around the systems containing sodium counterions (to achieve neutral systems) and 2482 (duplex structures) or 3278 (triplex structures) TIP3P water molecules.[21] The systems were energy minimized, heated, and equilibrated using our standard equilibration process.[20] The equilibrated systems were subjected to 1.5 **(1, 3)**, 10 **(2)**, and 3 **(4, 5)** ns unrestrained MD simulations.

Energy analysis was done using AMBER-5.0 as well as "in-house" programs. Helical parameters were calculated using CURVES.[22] Where CURVES yields both global and local versions of parameters, the local parameter values are quoted.

**Free Energy Calculations.** MD-TI calculations and standard thermodynamic cycles (see Figure 2) were used to determine the effect of the T→F mutation on the stability of duplexes and triplexes. Starting structures for the MD-TI calculations were determined after extensive MD equilibration (see Table 1). Mutations were typically performed in both the T→F and F→T directions to verify the reversibility of the mutation pathways. Simulations were performed using 21 windows of 10/20 ps of equilibration and 10/20 ps of collection for a total of 420 or 840 ps. In all cases, free energy estimates obtained using the first half (equilibration) and second half (data collection) of each window were collected. The existence of several estimates of the free energy change associated with a particular mutation allows us to obtain a good estimate of the statistical errors in the free energy estimates.

Trajectories for MD-TI calculations were performed considering long-range contributions as introduced by the PME method, and considering all inter- and nonbonded intramolecular contributions. All the technical details of simulations used for TI calculations are identical with those used in standard MD simulations (see above).

(20) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463.

(21) Jorgensen, W. L.; Chandresekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(22) Lavery, R.; Sklenar, J. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.

**Table 1.** Summary of MD/TI Simulations[a]

| mutation | system | structure | length of MD |
|---|---|---|---|
| F→T | SS (3 steeps) | equil. for 340 ps[b] | 420 ps |
| F→T | SS (3 steeps) | equil. for 340 ps[c] | 420 ps |
| F→T | SS (3 steeps) | equil. for 340 ps | 840 ps |
| F→T | SS (3 steeps) | annealing (520 ps)[d] | 420 ps |
| F→T | SS (5 steeps) | equil. for 340 ps[b] | 420 ps |
| T→F | duplex d(A·T) | equil. for 1500 ps | 420 ps |
| T→F | duplex d(A·T) | equil. for 1500 ps | 840 ps |
| F→T | duplex d(A·F) | equil. for 1500 ps | 420 ps |
| T→F(WC) | triplex d(A·T−T) | equil. for 1000 ps | 420 ps |
| T→F(H) | triplex d(A·T−T) | equil. for 1000 ps | 420 ps |
| F→T(WC) | triplex d(A·F−T) | equil. for 1000 ps | 420 ps |
| F→T(WC) | triplex d(A·F−T) | equil. for 1000 ps | 840 ps |
| F→T (H) | triplex d(A·T−F) | equil. for 1000 ps | 420 ps |
| F→T (H) | triplex d(A·T−F) | equil. for 1000 ps | 840 ps |

[a] For more details of the simulations see text. SS means single stranded, WC refers to the Watson−Crick position, and H refers to the Hoogsteen position. [b] B-type single stranded structures were generated, surrounded by counterions and around 850 water molecules. Structures were then minimized, heated, and equilibrated for 340 ps. [c] As in footnote *b*, but a large box of 2851 water molecules (similar to those used for duplex structures) was used. [d] As in footnote *b*, but a 180 ps simulated annealing (heating at 500 K for 80 ps and cooling at 300 K for 100 ps) was used to obtain a different starting conformation.

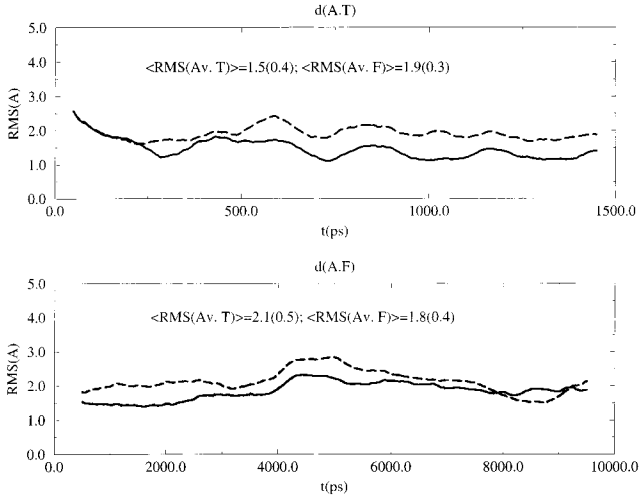


**Figure 3.** RMS deviations in the trajectories of the d(AAGAA**A**-GAAAAG) duplex containing T or F bound to the adenine in position 6. The continuous line shows the deviation with respect to the average structure obtained from the same trajectory. The dashed line shows the deviation with respect to the average structure from the other trajectory.

## Results and Discussion

**MD Simulations. (a) Duplex DNA.** MD simulations of the d(AAGAA**A**GAAAAG) DNA duplex with T or F paired to adenine at position six lead to stable trajectories in which the conformation of the duplex remains closer to B than to A DNA. Comparison of the two trajectories shows (Figure 3) that from a general structural point of view they are very similar. Thus, the average RMS deviation between the MD-averaged structure of **2**, and the trajectory of **1** is 1.9(±0.3) Å. Conversely, the

**Table 2.** Selected MD-Averaged Helical Parameters for the Central 10 Base Pairs of the Dodecamers Containing a D(A·T) and a D(A·F) Step[a]

| parameters | duplex with d(A·T) | duplex with d(A·F) |
|---|---|---|
| X-displacement (Å) | −1.7 (0.7) | −1.4(0.6) |
| inclination (deg) | −8.2(4.7) | −8.5((4.9) |
| rise (Å) | 3.5(0.2) | 3.5(0.2) |
| roll (deg) | 6.3(2.5) | 6.9(2.3) |
| twist (deg) | 30.3(2.0) | 30.0(1.5) |
| propeller twist (deg) | −11.2(5.3) | −11.1(4.5) |
| phase angle (deg) | 119(31) | 116 (35) |
| puckering amplitude (deg) | 40.9(6.0) | 40.7(6.2) |
| major groove width (Å) | 21.5(1.7) | 22.1(1.7) |
| minor groove width (Å) | 12.3(0.9) | 12.8(0.9) |

[a] Averages are done using the last 5 ns of the trajectory of the duplex with the d(A·F) step and the last 1 ns of the trajectory for the control duplex.

average RMS deviation between the MD-averaged structure of **1** and the trajectory of **2** is 2.1(±0.4) Å. These values are almost identical with those found when the MD-averaged structures of **1** and **2** are compared with their own trajectories (1.5(±0.4) Å for **1** and 1.8(±0.4) Å for **2**), demonstrating the identity between the different trajectories.

Results in Table 2 confirm the B-like nature of the two duplexes. It is also clear that the T→F mutation leads to very small changes in the helical parameters of the duplex. The grooves are also very similar in normal DNA and that containing F. All sugars show a B-type puckering with major population of the East and South regions. As typically found in simulations of DNA with AMBER-95 force field,[16] twist values are slightly underestimated with respect to a canonical B-type duplex, but in any case almost the same twist is found for both duplexes, confirming the small impact the T→F mutation has on the average structure of the duplex.

Despite the close similarity in the general structure of the two duplexes, there are detectable differences in the region of the mutation. Thus, the T→F mutation leads to small changes in the recognition characteristics of duplex DNA, as evident in the MIP [20,23] map (Figure 4). This shows that, compared to **1**, **2** has a reduced affinity for cationic probes in the minor groove. However, this altered affinity is very localized to the mutation site. MD-derived solvation maps[20] (Figure 4) also show a reduced solvation in the minor groove in the region close to the AF pair, as compared to the AT one. These differences may be rationalized by considering first the alteration in the molecular electrostatic potential (see Figure 5) due to the lack of the lone pair at position 2 in difluorotoluene and second the existence of breathing movements (see below) which lead to a partial disruption of the solvent atmosphere surrounding the mutated base pair.

Though we observe no significant difference between **1** and **2** in terms of their global structures, there is a clear difference between them in terms of their dynamics. The most important aspect of this difference relates to the existence of a breathing motion, which leads to a partial or total reversible opening of the A·F pair, in contrast to the rigid behavior of the A·T pair. The breathing events occur on a nanosecond time scale (see Figure 6 and ref 13), and have been shown to originate from the related to the reduced stability of the A·F Watson−Crick dimer compared to that of the A·T pair.

**(b) Triplex DNA.** Three triplexes have been studied, with the sequences shown in Figure 1. Triplex **3** contains the unmodified d(A·T−T) trio at position 6, while the other two contain a T→F mutation: d(A·F−T) in **4** and d(A·T−F) in **5**. In all cases MD simulations produce stable trajectories over the nanosecond time scale at room temperature and near-physiological conditions, suggesting that triplexes containing a single T→F mutation can be stable in physiological conditions.

The triplex conformations remain essentially B-type. Thus, the RMS deviations from a canonical B-type model average 1.3 Å while the RMS deviations from a canonical A-type model lie in the range 2.3−2.4 Å. In a manner similar to the duplexes, convergence of the three trajectories is evident from comparison of RMS deviations. This is shown in Table 3, where the cross RMS deviations between triplexes containing F and T are 1.0−1.1 Å. These values are similar in magnitude to the RMS deviations found between individual structures from a trajectory and the MD-averaged structure derived from it.

The T→F mutation does not have any major impact on the helical characteristics of the triplexes, which are those expected for a B-type triplex[20,24,2524,25] (see Table 4). In all the cases the twist is around 30° with a rise around 3.3−3.4 Å. No change in groove dimensions or sugar puckering (average phase angle around 110−112°) is apparent as a consequence of the T→F mutation either in the Watson−Crick or Hoogsteen positions.

The general recognition and hydration characteristics of the triplex remain unaltered upon the T→F mutation in the Watson−Crick position. However, such a mutation weakens the region of negative potential in the minor groove, and a partial disruption of the spine of hydration (see Figure 7). There is also a slight decrease in the apparent density of water in the minor part of the major groove (for nomenclature see ref 20). As for the duplexes (see above), these observations can be accounted for by considering two factors: first the loss of the lone pair at position 2 of the pyrimidine, resulting from the T→F mutation (see Figure 5), and second the breathing motions in the A·F pair (see below).

The T→F mutation in the Hoogsteen position does not lead to any important change in the electronegativity in the grooves or in their apparent water density (see Figure 7). The lack of difference in the minor part of the major groove is particularly interesting, since it should experience a partial loss in H-bonding capabilities as a result of the T→F mutation. Comparing with the results for triplex **4**, this suggests that the integrity of the spines of hydration is more sensitive to the existence of "breathing" and "opening" movements (see ref 13 for nomenclature) than to the partial loss of H-bonding capability.

Breathing movements are detected in triplexes containing F (**4** and **5**), while they are not found in the reference triplex **3**. The triplex with the T→F mutation in the Watson−Crick position (**4**) shows reversible partial opening and breathing movements (see ref 13 for nomenclature), which lead to partial or total loss of WC hydrogen bonds due to the displacement of F to the major groove. Analysis of the trajectory (see Supporting Information for details) suggests free energy changes of 0.9−1.3 kcal/mol for partial opening, and 2.0−2.4 kcal/mol for breathing. These values are only around 0.5 kcal/mol higher than those found for a duplex DNA with a single T→F mutation,[13] which suggests that the "breathing" of Watson−Crick base pairs is not dramatically modified by the presence of a third strand.

The triplex with the T→F mutation in the Hoogsteen position (**5**) shows reversible breathing movements, where the F is displaced to the major groove, leading to the partial or total

(23) Shields, G.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895.

(24) Ragunnathan, G.; Miles, H. T.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 455.

(25) Howard, F. N.; Miles, H. T.; Liu, K.; Frazier, J.; Raghunathan, G.; Sasisekharan, V. *Biochemistry* **1993**, *31*, 10671.

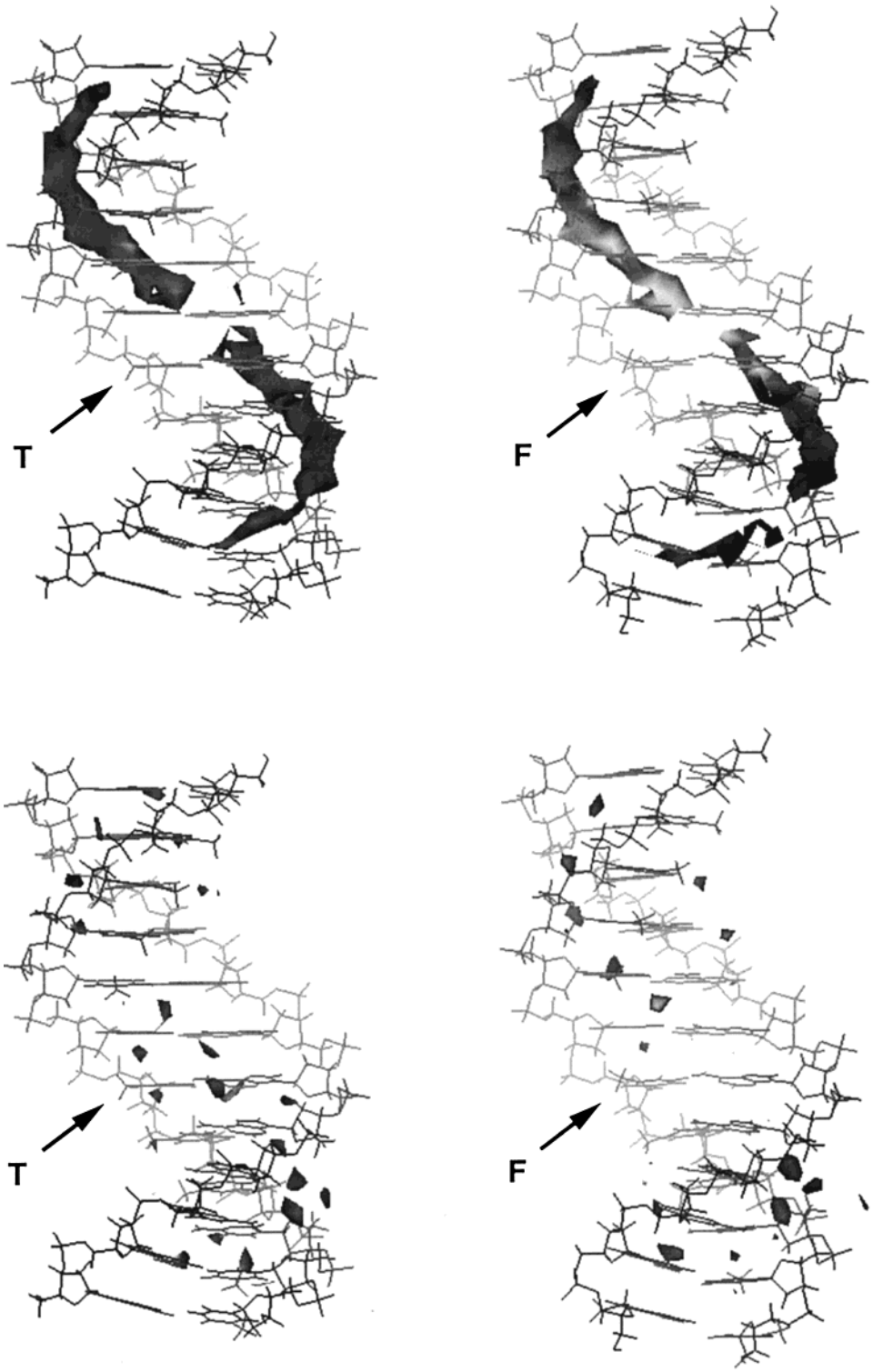


**Figure 4.** Molecular interaction potential (MIP) and solvation maps obtained for the duplexes containing T or F. The MIP map is contoured at −5.0 kcal/mol, and the solvation map is contoured at a water density of around 2.5 g/cm$^3$.

loss of Hoogsteen hydrogen bonds. The free energies for partial and total opening are estimated to be around 1.9−2.6 and 3.1−4.1 kcal/mol, respectively. These values are 1−2 kcal/mol larger than those found for Watson−Crick breathing of A·F pairs. This strongly suggests that Hoogsteen breathing (at least for A·F pairs) is a less common event than Watson−Crick breathing, which might suggest a greater stability for the d(A·T−F) trio compared to the d(A·F−T) trio.

**Free Energy Calculations.** Figure 8 shows the free energy profiles associated with the T↔F mutation in single-stranded oligonucleotides, duplex DNA, and triplex DNA, obtained considering different starting conformations, both the T→F and F→T directions, and using trajectories of different lengths (see Table 1).[26] All the free energy profiles (Figure 8) are smooth, without apparent discontinuities which could signal the existence of hysteresis. Moreover, the different free energy changes associated with a given mutation, which are determined from

(26) In all cases the "gas phase" profile is subtracted to produce a common reference state for all the simulations.

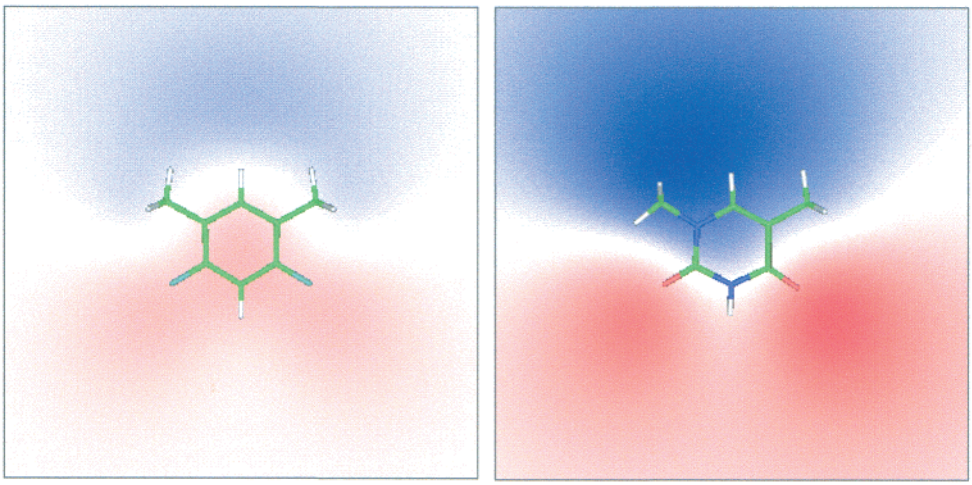

**Figure 5.** Ab initio HF/6-31G(d) molecular electrostatic potential map computed at 3.4 Å above the aromatic plane. Colored areas correspond to MEP values from −10 (deep red) to +10 (deep blue) kcal/mol.

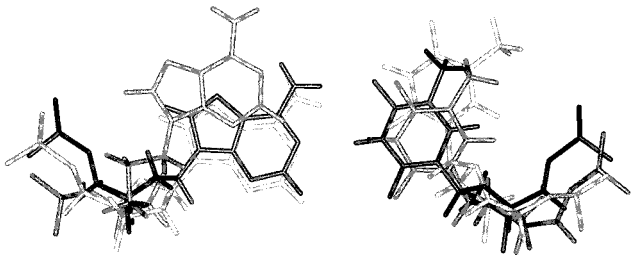

**Figure 6.** Structure of three snapshots showing normal A·F pair (black) and pairs with breathing of A (dark gray) and breathing of F (light gray).

**Table 3.** RMS Deviations (in Å) between Trajectories (rows) and MD-Averaged Structures (columns)[a]

| | d(A·T−T) | d(A·F−T) | d(A·T−F) |
|---|---|---|---|
| d(A·T−T) | 0.97(0.16) | 1.05(0.18) | 1.08(0.22) |
| d(A·F−T) | 1.11(0.21) | 1.10(0.21) | 1.11(0.19) |
| d(A·T−F) | 1.03(0.18) | 1.07(0.17) | 1.01(0.18) |

[a] Standard deviations in the averages are displayed in parentheses.

**Table 4.** Selected MD-Averaged Helical Parameters for the Central 10 Base Pairs of the Triplexes Containing a d(A·T−T), a d(A·F−T), and a d(A·T−F) Step[a]

| parameters | triplex d(A·T−T) | triplex d(A·F−T) | triplex d(A·T−F) |
|---|---|---|---|
| X-displacement (Å) | −3.0(0.5) | −3.0(0.6) | −3.1(0.5) |
| inclination (deg) | −3.2(4.7) | −1.6(5.0) | −1.4(4.3) |
| rise (Å) | 3.4(0.1) | 3.4(0.1) | 3.4(0.1) |
| roll (deg) | 3.2(1.8) | 3.6(1.8) | 3.1(1.8) |
| twist (deg) | 29.6(0.7) | 29.5(0.6) | 29.7(0.7) |
| propeller twist (deg) | −12.2(4.7) | −12.8(4.7) | −13.5(4.8) |
| phase angle (deg) | 111(29) | 110(38) | 113(26) |
| puckering amplitude (deg) | 41.5(5,8) | 41.5(5.8) | 41.7(5.7) |
| MM groove width (Å) | 15.6(1.2) | 15.1(1.0) | 15.3(0.8) |
| mM groove width (Å) | 9.0(0.3) | 9.0(0.3) | 9.0(0.3) |
| m groove width (Å) | 11.5(0.5) | 12.1(0.5) | 11.5(0.5) |

[a] Averages are done using the last 1.5 (reference triplex) and 2.0 (triplexes with mutation) ns of the trajectories. MM means major-major groove, mM means minor-major groove, and m refers to the minor groove. See ref 20 for nomenclature.

different simulations, are in very close agreement (below 0.3 kcal/mol), giving us confidence in the statistical quality of these estimates.

Free energy differences around −8.1 kcal/mol are found for mutations in single-stranded DNA (irrespective of the size of the oligonucleotide) favoring hydration of T in front of F (see Figures 2 and 8). This value agrees well with SCRF estimates (−8.6 kcal/mol) obtained at the AM1/MST level,[27] which gives us further confidence in the quality of the MD/TI calculations.

Simple algebra with free energy differences in Figure 2 gives us all the thermodynamic data associated with the effect of T→F mutations on duplex and triplex stability. Results in Table 5 show that the T→F mutation destabilizes the duplex DNA by about 5 kcal/mol, in good agreement with experimental values found by Kool and co-workers (values around 4 kcal/mol from refs 6, 7, and 9). Analysis of the central d(GpApA) sequence of the MD-averaged structures (see Table 6) shows that the largest difference between T and F in terms of duplex stability lies in the H-bonding term, which is near 10 kcal/mol more stable for duplexes containing the d(A·T) pair.[28]

In summary, MD and MD/TI simulations strongly suggest that the presence of a single d(A·F) step destabilizes the duplex, but the resulting DNA is still stable at room temperature, in good agreement with all known experimental evidence.[6,7,9] In practice, the largest differences between a DNA duplex containing a d(A·F) pair and the natural one should be related to (i) the change in the intrinsic reactivity in the minor groove, which might modify the ability of DNA to interact with small molecules or some DNA minor-groove binding proteins, and (ii) the existence of breathing events which might also affect the reactivity of DNA, and contribute to the activation of DNA-repairing systems.

The formation of most parallel-motif triplexes is a two-step process. First a DNA duplex is formed, and second the triplex-forming oligonucleotide is bound[29] to give the triplex. The transition from a duplex with a d(A·T) pair to a triplex with a d(A·T−T) trio is around 3.2 kcal/mol more favorable than the transition from the same duplex to a triplex with a d(A·T−F) trio (see Table 5). Interestingly, if the duplex contains a d(A·F) pair, the formation of a d(A·F−T)-containing triplex is not disfavored with respect to the formation of a triplex with a d(A·T−T) trio from a duplex with the d(A·T) pair (differential free

(27) (a) Luque, F. J.; Negre, M.; Orozco, M. *J. Phys. Chem.* **1993**, *97*, 4386. (b) Luque, F. J.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 847. (c) Orozco, M.; Bachs, M.; Luque, F. J. *J. Comput. Chem.* **1995**, *16*, 564−575. (d) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117. (e) Miertus, J.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.

(28) AM1/MST calculations show that hydration disfavors A·T and A·F pairings by around 8.1 and 3.6 kcal/mol. This implies that solvation favors relative binding of F to A, with respect to the binding of T to A, by around 4.5 kcal/mol, thus reducing the apparent large difference in stability suggested from H-bonding considerations alone.

(29) Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601.

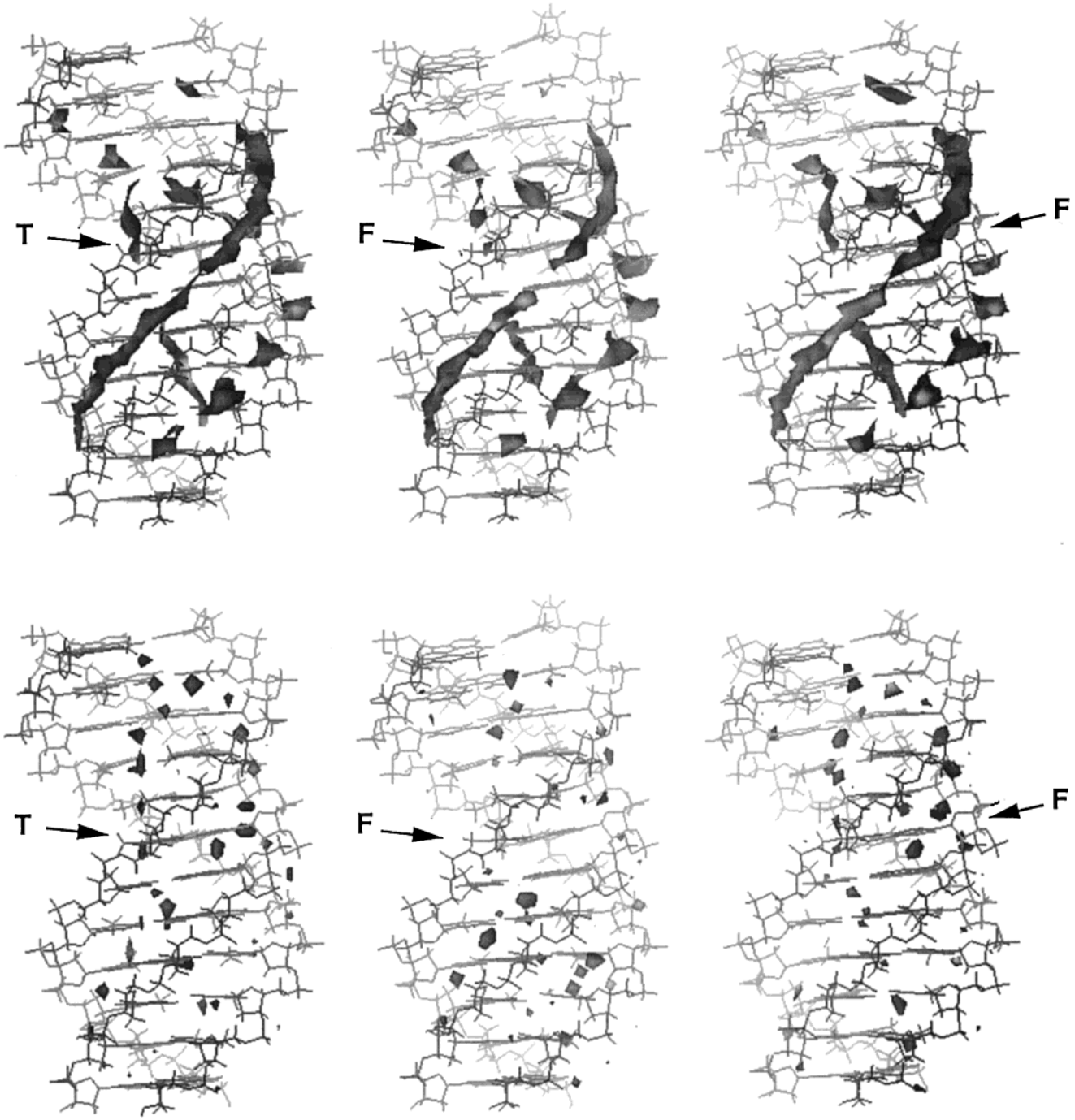


**Figure 7.** Molecular interaction potential (MIP) and solvation maps obtained for the triplexes containing T or F. The MIP maps are contoured at −6.0 kcal/mol, and the solvation maps are contoured at a water density of around 3.5 g/cm$^3$

energy −0.6 kcal/mol). This finding can be explained considering that the characteristics of the major groove of the duplex containing the d(A·F) pair are very similar to those of the reference duplex (see above and Table 2). The similarity of triplex-forming characteristics of the parent and T→F mutated duplex strongly suggests that a duplex DNA containing a d(A·F) pair would have, a priori, a similar ability to interact with most major groove-binding DNA proteins as the "parent" duplex.

Perhaps more interesting is the study of the effect of the T→F mutation on the thermodynamics of the whole assembly process, from single-stranded species to triplex. As seen in Figure 8 and Table 5, the presence of one F destabilizes the resultant triplex with respect to the reference triplex (d(A·T·T)) by around 3.2 kcal/mol when F is in the Hoogsteen position, and 4.7 kcal/mol when it is in the Watson−Crick position. Therefore, a triplex containing a d(A·T−F) trio will be around 1.5 kcal/mol more stable than that containing a d(A·F−T) trio. Energy analysis of the MD-averaged structures suggests (Table 7) that the largest difference in stability due to the presence of T instead of F lies in the H-bonding term. The T→F mutation in the Watson−Crick position adversely affects H-bonding to a greater extent than the same mutation in the Hoogsteen position, which may help to explain the higher stability of the d(A·T−F) triplex compared to the d(A·F−T) triplex found in our MD/TI calculations. The better hydrogen bonding of the d(A·T−F) triplex (unexpected from optimizations in the gas phase[30]) can be mostly attributed to backbone effects which make possible a better interaction geometry for the Hoogsteen A−F than for the Watson−Crick A·F pairing.

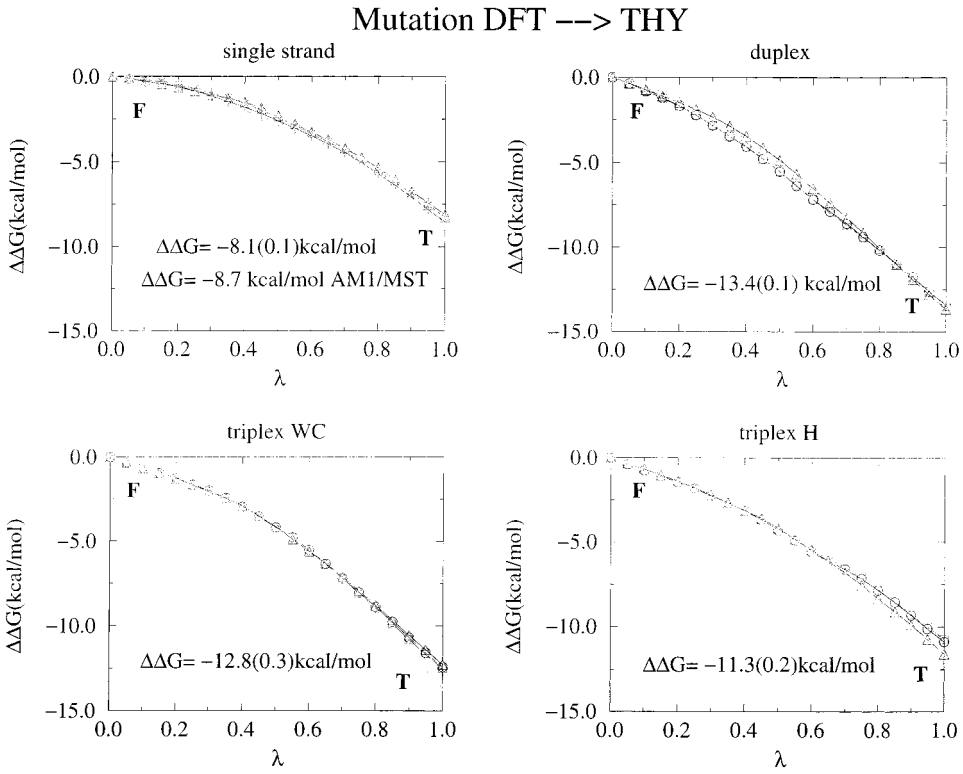


**Figure 8.** Free energy profiles obtained from different trajectories. Average values (in kcal/mol) are shown with their standard errors. See also Figure 2.

**Table 5.** Changes in the Free Energy of the Duplex→Triplex, Single→Duplex, and Single→Triplex Process Associated with the T→F Mutation[a]

| folding process | sequence | ΔΔ*G* (kcal/mol) | SE (kcal/mol) |
|---|---|---|---|
| duplex→ triplex | d(T)+d(A·T)→ d(A·T−T) | 0.0 | 0.0 |
| duplex→ triplex | d(T)+d(A·F)→ d(A·F−T) | −0.6 | 0.3 |
| duplex→ triplex | d(F)+d(A·T)→ d(A·T−F) | 3.2 | 0.2 |
| single→ duplex | d(T)+d(A)→d(A·T) | 0.0 | 0.0 |
| single→ duplex | d(F)+d(A)→d(A·F) | 5.3 | 0.1 |
| single→ triplex | d(T)+d(A)+d(T)→ d(A·T−T) | 0.0 | 0.0 |
| single→ triplex | d(F)+d(A)+d(T)→d(A·F−T) | 4.7 | 0.3 |
| single→ triplex | d(T)+d(A)+d(F)→d(A·T−F) | 3.2 | 0.2 |

[a] Standard errors for the different estimates are displayed. See Figure 8 and text for details.

**Table 6.** Interaction Terms (in kcal/mol) for the Central Three Steeps (d(Gp**A**pA)) of the Duplexes Containing the d(A·T) and d(A·F) Pair[a]

| interaction | d(A·T) | d(A·F) |
|---|---|---|
| stacking (intrastrand) | −27.6 | −24.8 |
| stacking (interstrand) | −4.6 | −6.2 |
| stacking (total) | −32.3 | −31.0 |
| hydrogen bonding | −54.5 | −44.2 |
| total interaction | −86.8 | −75.2 |

[a] Calculations are done using the MD-averaged structure.

MD-TI calculations strongly suggest that the T→F replacement in Watson−Crick or Hoogsteen positions destabilizes the triplex DNAs by 3−5 kcal/mol with respect to the parent triplex **3**. With use of the empirical approach of Roberts and Crothers,[31]

(30) B3LYP/6-31G(d, p) geometry optimization suggests stabilization energies (after BSSE correction) of −3.1, and −2.9 kcal/mol for Hoogsteen and Watson−Crick A·F pairs. These values are remarkably similar to those obtained using AMBER parameters: −3.3 (A−F, H) and −3.2 (A−F, WC) kcal/mol.

(31) Roberts, R. W.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4320.

**Table 7.** Interaction Terms (in kcal/mol) for the Central Three Steps (d(Gp**A**pA)) of the Triplexes Containing the d(A·T−T), d(A·F−T), and d(A·T−F) Trios[a]

| interaction | d(A·T−T) | d(A·F−T) | d(A·T−F) |
|---|---|---|---|
| stacking (intrastrand) | −40.1 | −35.9 | −38.7 |
| stacking (interstrand) | −20.6 | −20.1 | −17.5 |
| stacking (total) | −60.7 | −56.0 | −56.2 |
| hydrogen bonding | −115.0 | −96.5 | −104.9 |
| total interaction | −175.7 | −152.5 | −161.1 |

[a] Calculations are done using the MD-averaged structure.

the standard free energy for triplex formation at 37 °C and pH 5.0 (we model all third strand cytosines as fully protonated) is −8.85 kcal/mol. On this basis, we would expect both triplexes **4** and **5** to be stable at room temperature, as suggested by 3 ns MD trajectories (see above). Experimental work should be done to verify the stability of these triplexes.

## Conclusions

The replacement of a single thymine base by a difluorotoluene mimic clearly destabilizes the structures of duplex and triplex DNAs. However, at least for the sequences considered here, all structures are stable in MD simulations at room temperature for at least several nanoseconds. These findings agree with previous experimental data for F-containing DNA duplexes, but are a prediction for F-containing triplexes, for which no experimental data exist. These theoretical suggestions are expected to encourage experimental effort to confirm the stability of F-containing triplexes.

The T→F mutation does not lead to major modifications in the structure or recognition properties of either duplex or triplex DNA, but leads to important changes in dynamics, particularly the emergence of "breathing" events on the nanosecond time scale. These breathing movements appear for all F-modified

structures, but are more common when the T→F mutation occurs in the Watson−Crick position than when it occurs in the Hoogsteen position. Breathing movements might have important implications to modulate the interactions of DNAs containing difluorotoluene with proteins and drugs, and might be involved in the activation of damage signals for DNA-repairing systems.

State of the art MD/TI simulations provide robust estimates for the change in duplex/triplex stability produced by the T→F mutation which agree well with available experimental evidence. The calculations suggest that the T→F mutation is more destabilizing in the Watson−Crick than in the Hoogsteen position. This suggests that it is a reasonable strategy to consider the development of nonstandard and apolar bases to be used in triplex-forming oligonucleotides for "antigene" strategies, where the lack of H-bonding ability might represent a clear advantage to design stable triplexes in nonhomopurine sequences.

**Acknowledgment.** We are indebted to E. Sherer, Dr. G. Shields, J. C. Morales, and Dr. E. T. Kool for valuable suggestions and acknowledge the financial support of the Spanish DGICYT (PB96-1005), the Catalonian Supercomputer Center (CESCA), and the Wellcome Trust. This is a contribution from the Centre Especial de Recerca en Química Teórica.

**Supporting Information Available:** Figure showing the variation of the H-bond distances for triplexes **3**, **4**, and **5** (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA000117N