Observation of Spontaneous Base Pair Breathing Events in the Molecular Dynamics Simulation of a Difluorotoluene-Containing DNA Oligonucleotide

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

Departament de Bioquímica i Biología Molecular 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 U.K. Departament de Fisicoquímica Facultat de Farmacia, Universitat de Barcelona Avgda. Diagonal sn, Barcelona 08028, Spain

Received April 5, 1999

We report the results of extended molecular dynamics simulations of a DNA oligonucleotide containing an adenine-difluorotoluene (A·F) base pair, and of the corresponding "parent" oligonucleotide containing an adenine-thymine base pair. The observation in the former case of spontaneous breathing events involving the A·F base pair gives further insight into the controversial subject of the thymine-mimicking characteristics of difluorotoluene.

Difluorotoluene (F, Figure 1) has been designed as a nonpolar homologue of thymine and investigated extensively in an attempt to understand the origins of fidelity in DNA replication.[1−4] Studies in chloroform show no evidence that F forms hydrogen-bonding interactions with adenine,[5] but DNA polymerase I will incorporate F across from A, and A across from F, in a precise fashion.[5,6] Despite this specificity, thermal denaturation studies[7] show that replacing T by F destabilizes DNA duplexes by 3.0−3.6 kcal mol^{-1}. The significance of these results has been debated in quantum mechanical (QM) calculations.[8,9] Recently, the structure of an A·F-containing DNA dodecamer has been determined by NMR.[10] The structure refinement involved numerous short (∼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.

Extended MD simulations of oligonucleotides and related molecules, including solvent and considering long-range electrostatic effects, can give reliable structural and dynamic information.[11,12] Using these methods,[13] we have performed a total of 10 ns of MD simulations on an A·F-containing dodecamer and 1.5 ns on its "parent", A·T-containing sequence (Figure 1).

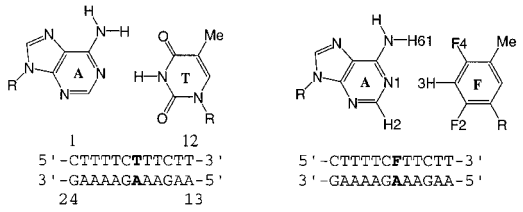

Figure 1. Structures of (left) A·T and (right) A·F base pairs and dodecamer sequences simulated, showing numbering scheme.

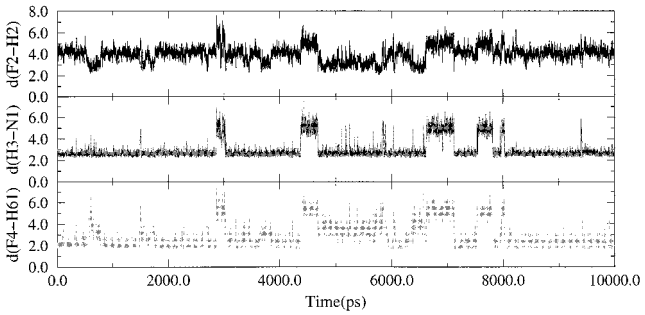

Figure 2. Variation in selected A·F nonbonded distances (in Å) through the trajectory.

The first 1.5-ns MD trajectories for both dodecamers were monitored by measuring the RMS deviation of the snapshots from reference canonical A- and B-form structures. Within 0.5 ns, both simulations reached a plateau about 4 Å RMS deviation from the B-form reference, but about 5 Å away from the A-form reference (not shown). An essentially B-form structure was maintained in both cases, the RMS deviation resulting largely from a reduction in helical twist.[12d] The last 1 ns of each simulation was used to generate time-averaged structures. These showed an RMS deviation of only 1.4 Å, indicating that both sequences were adopting similar conformations, in agreement with the NMR data[10]

However, plotting the lengths of the "hydrogen bonds" in the A·F base pair over the simulation (Figure 2) showed that the conventional orientation between these bases was lost for about 200 ps, beginning at about 600 ps. This event was found to involve the swinging out into the major groove of the A and F bases—a base pair breathing motion.

Cieplak et al.[12f] have previously reported the breathing of a terminal base pair, but "end effects" inevitably limit the generality of conclusions that can be drawn from that study. To our knowledge, this is the first time that a spontaneous breathing event of an "internal" base pair has been observed in the MD simulation of a DNA duplex, at least with the current generation of force fields and simulation protocols. This is not unexpected; for a natural Watson−Crick base pair, the breathing frequency is estimated to be in the microsecond range. The weaker nature of the A−F interaction would appear to have moved this process into a time scale accessible by atomistic MD. Solvent exchange data from the NMR studies support this conclusion.[10]

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

(1) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252.
(2) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.
(3) Johnson, K. A. *Annu. Rev. Biochem.* **1993**, *62*, 685.
(4) Guckian, K. M.; Kool, E. T. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2825.
(5) Moran, S.; Ren, R. X.-F. Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506.
(6) Liu, D. Y.; Moran, S.; Kool, E. T. *Chem. Biol.* **1997**, *4*, 919.
(7) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056.
(8) Evans, T. A.; Seddon, K. R. *J. Chem. Soc., Chem. Commun.***1997**, 2023.
(9) Wang, X.; Houk, K. N. *J. Chem. Soc., Chem. Commun.* **1998**, 2631.
(10) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nature Struct. Biol.* **1998**, *5*, 954.
(11) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463. Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895. Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226.
(12) (a)Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1998**, *281*, 675. (b) Weerasinghe, S.; Smith, P. E.; Mohan, V.; Cheng, Y.-K.; Pettit, B. M. *J. Am. Chem. Soc.* **1995**, *117*, 2147. (c)York, D. M.; Yang, W.; Lee, H.; Darden T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001. (d) Feig, M.; Pettitt, B. M.; *Biophys. J.* **1998**, *75*, 134. (e) Spackova, N.; Berger, I.; Egli, M.; Sponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147. (f) Cieplak, P.; Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6722. (g) Foloppe, N.; MacKerell, A. D. *J. Phys. Chem. B.* **1998**, *102*, 6669. (h) Duan, Y.; Wilkosz, P.; Crowley, M.; Rosenberg, J. M. *J. Mol. Biol.* **1997**, *272*, 553.
(13) All MD simulations were performed using the AMBER suite of programs.[14] Full details of the simulation protocols are included in the Supporting Information.
(14) 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.

10.1021/ja991067t CCC: $18.00
Published on Web 09/01/1999

Table 1. RMS Coordinate Fluctuations (in Å), Averaged over All Atoms within Each Base, over the Breathing Event, 2.8−3.1 ns

strand 1	RMS fluctuation	strand 2	RMS fluctuation
T5	0.27	A20	0.15
C6	0.45	G19	0.45
F7	0.69	A18	1.30
T8	0.41	A17	0.54
T9	0.28	A16	0.33

The simulation of the A·F-containing sequence has been extended to 10 ns. Several further breathing events have been observed (Figure 2). These events may be conveniently divided into two types. More common are partial openings, where the F(F4)−A(H61) distance increases greatly but the F(H3)−A(N1) distance is not greatly extended. Less common are full breathing events, where all "conventional" interactions between these two bases are lost. Taking 3.0 Å as the cutoff for both of these key distances, we calculate the equilibrium population of each state and thus a free energy of partial opening for the A·F pair of ∼0.4 kcal/mol and for complete breathing of ∼0.9 kcal/mol. The choice of cutoff distance is not critical; for instance, a cutoff of 3.2 Å yields ΔG values of 0.5 and 1.0 kcal/mol for partial opening and complete breathing, respectively, while a cutoff of 3.5 Å gives values of 0.6 and 1.0 kcal/mol.

Because of the close shape mimicry of F for T, the mechanics of the breathing process detected here should be applicable to natural base pairs. We have therefore examined these breathing events in some detail. First, we observe that the structural deformations involved are very localized. Table 1 shows the average RMS coordinate fluctuations for selected bases over the time of the breathing event around 3 ns. The bases on either side of the A·F pair are scarcely affected by this major structural transition. This is in general agreement with the studies of Chen et al.[15] on base opening mechanics, but we do not observe a clear correlation between the ζ torsion angle and base opening, as that study suggested; here, the mechanics of breathing are complex and variable from event to event. Second, as the RMS fluctuation data indicate, the breathing motion is asymmetric. In this particular event, the A base unstacks and protrudes from the helix to a much greater extent than the F. However, analysis of the event at around 4.5 ns reveals the opposite—it is the F base that unstacks and protrudes. Intrastrand purine-purine stacking interactions are generally stronger than pyrimidine-pyrimidine ones, favoring F unstacking. However, this process involves a major increase in the solvent exposure of the base concerned, disfavoring F unstacking. Over the 3-ns breathing event, the solvent-accessible surface area (SASA) of the adenine base increases from an average of 210 to about 290 Å^2. In the 4.5-ns event, the SASA for the F base increases from an average value of 220 to about 250 Å^2. In both events, the SASA of the partner base is largely unaffected. This analysis is confirmed by examination of the MD data. The number of water molecules less than 3.5 Å from the polar atoms of the bases (N1, N3, N6, N7, and N9 for A; F2, C3, and F4 for F) increases from about four to eight when A breathes, but only from about two to four when F breathes (results not shown). In conclusion, it would appear, therefore, that stacking and solvation factors are finely balanced in this case.

To complement the MD studies and help validate the AMBER parametrization of F, the thermodynamics of A·T and A·F dimer formation have been determined at the DFT level of theory using the B3LYP functional,[16] with the 6-31G(d) and 6-31G(d,p) basis sets.[17] Results in Table 2 show that the dimerization energy of A·T is around 9 kcal/mol more favorable than that of A·F, the difference being reduced if entropic effects are considered. The total interaction energy of A·F is around 3 kcal/mol, a value equivalent to a weak H-bond, in agreement with previous estimates.[18] Note that free energy calculations suggest that, while the A·T pairing is stable for a 1 M gas-phase reference state, A·F H-bonding is disfavored, as suggested from experimental studies in apolar solvents.[5]

Table 2. Dimerization Energies and Free Energies (Both in kcal/mol) for the H-Bond Dimerization of A·T and A·F[a]

dimer	ΔE (QM)	ΔG (1 atm)	ΔG (1 M)	ΔE (AMBER)
A·T	−12.0/*−12.2*	1.1/*0.6*	−0.9/*−1.3*	−13.9
A·F	−2.9/*−3.0*	7.8/*6.9*	5.9/*5.0*	−3.2

[a] B3LYP/6-31G(d) results are in roman, B3LYP/6-31G(d,p) results in italics.

Table 3. Intrastrand Stacking Energies for T and F with Their Adjacent Bases from the Time-Averaged Structure[a]

stacked pair	ΔE (QM)	ΔE (AMBER)	stacked pair	ΔE (QM)	ΔE (AMBER)
5′-d(CF)-3′	−2.5	−2.8	5′-d(FT)-3′	−3.2	−4.2
5′-d(CT)-3′	−6.0	−7.4	5′-d(TT)-3′	−2.2	−2.5

[a] All values are in kcal/mol.

Table 4. B3LYP Vibrational Frequency Analysis for A·F and A·T

	A·F	A·T	A·F	A·T	A·F	A·T
motion[a]	opening		shear		stretch	
frequency (cm^{-1})	28.5	54.5	56.8	96.2	67.1	109.3

[a] The descriptions of the motions relate to the Cambridge Conventions but are not exact.

Calculation of stacking energies[17] (Table 3) shows that there is no dramatic and consistent difference in the stacking of T and F. These results suggest that the intrinsic stacking abilities of T and F are not dramatically different, and that important sequence effects in the determination of the stability of DNA containing F substitutions can be expected. Future calculations will be focused on the analysis of solvent and environmental effects on the stability of T → F substitutions. Tables 2 and 3 also contain data calculated using the AMBER nonbonded parametrization of F. These help to confirm the accuracy of these parameters and of the MD simulations based on them.

Analysis of B3LYP frequencies has provided a quasi-dynamic picture of the flexibility of A·T and A·F pairs. We observe three normal modes for each that relate to base pair breathing movements (Table 4); all are shifted to considerably lower frequencies in the A·F pair. The QM studies thus explain, at least in part, the frequent occurrence of breathing events found in our MD simulations.

In conclusion, these simulations provide more information on the structure and dynamics of DNA containing A·F base pairs. Furthermore, they give insight into the mechanics of base pair breathing in general. Breathing events can be very localized, and the degree and asymmetry of base pair opening depends on a delicate balance between hydrogen bonding, stacking, and solvation effects. All three, not just H-bonding, may be important in how F can substitute for T in replication. We observe that the stacking interactions of F are not greatly different from those of T, but solvation effects are almost certainly important, as has been recently proposed.[19]

Acknowledgment. We thank the Fulbright Commission, the BBSRC, the Centre de Supercomputacio de Catalunya, and DGICYT for financial support. This is a contribution from the Centre Especial de Recerca en Química Teòrica.

Supporting Information Available: Details of the MD simulation protocols and AMBER parameter files for F and the quantum mechanical calculation protocols (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA991067T

(15) Chen, Y. Z.; Mohan, V.; Griffley, R. H. *J. Biomol. Struct. Dyn.* **1998**, *15*, 765.
(16) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Lee, C.; Yang W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(17) Full details of the protocols used for the quantum mechanical calculations are included in the Supporting Information.
(18) Meyer, M.; Sühnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619.
(19) Kool, E. T. *Biopolymers* **1988**, *48*, 3.