

## Fe de erratas

- Página 189. *Energy calculations*. (1er párrafo, línea 6 empezando por el final)
  - ✗ intramolecular
  - ✓ intermolecular
  
- Página 190. *Energy calculations*. (3er párrafo, línea 4 empezando por el principio)
  - ✗ 8AG-C<sup>+</sup>
  - ✓ 8AI-C<sup>+</sup>
  
- Página 192. **CONCLUSIONS** (Línea 5 empezando por el principio)
  - ✗ 2-aminohypoxanthine
  - ✓ 8-aminohypoxanthine

4.3.2 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 &

Charles A. Laughton

*J. Am. Chem. Soc.* **1999**, *121*, 8653-8654

*(Esta página está intencionadamente en blanco)*

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

Elena Cubero,<sup>†</sup> Edward C. Sherer,<sup>‡</sup> F. Javier Luque,<sup>§</sup> Modesto Orozco,<sup>†</sup> and Charles A. Laughton<sup>\*‡</sup>

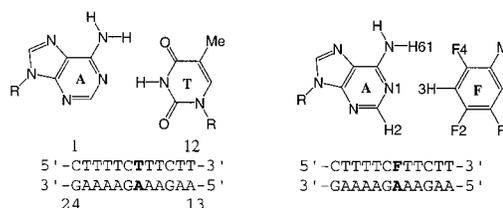
Departament de Bioquímica i Biologia Molecular  
Facultat de Química, Universitat de Barcelona  
Martí i Franquès 1, Barcelona 08028, Spain  
Cancer Research Laboratories  
School of Pharmaceutical Sciences  
University of Nottingham, NG7 2RD U.K.  
Departament de Fisicoquímica  
Facultat de Farmàcia, Universitat de Barcelona  
Avda. Diagonal sn, Barcelona 08028, Spain

Received April 5, 1999

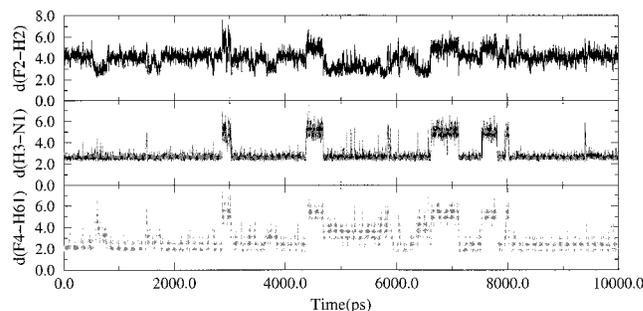
We report the results of extended molecular dynamics simulations of a DNA oligonucleotide containing an adenine-difluorotoluene (A·F) base pair, and of the corresponding “parent” oligonucleotide containing an adenine-thymine base pair. The observation in the former case of spontaneous breathing events 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.<sup>1–4</sup> Studies in chloroform show no evidence that F forms hydrogen-bonding interactions with adenine,<sup>5</sup> but DNA polymerase I will incorporate F across from A, and A across from F, in a precise fashion.<sup>5,6</sup> Despite this specificity, thermal denaturation studies<sup>7</sup> show that replacing T by F destabilizes DNA duplexes by 3.0–3.6 kcal mol<sup>-1</sup>. The significance of these results has been debated in quantum mechanical (QM) calculations.<sup>8,9</sup> Recently, the structure of an A·F-containing DNA dodecamer has been determined by NMR.<sup>10</sup> 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 electro-



**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.

static effects, can give reliable structural and dynamic information.<sup>11,12</sup> Using these methods,<sup>13</sup> 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).

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.<sup>12d</sup> 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<sup>10</sup>

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.<sup>12f</sup> 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.<sup>10</sup>

(13) All MD simulations were performed using the AMBER suite of programs.<sup>14</sup> 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.

<sup>†</sup> Facultat de Química, Universitat de Barcelona.

<sup>‡</sup> University of Nottingham.

<sup>§</sup> Facultat de Farmàcia, 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.; Rummey, 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.

**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 two 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  $\Delta G$  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.<sup>15</sup> on base opening mechanics, but we do not observe a clear correlation between the  $\zeta$  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<sup>2</sup>. In the 4.5-ns event, the SASA for the F base increases from an average value of 220 to about 250 Å<sup>2</sup>. 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,<sup>16</sup> with the 6-31G(d) and 6-31G(d,p) basis sets.<sup>17</sup> 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

**Table 2.** Dimerization Energies and Free Energies (Both in kcal/mol) for the H-Bond Dimerization of A•T and A•F<sup>a</sup>

dimer	$\Delta E$ (QM)	$\Delta G$ (1 atm)	$\Delta G$ (1 M)	$\Delta E$ (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

<sup>a</sup> 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<sup>a</sup>

stacked pair	$\Delta E$ (QM)	$\Delta E$ (AMBER)	stacked pair	$\Delta E$ (QM)	$\Delta E$ (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

<sup>a</sup> 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 <sup>a</sup>						
frequency (cm <sup>−1</sup> )	opening	opening	shear	shear	stretch	stretch
	28.5	54.5	56.8	96.2	67.1	109.3

<sup>a</sup> The descriptions of the motions relate to the Cambridge Conventions but are not exact.

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.<sup>18</sup> 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.<sup>5</sup>

Calculation of stacking energies<sup>17</sup> (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.<sup>19</sup>

**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>.

(15) Chen, Y. Z.; Mohan, V.; Grifflay, 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.

### 4.3.3 Molecular dynamics study of oligonucleotides containing difluorotoluene

**Elena Cubero**, Charles A. Laughton, F. Javier Luque & Modesto Orozco  
*J. Am. Chem. Soc.* **2000**, *122*, 6891-6899

*(Esta página está intencionadamente en blanco)*

# Molecular Dynamics Study of Oligonucleotides Containing Difluorotoluene

Elena Cubero,<sup>†</sup> Charles A. Laughton,<sup>‡</sup> F. Javier Luque,<sup>§</sup> and Modesto Orozco<sup>\*,†</sup>

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 Físicoquímica, 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.<sup>1,2</sup> 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<sup>3</sup> to range between 0.2 and 0.4 kcal mol<sup>-1</sup>. 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.<sup>4</sup> However, other workers<sup>5</sup> 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,<sup>6</sup> which lacks H-bonding capability, even in apolar solvents such as chloroform.<sup>7</sup> 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.<sup>7,8</sup> Despite this specificity, thermal denaturation studies<sup>9</sup> show that replacing T by F destabilized DNA duplexes by 3.0–3.6 kcal mol<sup>-1</sup>. The significance of these results was challenged by Evans and Seddon,<sup>10</sup> 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,<sup>11</sup> 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.<sup>12</sup> 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<sup>13</sup> 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

<sup>†</sup> Facultat de Química, Universitat de Barcelona.

<sup>‡</sup> University of Nottingham.

<sup>§</sup> 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.

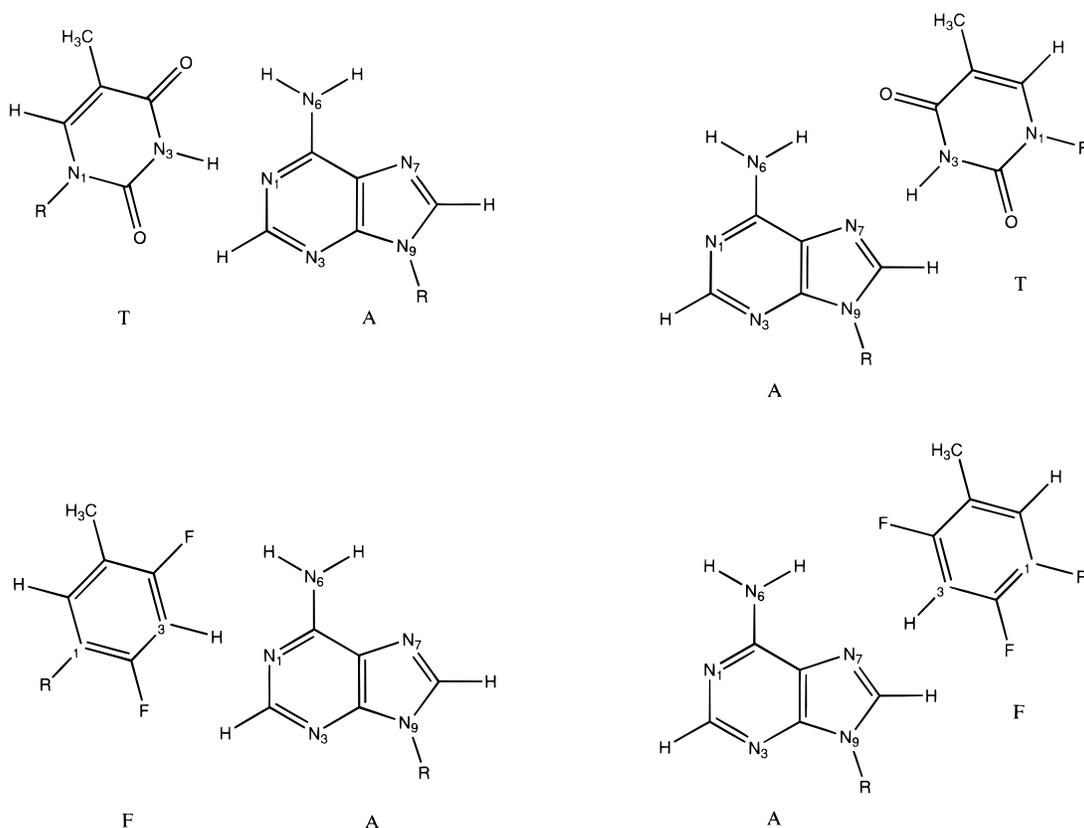
(9) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056.

(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.

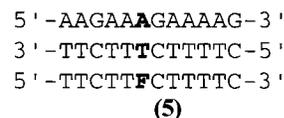
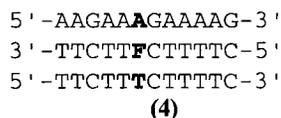
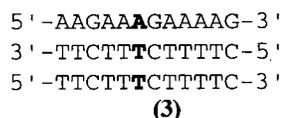
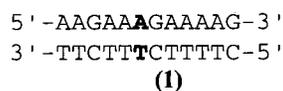
(12) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nature Struct. Biol.* **1998**, *5*, 954.

(13) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653.



**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



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,<sup>14,15</sup> 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<sup>16</sup> in association with the AMBER 95 force field. Missing AMBER parameters for difluorotoluene were those previously determined in our previous studies.<sup>13</sup> 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)<sup>17</sup> method. SHAKE<sup>18</sup> 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.<sup>19</sup> Triplexes **3**, **4**, and **5** were generated using our equilibrated structure

(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.

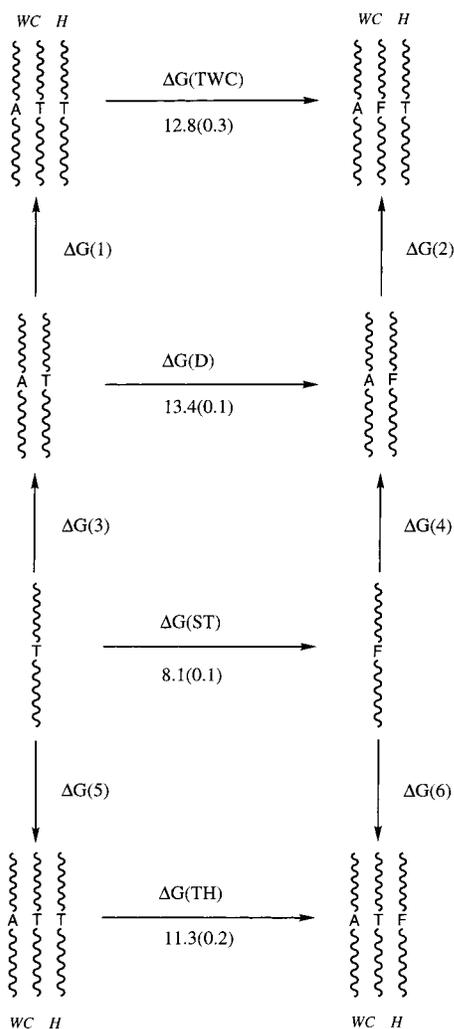
(17) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.

(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.

(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.



**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.<sup>20</sup> 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.<sup>21</sup> The systems were energy minimized, heated, and equilibrated using our standard equilibration process.<sup>20</sup> The equilibrated systems were subjected to 1.5 (**1**, **3**), 10 (**2**), and 3 (**4**, **5**) ns unrestrained MD simulations.

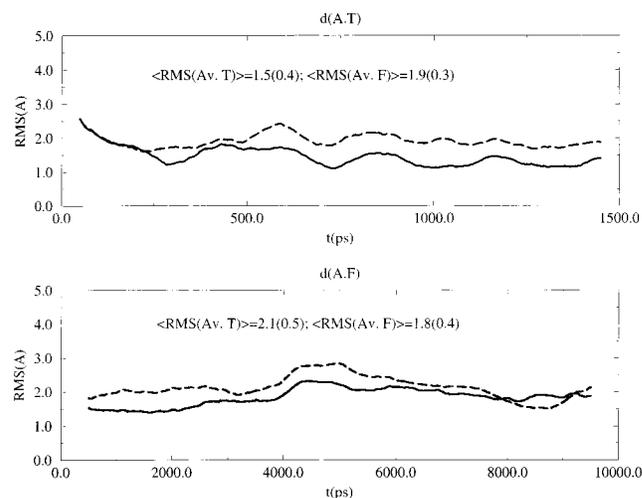
Energy analysis was done using AMBER-5.0 as well as “in-house” programs. Helical parameters were calculated using CURVES.<sup>22</sup> Where CURVES yields both global and local versions of parameters, the local parameter values are quoted.

**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

**Table 1.** Summary of MD/TI Simulations<sup>a</sup>

mutation	system	structure	length of MD
F→T	SS (3 steeps)	equil. for 340 ps <sup>b</sup>	420 ps
F→T	SS (3 steeps)	equil. for 340 ps <sup>c</sup>	420 ps
F→T	SS (3 steeps)	equil. for 340 ps	840 ps
F→T	SS (3 steeps)	annealing (520 ps) <sup>d</sup>	420 ps
F→T	SS (5 steeps)	equil. for 340 ps <sup>b</sup>	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

<sup>a</sup> 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. <sup>b</sup> 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. <sup>c</sup> As in footnote *b*, but a large box of 2851 water molecules (similar to those used for duplex structures) was used. <sup>d</sup> 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(AAGAAA-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.

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).

## Results and Discussion

**MD Simulations. (a) Duplex DNA.** MD simulations of the d(AAGAAAGAAAAG) 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

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

(21) Jorgensen, W. L.; Chandrasekhar, 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 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<sup>a</sup>

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)

<sup>a</sup> 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,<sup>16</sup> 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<sup>20,23</sup> 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<sup>20</sup> (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<sup>20,24,25,24,25</sup> (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.

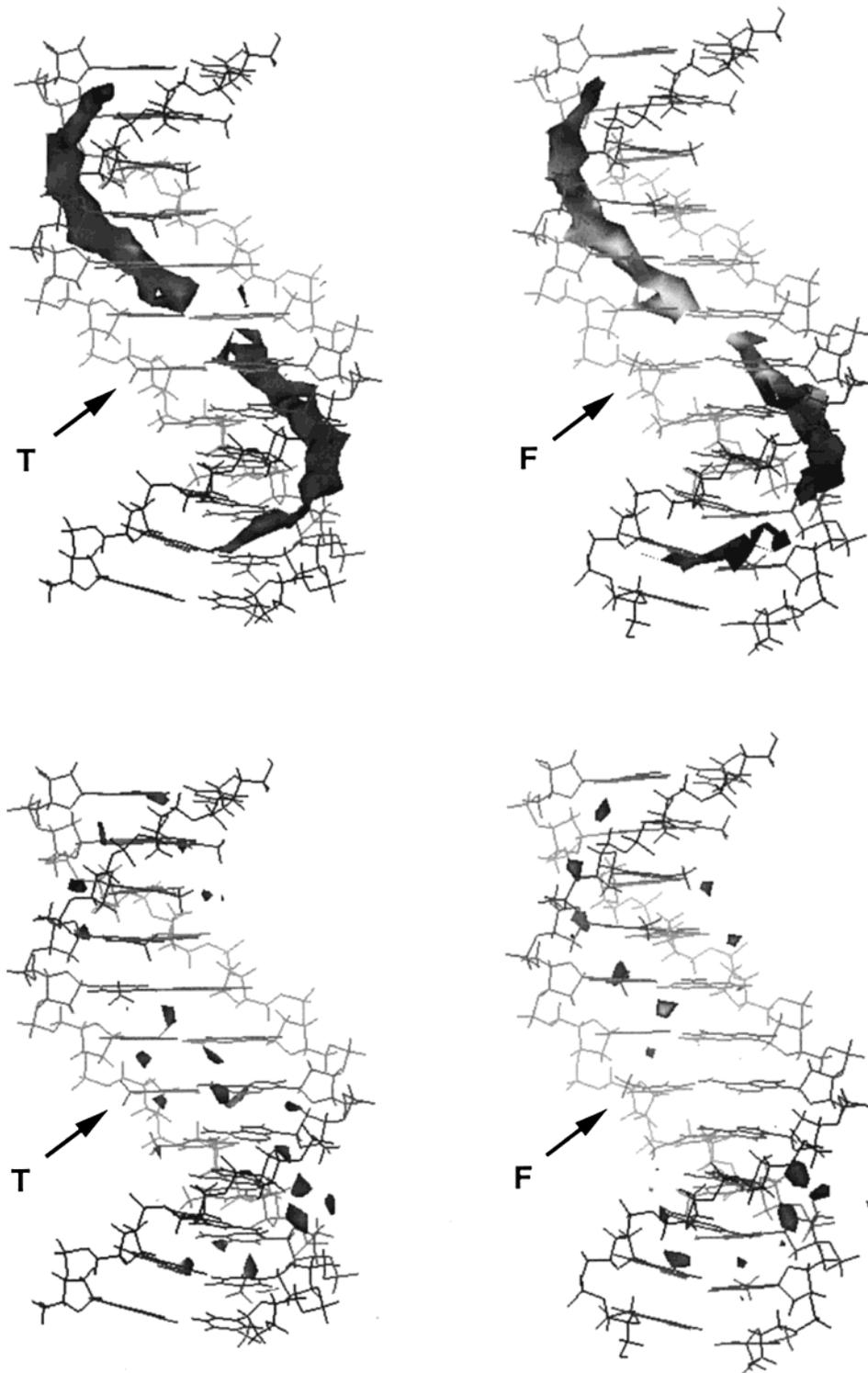
Breathing movements are detected in triplexes containing F (**4** and **5**), while they are not found in the reference triplex **3**. The triplex with the T→F mutation in the Watson–Crick position (**4**) shows reversible partial opening and breathing movements (see ref 13 for nomenclature), which lead 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,<sup>13</sup> 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

(24) Ragunathan, 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.

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



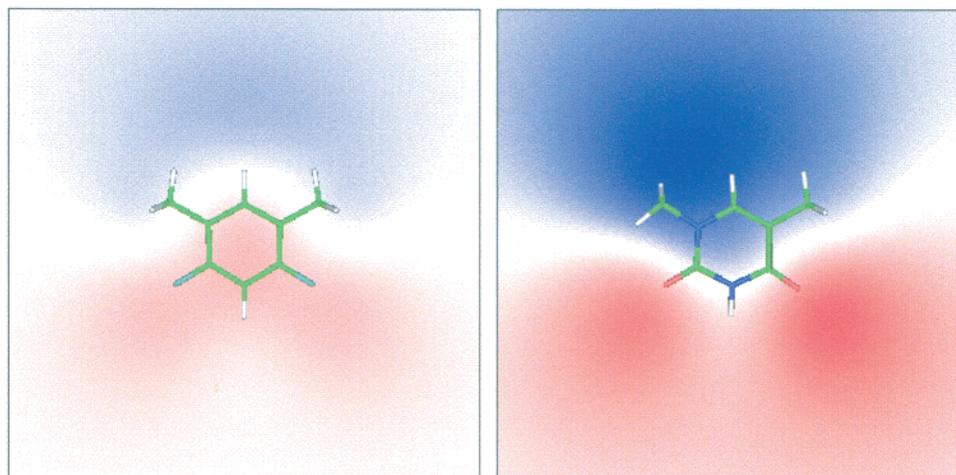
**Figure 4.** Molecular interaction potential (MIP) and solvation maps obtained for the duplexes containing T or F. The MIP map is contoured at  $-5.0$  kcal/mol, and the solvation map is contoured at a water density of around  $2.5$  g/cm<sup>3</sup>.

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.

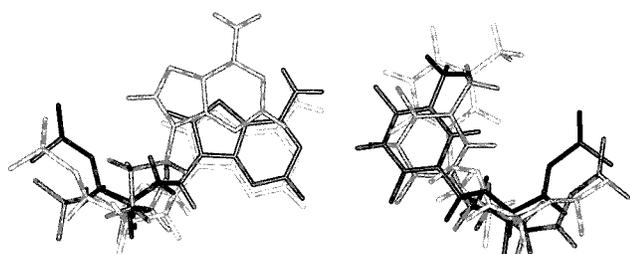
**Free Energy Calculations.** Figure 8 shows the free energy profiles associated with the T↔F mutation in single-stranded

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).<sup>26</sup> 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)<sup>a</sup>

	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)

<sup>a</sup> 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<sup>a</sup>

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)

<sup>a</sup> 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,<sup>27</sup> 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.<sup>28</sup>

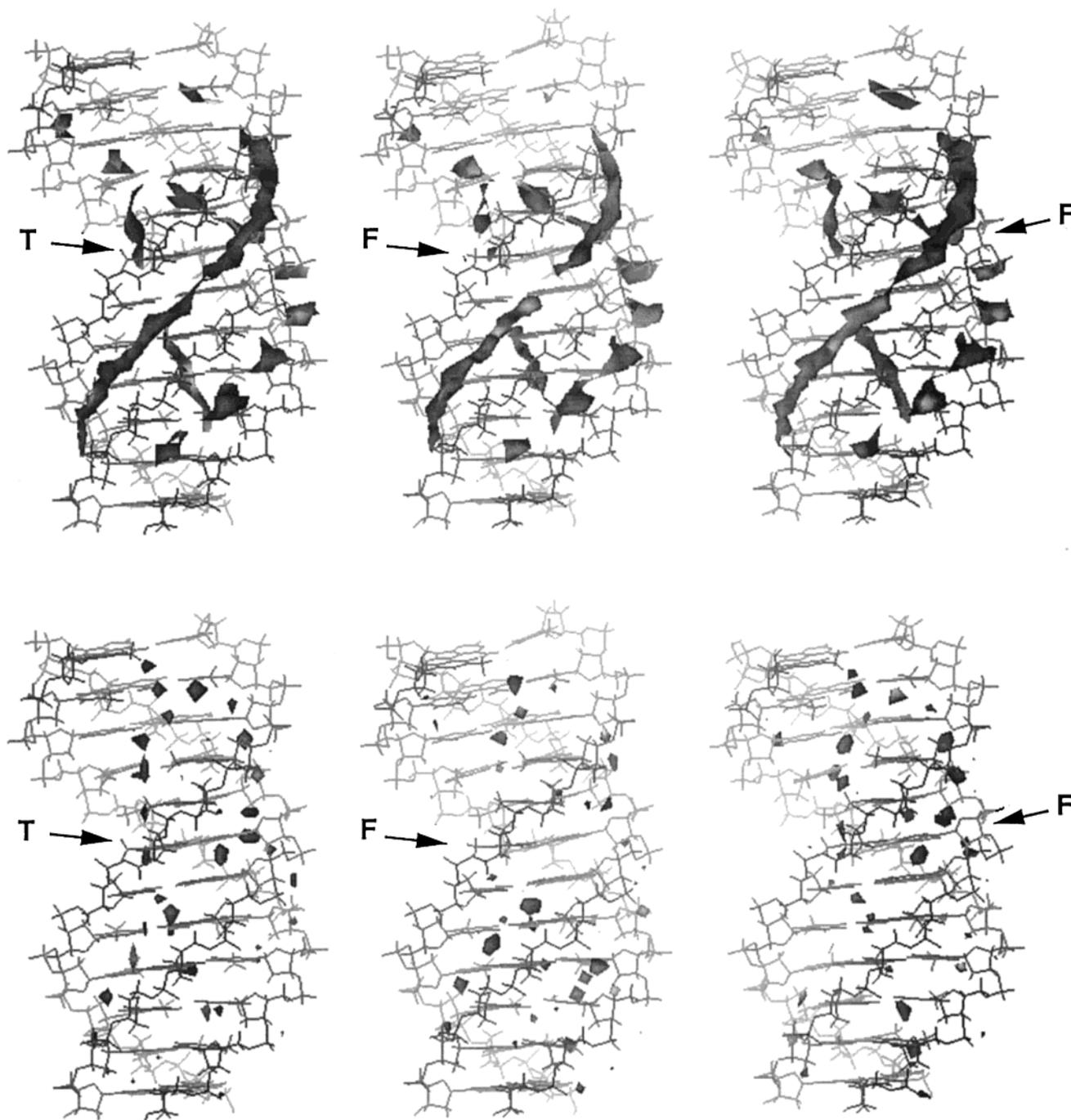
In summary, MD and MD/TI simulations strongly suggest that the presence of a single d(A·F) step destabilizes the duplex, but the resulting DNA is still stable at room temperature, in good agreement with all known experimental evidence.<sup>6,7,9</sup> In practice, the largest differences between a DNA 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<sup>29</sup> 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.

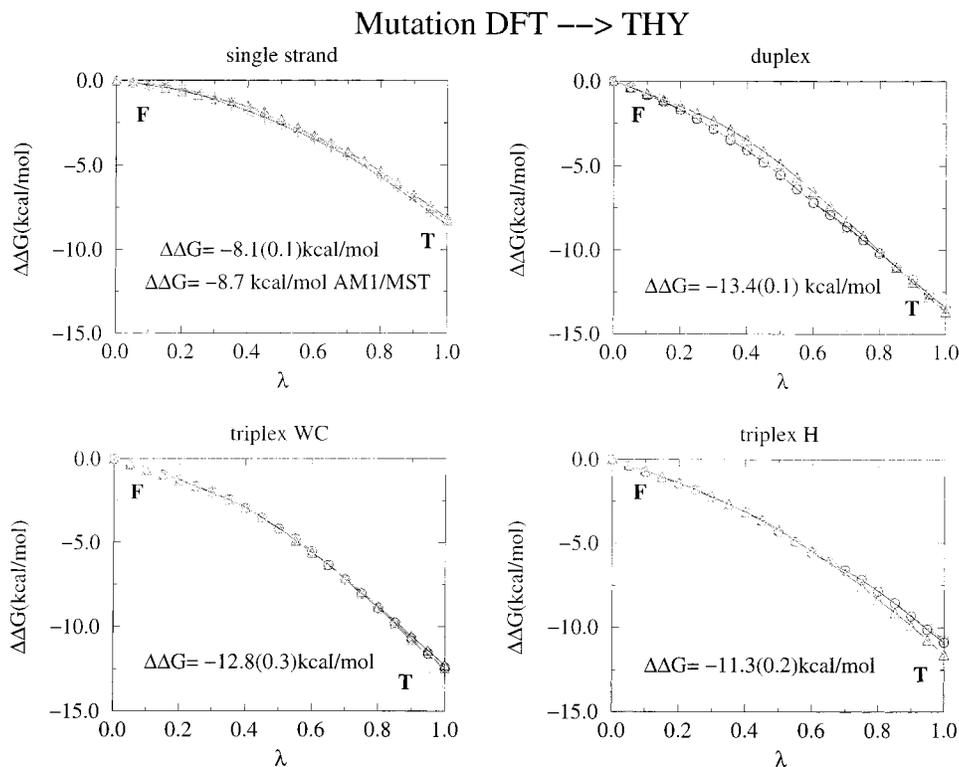


**Figure 7.** Molecular interaction potential (MIP) and solvation maps obtained for the triplexes containing T or F. The MIP maps are contoured at  $-6.0$  kcal/mol, and the solvation maps are contoured at a water density of around  $3.5$  g/cm<sup>3</sup>

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<sup>30</sup>) 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<sup>a</sup>

folding process	sequence	$\Delta\Delta G$ (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

<sup>a</sup> 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 Steps (d(GpApA)) of the Duplexes Containing the d(A·T) and d(A·F) Pair<sup>a</sup>

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

<sup>a</sup> 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,<sup>31</sup>

(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(GpApA)) of the Triplexes Containing the d(A·T-T), d(A·F-T), and d(A·T-F) Trios<sup>a</sup>

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

<sup>a</sup> 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 Catalanian 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

*(Esta página está intencionadamente en blanco)*

## 4.4 Interacciones no canónicas entre bases<sup>1,22</sup>

Las interacciones no canónicas entre bases dan lugar a toda una serie de estructuras no estándar de los ácidos nucleicos. En este bloque se estudiarán los dúplexes de ADN paralelos, que pueden formarse a partir de dos patrones diferentes de puentes de hidrógeno: tipo Hoogsteen (H) y tipo Watson-Crick reverso (rWC), en lugar del apareamiento por puente de hidrógeno Watson-Crick (WC).

Los primeros estudios de dúplexes paralelos ricos en d(A·T) apoyan el modelo de interacción rWC, los cuales son menos estables que los correspondientes dúplexes paralelos<sup>23</sup>. Sin embargo, este modelo ha sido puesto en cuestión por datos experimentales que indican que la interacción H es más estable cuando las purinas están modificadas en posición 2 y en secuencias con mezcla de pares d(A·T) y d(G·C)<sup>24,25</sup>.

En el apartado 4.4.1 se presenta un estudio teórico sistemático de la estructura, la flexibilidad, la estabilidad y las propiedades de reconocimiento molecular de los dúplexes paralelos basados en el motivo d(A·T). Los resultados se comparan con los obtenidos en los dúplexes antiparalelos con la misma secuencia. Se describen las características mediante simulaciones de dinámica molecular (MD)<sup>16</sup> y mecánica molecular acopladas a cálculos de Poisson-Boltzman (MM-PB/SA)<sup>26</sup>.

En el apartado 4.4.2 se describe la estabilidad de los apareamientos WC y H entre bases canónicas A·T, G·C y sus derivados 8-aminopurina, partiendo de su geometría estándar en un tríplex tipo B-DNA.

La síntesis de dúplexes paralelos que contienen 8-aminopurinas y los estudios experimentales de estabilidad han sido realizados por el grupo del Dr. R. Eritja.

El estudio de unión mediante experimentos de *gel-shift* ha sido realizado por el Dr. F. Azorín.

En el apartado 4.4.3 se analiza la estructura de una serie de dúplexes paralelos que contienen una mezcla de pares G·C y A·T. También se analiza la habilidad de los derivados 8-aminopurina para estabilizar dúplexes paralelos. Este estudio teórico se ha llevado a cabo mediante simulaciones de dinámica molecular (MD) e interacción termodinámica (TI)<sup>16</sup>.

La síntesis de dúplexes paralelos que contienen 8-aminopurinas, los experimentos de fusión y los estudios espectroscópicos de dicroísmo circular (CD) han sido realizados por el grupo del Dr. R. Eritja.

El estudio de los dúplexes paralelos por resonancia magnética molecular (RMN) ha sido realizado por el Dr. C. González.

4.4.1 Theoretical studies of d(A:T)-based parallel-stranded DNA duplexes

**Elena Cubero**, F. Javier Luque & Modesto Orozco

Aceptado en *J. Am. Chem. Soc.* **2001**

*(Esta página está intencionadamente en blanco)*

# THEORETICAL STUDIES OF d(A:T)-BASED PARALLEL-STRANDED DNA DUPLEXES

Elena Cubero<sup>1</sup>, F. Javier Luque<sup>2\*</sup> and Modesto Orozco<sup>1\*</sup>

Poly d(A:T) parallel-stranded DNA duplexes based on the Hoogsteen and reverse Watson-Crick hydrogen bond pairing are studied by means of extensive molecular dynamics (MD) simulations and molecular mechanics coupled to Poisson-Boltzman (MM-PB/SA) calculations. The structural, flexibility and reactivity characteristics of Hoogsteen and reverse Watson-Crick parallel duplexes are described from the analysis of the trajectories. Theoretical calculations show that the two parallel duplexes are less stable than the antiparallel Watson-Crick duplex. The difference in stability between antiparallel and parallel duplexes increases steadily as the length of the duplex increases. The reverse Watson-Crick arrangement is slightly more stable than the Hoogsteen duplex, the difference being also increased linearly with the length of the duplex. A subtle balance of intramolecular and solvation terms is responsible for the preference of a given helical structure.

---

<sup>1</sup> Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain.

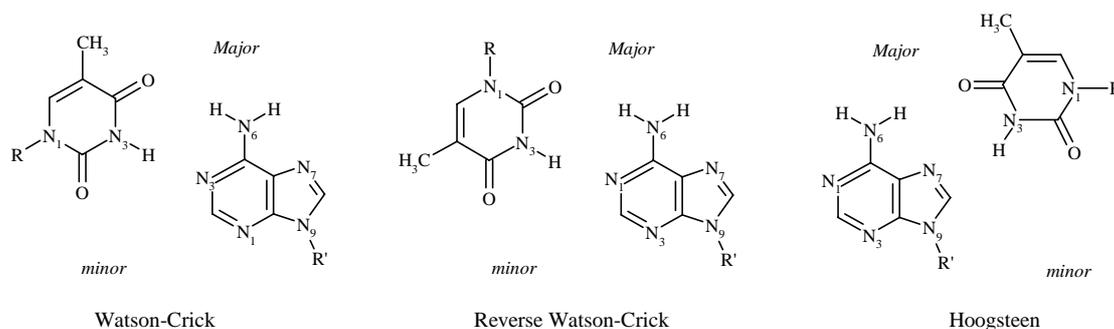
<sup>2</sup> Departament de Físicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda. Diagonal s/n, Barcelona 08028, Spain.

\* Corresponding authors.

## INTRODUCTION

DNA duplexes in physiological environments, and under most laboratory conditions, are antiparallel (i.e. one strand runs 5'→3' and the complementary 3'→5'). However, it is known since the early eighties that parallel arrangements are also possible (1). Thus, parallel DNAs have been found in several hairpins and linear DNAs (2-13), and regions with propensity to form parallel stranded DNA have been detected in specific chromosome regions (14-19).

Instead of the Watson-Crick (WC) hydrogen-bond (H-bond) pairing, parallel stranded DNAs can be formed following two different H-bond patterns (see Figure 1): i) the Hoogsteen (H) scheme, and ii) the reverse Watson-Crick (rWC) one. Interestingly, the H parallel duplex can be used as template for the formation of triplexes (see Figure 2), which can have relevant implications for biotechnological purposes, as well as for the design of antigene and antisense therapies (13,20-22).

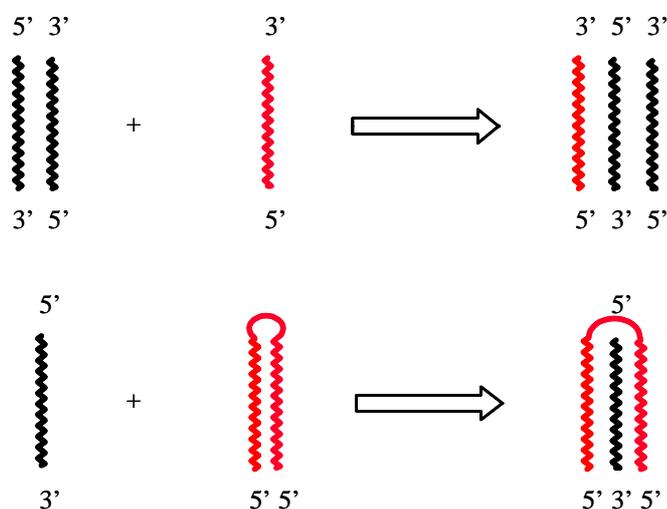


**Figure 1.** Schematic representation of the Watson Crick (WC), reverse Watson Crick (rWC) and Hoogsteen (H) pairings A:T. Nomenclature of grooves for non-standard DNAs is based on homology with WC duplexes (rWC) and on triplex DNA (H).

The first studies on parallel DNA duplexes rich in d(A:T) supported the rWC model for parallel duplexes, which were shown to be less stable than the corresponding antiparallel duplexes (2-8,23). Structural studies using modeling methods (24) and high-resolution NMR data (8,23,25) further validated the rWC pairing of parallel duplexes. However, this model has been challenged by recent experimental data, which indicates that the H pairing is more stable in duplexes where purines are modified at position 2 (27), and in DNAs with mixed d(A:T),d(G:C) sequences (9,13,20-22,27-29). In addition, it has been found that H-type parallel duplexes can be stabilized by the binding of drugs (28). More surprising, it has been shown that antiparallel duplexes based on the H-motif can be even more stable than the parent antiparallel duplex in certain experimental conditions (9,27).

Sequences rich in d(G:C) pairs have an intrinsic preference for the H-form, specially at acidic pH (9,13,20-22,27-29). This arises probably from: i) the problems of G and C to form rWC H-bond pairs, and ii) the formation of strong ionic H-bond between the guanine and the cytosine in the Hoogsteen

arrangement. The preference of sequences d(A:T) for the rWC pairing is unclear, considering that the Hoogsteen side is more polar than the Watson-Crick side, and the H-pairing of isolated A and T is slightly more stable than the WC and rWC ones (30-32).



**Figure 2.** Strategies for the formation of triplexes. TOP: single stranded DNA as triplex-forming oligonucleotide; BOTTOM: parallel stranded duplex (hairpin) as triplex-forming oligonucleotide.

In this paper we present the first systematic theoretical study on the structure, flexibility, stability, and molecular recognition properties of parallel duplexes based on the d(A:T) motif. Results are compared with those obtained for antiparallel duplexes of the same sequence.

## METHODS

Parallel poly d(A:T) DNA duplexes based on the H and rWC H-bond pairings have been studied using molecular dynamics (MD) simulations. In order to obtain statistically-significant conclusions and to reduce the noise intrinsic to energy analysis, we have considered nine poly d(A:T) duplexes of different length (the shortest and the largest being 5-mer and 15-mer duplexes), which were built up imposing the H and rWC H-bond motifs. For comparison purposes, antiparallel duplexes of the same sequence and lengths (5-, 7-, 9-, 11- and 15-mer duplexes) were also analyzed.

Starting (canonical) models for the rWC duplexes were defined using Pattabiraman's canonical model (24), which is known to reproduce accurately NMR data (8,23,25). Starting models for the H duplexes were defined from the canonical structure of a poly d(A-T-T) triplex (33). Finally, B-type models for the antiparallel DNA duplexes were defined from Arnott's canonical data (34). Sodium counterions were added using the iterative cMIP approximation (33,35) to maintain neutrality. This approach locates the ions in the preferred positions according to a classical interaction potential computed in a grid around the DNA using van der Waals interactions, and an electrostatic term determined by

solving Poisson-Boltzman equation. Every time an ion is placed the interaction potential is recomputed considering the perturbation introduced by its presence. The procedure was repeated until neutralization of the system.

The neutral DNA systems were hydrated by adding 1340-3734 water molecules, defining simulation boxes ranging from 54872 to 136500 Å<sup>3</sup>. The solvated systems were then optimized, heated (298 K) and equilibrated for 130 ps using our standard multi-stage process (33,36,37). Finally, the 23 duplexes analyzed here were subjected to 1 ns of unrestrained MD simulation at constant pressure (1 atm) and temperature (298 K). In order to verify the convergence in the results trajectories of selected duplexes were extended up to 3 ns. Long-range effects were introduced using periodic boundary conditions and the particle Mesh Ewald technique (PME; 38,39). PME calculations were performed using a grid spacing around 1 Å, a 4<sup>th</sup> order spline, and a tolerance of  $5 \times 10^{-6}$ . All van der Waals interactions beyond 8 Å were ignored. SHAKE (40) was used to maintain all the chemical bonds at their equilibrium distances, which allowed us to use an integration step of 2 fs. The AMBER-99 force field (41,42), in conjunction with the TIP3P (43) water model, was used to describe molecular interactions. The AMBER-5.1 computer program was used for all the MD simulations (44).

Solvation and Molecular Interaction Potential (MIP) calculations were used to examine the molecular recognition properties of the duplexes following the procedure explained in detail elsewhere (33,35-37). The strategy is based on the calculation of the interaction potential between a classical probe particle (typically O<sup>+</sup>) and the DNA in thousands of grid points around the DNA. The interaction energy is computed using a classical Lenard-Jones term, and a solvent-screened molecular electrostatic potential obtained by solving Poisson-Boltzman equation. The average structures obtained during the last 0.5 ns of trajectories were used for MIP calculations. Solvation maps were determined by integrating the water population during the last 0.5 ns of the trajectories.

The MD trajectories were analyzed to obtain the intramolecular energy contribution. Intramolecular energy analysis was performed using the corresponding modules in AMBER-5.1 (44), as well as *in-house* developed programs. The free energy of solvation of the duplexes was determined as the addition of electrostatic and steric contributions. Following the PB/SA method, the electrostatic component was determined by solving the Poisson-Boltzman equation (45) using MEAD program (46,47), an initial grid spacing of 1 Å, and a final (focusing) grid of 0.4 Å. An external dielectric of 80 and an ionic strength of 0.145 M were used to simulate aqueous environment, while the interior of the DNA was simulated by a dielectric constant of 2, which is expected to capture the electronic response of the macromolecule. The solute/solvent boundary was determined using standard van der Waals parameters (46,47) in conjunction with exclusion radii of 1.4 (water) and 2.0 Å (ions). The steric component to solvation was determined by scaling the solvent accessible surface by 0.005 kcal/(mol Å<sup>2</sup>) following Honig and coworkers (48). Solvation calculations were typically done every 10 ps (100 structures), but in selected cases the calculations were done every 5, 2 and 1 ps (see below) to verify the

convergence of the results. In all the cases studied the differences between solvation free energy using short (100 structures) and long (1000 structures) averages were below 0.5 kcal/mol.

Analysis of molecular flexibility was performed using principal component analysis (PCA) following the protocol explained in detail elsewhere (49). This technique allows us to obtain the essential dynamics of a macromolecule, that is the “normal modes” explaining the largest part of the structural variance of molecule along the trajectory. Technically this is achieved by diagonalization of the covariance matrix, i.e. that containing the fluctuation of all the atoms of the system around average positions. Helical analysis was performed using Curves (50). For all the analysis the terminal base pairs were excluded to avoid artifactual results arising from fraying effects.

## RESULTS AND DISCUSION

### Structural description

The trajectories of the d(A:T)<sub>n</sub> duplexes in the WC, rWC and H helical models are stable, as noted in the RMS deviations with respect to the average structure (RMSd<sub>av</sub>) for each simulation, which are clearly below 2 Å for all the simulations (Table 1). As expected for a polymer, the RMSd<sub>av</sub> increases with the size (around 0.08 Å/base pair for WC duplex and 0.06 Å/base pair for rWC and H helices). Comparison of RMSd<sub>av</sub> for the three families of trajectories (Table 1) shows that the H helix exhibits the smallest fluctuations with respect to the average structure, suggesting that it is slightly more rigid than either the WC or rWC duplexes (see below).

Number of residues	Antiparallel Watson-Crick	Parallel Reverse Watson-Crick	Parallel Hoogsteen
5	0.7(0.2) <i>1.1(0.1)</i>	0.8(0.2) <i>1.4(0.3)</i>	0.6(0.1) <i>0.7(0.1)</i>
6		0.8(0.2) <i>1.3(0.2)</i>	0.7(0.1) <i>0.8(0.2)</i>
7	0.9(0.2) <i>1.5(0.2)</i>	0.8(0.2) <i>1.4(0.2)</i>	0.8(0.1) <i>0.8(0.2)</i>
8		1.1(0.2) <i>1.6(0.4)</i>	0.8(0.2) <i>0.9(0.1)</i>
9	1.0(0.2) <i>2.1(0.3)</i>	1.2(0.3) <i>1.8(0.3)</i>	0.8(0.2) <i>0.9(0.1)</i>
10		1.1(0.2) <i>1.7(0.2)</i>	0.9(0.2) <i>1.0(0.2)</i>
11	1.0(0.2) <i>2.2(0.3)</i>	1.2(0.4) <i>2.1(0.4)</i>	1.0(0.3) <i>1.2(0.3)</i>
12		1.3(0.3) <i>2.3(0.4)</i>	1.1(0.3) <i>1.4(0.3)</i>
15	1.5(0.4) <i>2.8(0.6)</i>	1.4(0.3) <i>2.9(0.5)</i>	1.2(0.3) <i>1.4(0.2)</i>

**Table 1.** Root mean square deviation (RMSd in Å) between the trajectories of the 9 d(A:T) duplexes considered in the study in the Watson-Crick, reverse Watson-Crick and Hoogsteen helical structures and: i) the average structure of the trajectory (in roman) and ii) the starting (canonical) structure (*in italics*). The base pairs at the ends of the helices are eliminated from the analysis.

Table 1 also shows the RMS deviations with respect to the starting (canonical; see Methods) structure ( $\text{RMSd}_{\text{can}}$ ) for each duplex. All the  $\text{RMSd}_{\text{can}}$  are reasonable (clearly below 3 Å even for the largest duplexes) and, as expected, increase with the length of the duplex (0.20, 0.05 and 0.08 Å/base pair for WC, rWC and H helices). The small  $\text{RMSd}_{\text{can}}$  values for the H-duplex (where the canonical models were derived directly from  $d(\text{A}\cdot\text{T}\cdot\text{T})_n$  triplexes (33)) is specially noticeable, suggesting that the H parallel stranded duplex is pre-organized to recognize a pyrimidine strand and form a triple helix (the canonical H-duplex was defined from the Hoogsteen strands of a triplex). This opens an important range of possibilities for these duplexes in antigene and antisense therapies (22).

The helical characteristics of the duplexes are well preserved for all the oligonucleotides, even for the shortest ones. The only local distortions are found for the terminal base pairs, which often display “fraying” movements with loss of H-bonds. These movements, which are common in  $d(\text{A}:\text{T})$  sequences, are however limited to the ends of the helices, and do not introduce major distortions in the rest of the structure. This is demonstrated by the fact that the helical parameters for a family of structures are very similar, irrespective of the length of the oligonucleotide, as noted in Table 2, where helical parameters for the 11 and 15-mer are shown. Helical parameters for B-DNA agrees well with known crystal data for similar sequences, besides a slight underestimation of twist which is common of MD simulations in pure solvent, but is not found when the crystal environment is considered.

Parameter	Antiparallel Watson-Crick	Parallel Reverse Watson-Crick	Parallel Hoogsteen
Twist	33.8±1.0 <i>33.9±1.0</i>	35.1±0.8 <i>34.4±1.2</i>	32.6±2.7 <i>32.5±1.6</i>
Rise	3.4±0.1 <i>3.4±0.1</i>	3.3±0.1 <i>3.4±0.1</i>	3.4±0.1 <i>3.4±0.1</i>
Roll	1.4±1.7 <i>1.9±1.6</i>	0.7±2.2 <i>-0.5±1.5</i>	-4.1±1.4 <i>-4.2±1.0</i>
X-disp	-1.2±0.4 <i>-1.3±0.4</i>	0.7±0.5 <i>0.9±0.4</i>	-0.2±0.4 <i>-0.1±0.4</i>
Phase angle	138±38 138±31	148±37 152±41	109±45 108±51
Minor Groove width	5.3±0.5 <i>5.7±0.6</i>	10.3±0.8 <i>10.6±0.8</i>	3.9±0.5 <i>4.0±0.4</i>
Major Groove width	13.7±1.2 <i>14.1±1.0</i>	7.4±0.5 <i>7.7±0.4</i>	19.8±1.1 <i>19.2±1.2</i>

**Table 2.** Selected MD-averaged helical parameters for the  $d(\text{A}:\text{T})_n$  sequence ( $n=11$  in roman,  $n=15$  in italics) in the WC, rWC and H conformations (standard deviations in parenthesis). The width of the grooves is noted as P-P distances minus 5.8 Å. The rest of helical parameters were determined using CURVES (49). When local and global parameters are available local values are shown. The base pairs at the ends of the helices are eliminated from the analysis. Twist, roll and phase angles are in degrees, the rest in Å.

Helical parameters of the three duplex families do not show large differences, as can be noted from inspection of Table 2. The rise is around 3.4 Å, and X-disp and roll are small in all cases. The twist for H duplex is smaller than for the WC or rWC ones, which suggests that the small twist found in DNA

triplexes (around 29 degrees from ref. 33) stems from the intrinsic low twist of the H duplex. The differences in twist between WC and rWC duplexes are probably within the statistical noise of the simulation. All the sugar puckerings are in the *South-East* region of the pseudorotational cycle, but there is clear a displacement towards the *East* region for the H duplex, and towards the *South* region for the WC and rWC duplexes. Major differences between the different helices are found in the structure of the grooves (see Figures 1 and 3). The WC helix has the well-known narrow minor (around 11 Å), and wide major grooves (around 20 Å), while the rWC duplex shows two similar, but not identical, grooves (around 13 and 16 Å for the major and minor grooves). Finally, the H helix has a completely different pattern of grooves, with a very narrow minor groove (around 10 Å), which reproduces the minor-Major groove of triplexes (33,36,37), and a very wide major groove (around 25 Å), which is related to the Major-Major groove of DNA triplexes (34,36,37).

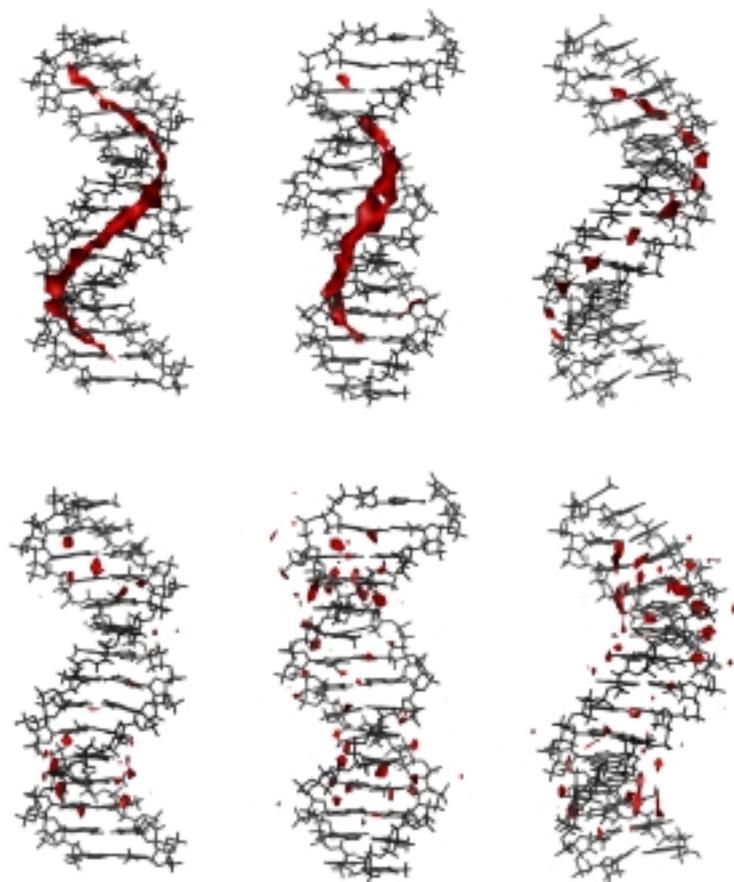
### **Molecular Recognition properties**

The ability of the duplexes to interact with small cationic molecules and with water was analyzed by means of MIP and solvation maps (see Methods; detailed explanations can be found in refs. 33,35-37). Because similar results were found for the different oligonucleotides examined in a given helical family, we limit the discussion to the values obtained for the 15-mer duplexes.

There are remarkable differences between the three helical duplexes concerning the location of the regions most favorable for interaction with small cations (Figure 3). The MIP map for the WC duplex exhibit the expected recognition profile for a B-DNA, with a wide and continuous region corresponding to favorable interaction along the minor groove. For the parallel rWC helix most of the negative valued MIP regions are found in the minor groove, where cations can interact simultaneously with N3 (A) and O4 (T) (see Figure 1). Regions of favorable interaction are also found in the major groove, but they are smaller than those located in the minor groove owing to the destabilizing contribution due to the amino group of adenine (see Figure 1). Finally, for the H duplex the proximity of the phosphate groups and the presence of the O2 atom of T generate a very favorable interaction site in the very narrow minor groove. The contour is, however, discontinuous owing to the small size of the groove, which might lead to notable steric hindrance for the interaction with large, positively charged molecules.

Solvation maps in Figure 3 illustrate the ability of the three helices to interact with water molecules. As expected (51), the minor groove is the best hydrated region of the WC duplex owing to the negative electrostatic potential at that region, and to the existence of H-bond acceptors (O2 and N3 atoms) at the bottom of the groove. The situation for the rWC duplex is different, as both minor and major grooves appear equally solvated. This correlates with the similar size of the two grooves (see above) and to the presence of H-bond donor and acceptors in the bottom of the two grooves (O4 and N3 coordinate waters in the minor groove; N7 and specially N6 are H-bonded to waters in the major groove). It is worth noting that the H-bond donor properties of N6, which handicapped the interaction with a small cation (see above), are favorable for hydration. Finally, the H duplex shows a major region of preferential solvation

along the minor groove, which corresponds to waters bound to O2 (T), and a minor region spread along the major groove, which reflects waters bound to O4 (T) and N6 (A).



**Figure 3.** TOP: Classical molecular interaction potential (cMIP) of the three helical models when the probe molecule is  $O^+$  (contour level  $-5.0$  kcal/mol); BOTTOM: solvation maps for the three helical models (contour level correspond to a density of 2 g/ml, i.e. to a preferential solvation of  $-0.4$  kcal/mol). See text for details.

### Molecular flexibility

Principal component analysis (PCA) from the covariance matrices obtained during the trajectories was performed to analyze the flexibility of the three duplexes. In general, the flexibility of the double helix arises from a complex, wide range of low and high frequency motions. As found in previous studies (49), the first modes, which correspond to low frequency motions, explain a very significant part of the structural variance of the trajectories (see Table 3). Inspection of the eigenvectors (modes) associated with the eigenvalues (frequencies) shown in Table 3 demonstrates that the essential dynamics of the three duplexes is similar in that they are controlled by global bending and twisting of the helix. The WC is the most flexible structure, as noted in configurational entropies (52-54) around  $2.205 \text{ kcal mol}^{-1} \text{ K}^{-1}$  (15-mer duplex), which compare with entropy values of  $2.181 \text{ kcal mol}^{-1} \text{ K}^{-1}$  for the H-duplex and  $2.145 \text{ kcal mol}^{-1} \text{ K}^{-1}$  for the rWC one (52). It is worth noting that the larger flexibility of the WC duplex detected in entropy calculations performed using all the modes (52-54) is also detected in the analysis of

the lowest frequencies, which are 5-9 cm<sup>-1</sup> smaller than the corresponding values in the parallel stranded duplexes (see Table 3).

Helix	Frequencies (cm <sup>-1</sup> )	% Variance explained	Mode description
Antiparallel	19	39	Global bending
	26	20	Global bending
	39	9	Twisting
RWC parallel	29	32	Twisting and bending
	35	20	Global bending
	44	10	Twisting
H parallel	25	25	Global bending
	32	17	Twisting and bending
	46	11	Twisting

**Table 3.** Frequencies (cm-1), percentage of variance explained, and physical description of the modes corresponding to the first, second and third principal components obtained after diagonalization of the covariance matrix for the three helices in the 15-mer trajectories. See text for details.

In summary, the essential dynamics of the three helices are similar, which suggests that the general helical structure, rather than the specific sequence pattern, determine the major conformational movements of the DNA. However, it is clear that the three helices have different flexibility, the WC antiparallel structure being more flexible than the two parallel duplexes.

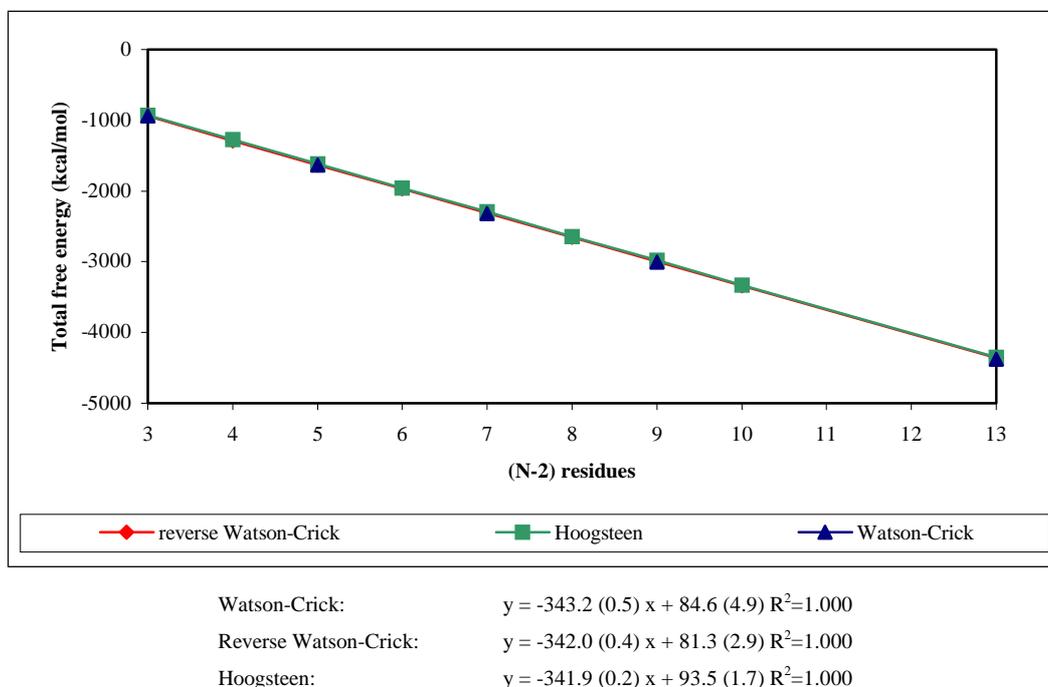
### Global energy analysis

The analysis of the trajectories allowed us to study the energetic characteristics of the three duplexes for a common d(A:T)<sub>n</sub> sequence. The stability of a helical structure can be determined as shown in equation 1 where E is the intra-molecular energy (as computed by AMBER force field), and G<sub>sol</sub> is the solvation (free) energy obtained using the PB/SA technique (see Methods), and the MD trajectories obtained considering explicit solvent. If the entropy corrections are ignored (52; for a qualitative discussion on entropy contributions see above) the free energy difference between two helical structures can be approximated using equation 2.

$$G_{TOT} = E_{intra} + G_{solv} - T S_{intra} \quad (1)$$

$$\Delta G_{TOT}^{A-B} = (E_{intra}^A - E_{intra}^B) + (G_{solv}^A - G_{solv}^B) \quad (2)$$

The use of equation 2 implies the comparison of very large numbers, which makes it necessary to verify the statistical significance of the results. For this purpose, the energy analysis was repeated for 5 (WC) and 9 (rWC and H) different helices, and the standard errors were determined for all the averaged values (see Table 4). For the 9-mer duplex simulations were extended to 3 ns to analyze *bias* in the results derived from the use of too short simulations (see Table 4). The small standard errors (56) and the excellent agreement between values for 1 and 3 ns trajectories support the quality of the MD-averaged results presented in Tables 4-6. However, to even increases the statistical confidence in the results, all the energy estimates in Tables 4-6 were subjected to regression analysis to obtain general trends for the different families. This expensive and lengthy strategy is expected to reduce drastically the noise of the results. All this accurate statistical analysis allowed us to be confident on the estimates of the relative stability between helices obtained by manipulation of very large energy values.



**Figure 4.** Dependence of the total free energy on the length of the duplexes. The regression equations (standard errors in the fitted parameters in parenthesis) and the determination coefficient are displayed. The error bars in the Figure correspond to the standard errors in the average energy estimates.

Figure 4 displays the total free energy of the helical structures as a function of the helix length (without the terminal base pairs). There are perfect ( $r^2=1.0000$ ) linear relationships between the length of the duplex and its free energy. This allows us to obtain accurate estimates of the relative nucleation free energy (the intercept of the equations), as well as to determine the relative stability as the length of the duplex increases (the slopes of the equations). The small magnitude of the errors (Figure 4) in the intercepts and slopes, and the perfect determination coefficient ( $r^2=1.0000$ ) guarantee the statistical quality of the fitted equations. At this point it is worth noting that caution is necessary when comparing

nucleation energies reported here cannot with experimental values, since the “unfolded” form of the duplex is not considered as reference in our calculations. However, taking advantage of the fact that the unfolded structure should be identical for all the duplexes of a given length, relative values of the three helical forms can be rigorously compared with experimental values.

Inspection of the regression equations, and energy values in Table 4 shows that the antiparallel WC helix is the most stable structure, followed by the parallel rWC helix, while the H helix is the least stable arrangement for a poly d(A:T) duplex. The best nucleation free energy (intercepts in Figure 4) is found for the rWC helix, the WC helix displays only slightly worse values, while the nucleation of the H helix is clearly less favored. The WC helix shows the larger gain in stability when the length of the duplex increases, as noted in the slope (helix growth) of the regression equation in Figure 4, which justify the preference for the antiparallel duplex found even for small duplexes. Interestingly, the slopes of the regression equations for rWC and H helices are identical, which indicates that the difference in stability between the two helices does not stem from the length of the duplex, but from the intrinsic differences in the nucleation free energies.

Number of residues	Antiparallel Watson-Crick	Parallel Reverse Watson-Crick	Parallel Hoogsteen
5	-941(0.5)	-938(0.7)	-933(0.8)
6		-1287(0.7)	-1275(0.9)
7	-1632(1.2)	-1632(1.1)	-1614(1.2)
8		-1970(1.6)	-1959(1.4)
9	-2321(1.9) <i>-2321(1.2)<sup>&amp;</sup></i>	-2314(1.8) <i>-2315(1.6)</i>	-2297(1.9) <i>-2298(1.1)</i>
10		-2657(2.6)	-2642(2.3)
11	-3009(2.8)	-2999(3.0)	-2981(3.3)
12		-3338(3.4)	-3328(3.0)
15	-4373(6.4)	-4361(5.2)	-4351(6.0)

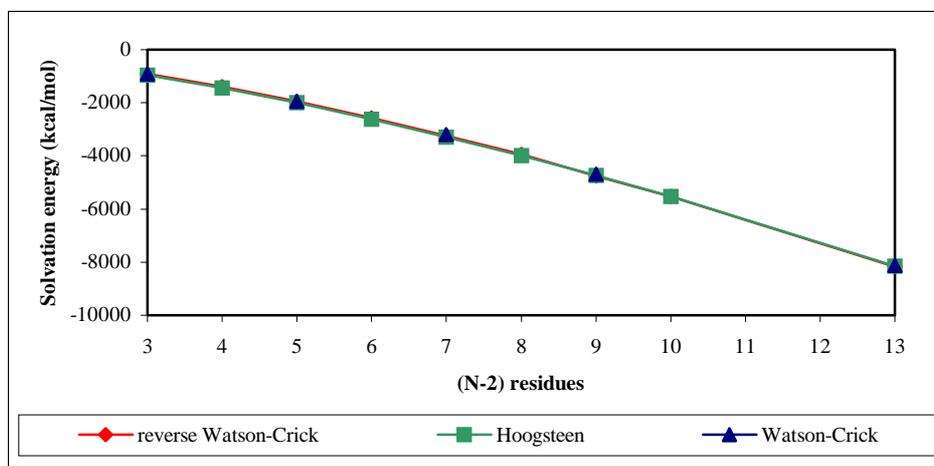
**Table 4.** Total Energy for the three families of helical structures considered in this study. Standard errors in the averages are displayed in parenthesis. All the values are in kcal/mol.

<sup>&</sup> Values in italics correspond to averages obtained during the last 2 ns of unrestrained trajectories of 3 ns.

The preceding results show a qualitatively correct picture of the stability of the d(A:T)<sub>n</sub> duplexes, since the larger stability of the antiparallel helix with respect to the parallel ones and the greater stability of the rWC helix with respect to the H one (2-8,23,25) are accurately predicted (the differences are expected (see above) to be magnified if entropic considerations are included). It is worth noting that a part of our results are indirectly supported by experimental data by Germann et al., who found similar nucleation values for antiparallel and rWC helices, but a much better helix growth factor for the antiparallel helix (8). We hope that our calculations will encourage experimentalists to verify that the difference between rWC and H helices originates in their different nucleation energies.

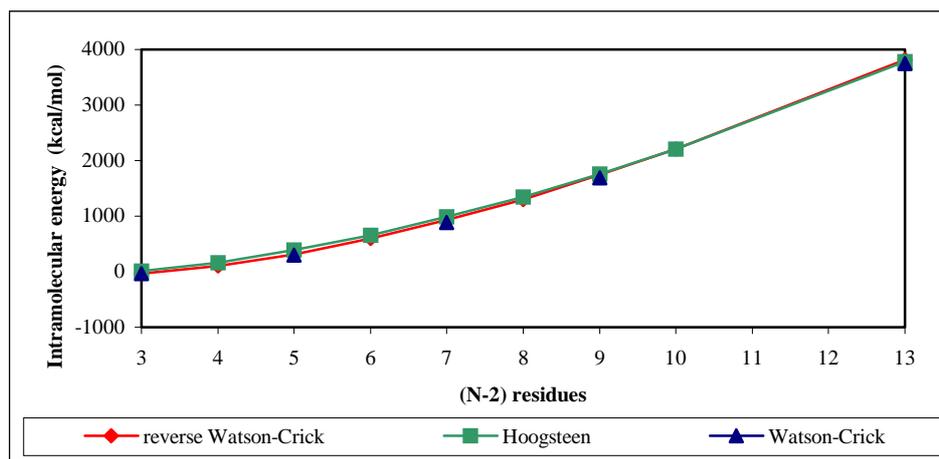
### Components of the molecular energy

The total (free) energy of the three duplexes can be divided into intramolecular and solvation components (see eqs 1-2). For almost all the duplexes (see Table 4) the intramolecular energy is positive, and the solvation term is large and negative, as expected for a very charged macromolecule. Interestingly, there is a second order polynomial dependence of both energy components with the length of the helix (see Figure 5). This indicates strong cooperative effects in both intramolecular and solvation components, which was unexpected considering the linear dependence of the total free energy with the helix length (see Figure 4). In turn, this suggests that there is cancellation of the positive and negative cooperativity of intramolecular and solvation terms (note the similar coefficient of the quadratic term in the fitted equations for the intramolecular and solvation contributions in Figures 5A and 5B)



Watson-Crick:	$y = -23.62 (1.2) x^2 - 346.05 (20.1) x + 363.08 (72.2) R^2=1.000$
Reverse Watson-Crick:	$y = -22.97 (0.9) x^2 - 365.04 (15.2) x + 430.72 (55.0) R^2=1.000$
Hoogsteen:	$y = -22.12 (0.5) x^2 - 368.13 (9.3) x + 374.83 (33.9) R^2=1.000$

**Figure 5A.** Dependence of the solvation free energy components on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parenthesis) and the determination coefficient are displayed. The error bars in the Figure correspond to the standard errors in the average energy estimates.



Watson-Crick:  $y = 23.94 (1.1) x^2 - 2.329 (19.7) x - 261.34 (70.4) R^2=0.9998$

Reverse Watson-Crick:  $y = 23.22 (0.9) x^2 + 19.11 (15.2) x - 336.30 (57.0) R^2=0.9998$

Hoogsteen:  $y = 22.05 (0.6) x^2 + 27.29 (9.2) x - 284.85 (33.3) R^2=0.9999$

**Figure 5B.** Dependence of the intramolecular free energy components on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parenthesis) and the determination coefficient are displayed. The error bars in the Figure correspond to the standard errors in the average energy estimates.

Inspection of Tables 5 and 6, and Figure 5 shows the interdependence between solvation and intramolecular terms. In general, a very stable helix from intramolecular considerations is not well solvated, and *viceversa*. Thus, the antiparallel WC helix has the most stable intramolecular interactions for duplexes studied greater than 5 base pairs (3 central base pairs), but it has also the worst solvation. On the contrary, the H helix leads to the least stable intramolecular interactions for helices shorter than 15 base pairs, but in this range it is also the best solvated structure.

The second order polynomial relationship of the solvation free energy with the length of the duplex can be easily understood considering the dependence of the solvation free energy with the square of the charge (45). The origin of the dependence of the intramolecular energy with the length of the duplex is less clear, which led us to analyze selected components of the intramolecular energy: i) H-bonding, ii) stacking, and iii) phosphate-phosphate repulsion. Results in Figure 6 shows the linear growth ( $r^2 > 0.999$  in all the cases) of the H-bond energy with the length of the duplexes for all the helical structures. The fastest growth and accordingly the strongest H-bond stabilization per base pair is found for the H helix, while WC and rWC display similar values. These results indicate that Hoogsteen H-bonds are stronger than WC and rWC ones, and suggests that H-bonding favors the parallel H helix instead of the WC and rWC ones (30-32), due probably to a greater polarity of the Hoogsteen-side of the purine.

Number of residues	Antiparallel Watson-Crick	Parallel Reverse Watson-Crick	Parallel Hoogsteen
5	-902(0.5)	-904(0.7)	-950(0.7)
6		-1386(0.6)	-1442(0.8)
7	-1936(1.1)	-1943(1.0)	-2003(1.0)
8		-2563(1.4)	-2617(1.3)
9	-3202(1.8) <i>-3206(1.2)<sup>&amp;</sup></i>	-3248(1.6) <i>-3222(1.5)</i>	-3285(1.7) <i>-3282(1.0)</i>
10		-3951(2.4)	-3981(2.1)
11	-4693(2.6)	-4749(2.7)	-4741(3.1)
12		-5539(3.2)	-5539(2.8)
15	-8120(6.0)	-8179(4.9)	-8137(5.7)

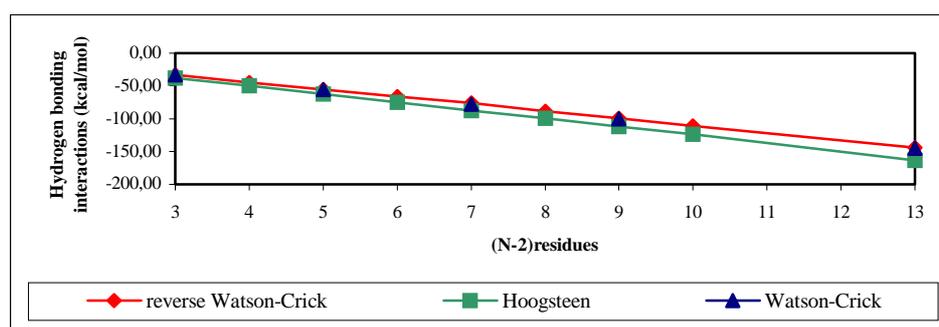
**Table 5.** Solvation free energy for the three families of helical structures considered in this study. Standard errors in the averages are displayed in parenthesis. All the values are in kcal/mol.

<sup>&</sup> Values in italics correspond to averages obtained during the last 2 ns of unrestrained trajectories of 3 ns.

Number of residues	Antiparallel Watson-Crick	Parallel Reverse Watson-Crick	Parallel Hoogsteen
5	-39(0.3)	- 35(0.3)	17(0.3)
6		98(0.4)	167(0.4)
7	304 (0.5)	310(0.5)	389(0.5)
8		593(0.6)	659(0.6)
9	881(0.7)	934(0.6)	989(0.7)
10		1295 (0.9)	1339(0.8)
11	1685(1.0)	1749(1.1)	1760(1.1)
12		2201(1.2)	2212(1.1)
15	3747(2.2)	3818(1.7)	3786(2.0)

**Table 6.** Internal energy for the three families of helical structures considered in this study. Standard errors in the averages are displayed in parenthesis. All the values are in kcal/mol.

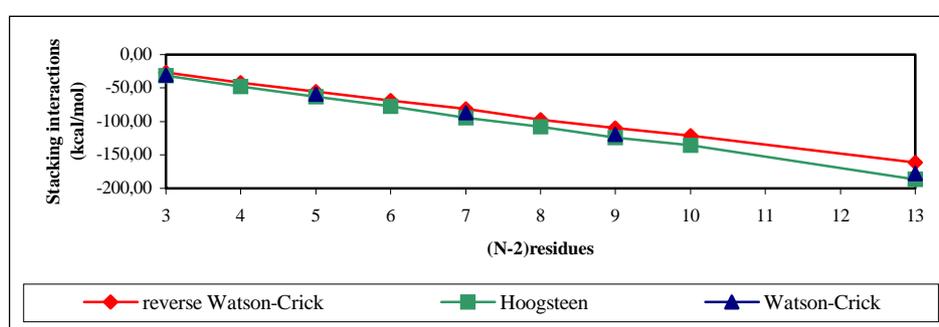
<sup>&</sup> Values in italics correspond to averages obtained during the last 2 ns of unrestrained trajectories of 3 ns.



Watson-Crick:  $y = -11.25 (0.02) x + 0.63 (0.1) R^2=1.0000$

Reverse Watson-Crick:  $y = -11.11 (0.05) x + 0.34 (0.4) R^2=0.9998$

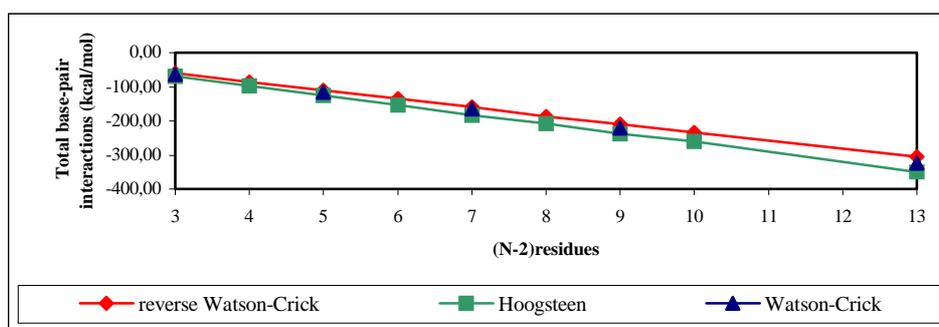
Hoogsteen:  $y = -12.54 (0.08) x + 0.08 (0.5) R^2=0.9998$



Watson-Crick:  $y = -14.80 (0.2) x + 14.59 (1.6) R^2=0.9995$

Reverse Watson-Crick:  $y = -13.42 (0.1) x + 11.85 (1.0) R^2=0.9993$

Hoogsteen:  $y = -15.27 (0.2) x + 13.73 (1.2) R^2=0.9993$



Watson-Crick:  $y = -26.05 (0.2) x + 15.23 (1.6) R^2=0.9998$

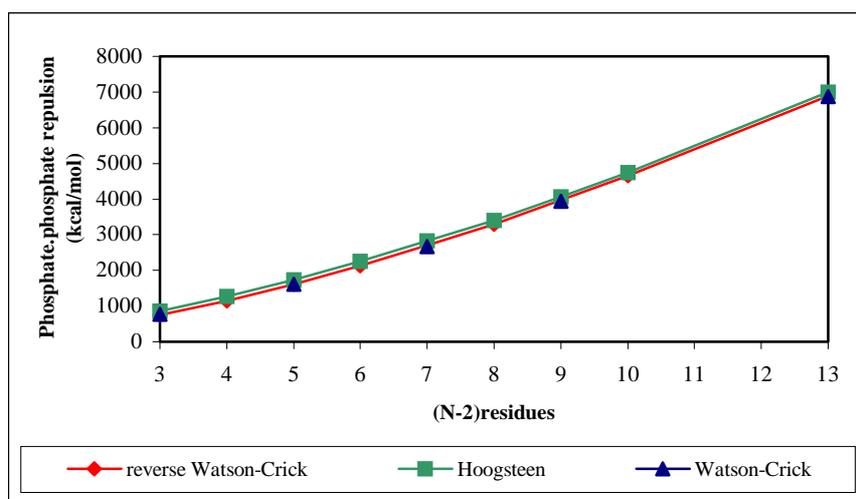
Reverse Watson-Crick:  $y = -24.53 (0.1) x + 12.18 (1.1) R^2=0.9998$

Hoogsteen:  $y = -27.81 (0.2) x + 13.81 (1.7) R^2=0.9996$

**Figure 6.** Dependence of the hydrogen bonding, stacking and total base-pair interactions on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parenthesis) and the determination coefficient are displayed. The error bars in the Figure correspond to the standard errors in the average energy estimates.

The phosphate-phosphate repulsion is clearly larger in magnitude than stacking and H-bond, which indicates that any helicoidal structure must be defined to minimize phosphate-phosphate repulsion,

even when this implies a certain loss of stabilizing (H-bond or stacking) interactions. The phosphate-phosphate repulsion term grows with the length of the oligonucleotide following a second order polynomial ( $r^2 > 0.9999$  in all the cases), indicating that long-range Coulombic repulsion are responsible for the negative cooperativity of the intramolecular energy in DNA duplexes. For the range of oligonucleotides studied the largest phosphate-phosphate repulsions are found for the H-helix, while WC and rWC helices display similar destabilizing phosphate-phosphate interactions. Assuming the goodness of equations in Figure 7 out of the fitted range, the situation might however change for very large oligonucleotides.



Watson-Crick:	$y = 21.77 (1.2) x^2 + 263.66 (20.0) x - 232.43 (71.0) R^2=1.000$
Reverse Watson-Crick:	$y = 20.49 (0.8) x^2 + 290.54 (12.4) x - 330.91 (44.8) R^2=1.000$
Hoogsteen:	$y = 20.53 (0.4) x^2 + 287.77 (6.2) x - 204.17 (22.3) R^2=1.000$

**Figure 7.** Dependence of the phosphate-phosphate interactions on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parenthesis) and the determination coefficient are displayed. The error bars in the Figure correspond to the standard errors in the average energy estimates.

## CONCLUSIONS

- 1) Double helices of DNAs based on the polyd(A:T) motif are intrinsically stable in dilute aqueous solution both in the parallel and antiparallel arrangements. The problems found in detecting parallel helices for linear DNAs seem to be related to the greater stability of the antiparallel helix, rather than to an intrinsic instability of the parallel structures.
- 2) The structural and molecular recognition characteristics of the three helices are quite different, even those of the WC and rWC helices. Though the nature of the essential dynamics of the three duplexes is similar, principal component analysis demonstrates that the canonical antiparallel helix is more flexible than the parallel ones.

- 3) The free energy of the different DNA helices increases linearly with the length of the duplex, while the solvation and intramolecular contributions display a second order polynomial dependence.
- 4) The antiparallel helix is the most stable helical structure. The difference in stability between the antiparallel and parallel helices increases linearly with the length of the oligonucleotide. The rWC helix is more stable than the H one, the difference being related to the better nucleation energy of the rWC helix.
- 5) Combination of nanosecond MD simulations, exhaustive analysis of oligonucleotides of different sizes, and MM/PB-SA calculations allow us to obtain a surprisingly accurate theoretical representation of the relative stability of different helical structures of DNA.

### ACKNOWLEDGEMENTS

This work has been supported by the Spanish DGICYT (PB98-1222 and PM99-0046), and the Centre de Supercomputació de Catalunya (CESCA).

### REFERENCES

- 1) Westhof, E.; Sundaralingam, M. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1852-1856.
- 2) Rippe, K.; Jovin, T. M. *Methods in Enzymology* **1992**, *211*, 199-220.
- 3) van de Sande, J. H.; Ramsing, N. B.; Germann, M. W.; Elhorst, W.; Kalisch, B. W.; Kitzing, E. V.; Pon, R. T.; Clegg, R. C.; Jovin, T. M. *Science* **1988**, *241*, 551-557.
- 4) Ramsing, N. B.; Rippe, K.; Jovin, T. M. *Biochemistry* **1989**, *28*, 9528-9535.
- 5) Germann, M. W.; Vogel, H. J.; Pon, R. T.; van de Sande, J. H. *Biochemistry* **1989**, *28*, 6220-6228.
- 6) Rippe, K.; Ramsing, N. B.; Jovin, T. M. *Biochemistry* **1989**, *28*, 9536-9541.
- 7) Rentzeperis, D.; Rippe, K.; Jovin, T. M.; Marky, L. A. *J. Am. Chem. Soc.* **1992**, *114*, 5926-5928.
- 8) Germann, M. W.; Kalisch, B. W.; Pon, R. T.; van de Sande, J. H. *Biochemistry* **1990**, *29*, 9426-9432.
- 9) Lavelle, L.; Fresco, J. R. *Nucleic Acids Res.* **1995**, *23*, 2692-2705.
- 10) Scaria, P. V.; Shafer, R. H. *Biochemistry* **1996**, *35*, 10985-10994.
- 11) Bhaumik, S. R.; Kandala, V. R.; Govil, G.; Liu, K.; Miles, H. T. *Nucleic Acids Res.* **1995**, *23*, 4116-4121.
- 12) Liu, K.; Miles, T.; Frazier, J.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 11802-11809.
- 13) Cubero, E.; Aviño, A.; de la Torre, B. G.; Frieden, M.; Eritja, R.; Luque, F. J.; González, C.; Orozco, M. *Submitted to publication* **2001**.
- 14) Kan, C. H.; Zhang, R.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126-131.

- 15) Sen, D.; Gilbert, D. *Biochemistry* **1992**, *31*, 65-70.
- 16) Kremer, E. J.; Pritchard, M.; Linch, M.; Yu, S.; Holman, K.; Baker, E.; Warren, S. T.; Schlessinger, D.; Sutherland, G. R.; Richards, R.I. *Science* **1991**, *252*, 1711-1714.
- 17) Tchurikov, N. A.; Ebralidze, A.K.; Georgiev, G. P. *EMBO J.* **1986**, *5*, 2341-2347.
- 18) Tchurikov, N. A.; Chernov, B. K.; Golova, Y. B.; Nechipurenko, Y. D. *FEBS Lett.* **1989**, *257*, 415-418.
- 19) Tchurikov, N. A.; Shchyolkina, A. K.; Borisova, O. F.; Chernov, B. K. *FEBS Lett.* **1992**, *297*, 233-236.
- 20) Kandimalla, E. R.; Agrawal, S.; Venkataraman, G.; Sasisekharan, V. *J. Am. Chem. Soc.* **1995**, *117*, 6416-6417.
- 21) Kandimalla, E. R.; Agrawal, S. *Biochemistry* **1996**, *35*, 15332-15339.
- 22) Aviñó, A.; Morales, J. C.; Frieden, M.; de la Torre, B. G.; Güimil-García, R.; Cubero, E.; Luque, F. J.; Orozco, M.; Azorín, F.; Eritja. *Bioorganic Med. Chem. Lett.* In Press **2001**.
- 23) Mohammadi, S.; Klement, R.; Shchyolkina, A. K.; Liquier, J.; Jovin, T. M.; Taillandier, E. *Biochemistry* **1998**, *37*, 16529-16537.
- 24) Pattabiraman, N. *Biopolymers* **1986**, *25*, 1603-1606.
- 25) Zhou, N.; Germann, M. W.; van de Sande, J. H.; Pattabiraman, N.; Vogel, H. J. *Biochemistry* **1993**, *32*, 646-656.
- 26) Hakoshima, T.; Fukui, T.; Ikehara, M.; Tomita, K. I. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 7309-7313.
- 27) Hashem, G. M.; Wen, J. D.; Do, Q.; Gray, D. M. *Nucleic Acids Res.* **1999**, *27*, 3371-3379.
- 28) Escudé, C.; Mohammadi, S.; Sun, J. S.; Nguyen, C-H.; Bisagni, E.; Liquier, J.; Taillandier, E.; Garestier, T.; Hélène, C. *Chemistry and Biology* **1996**, *3*, 57-65.
- 29) Germann, M. W., Kalisch, G. T.; van de Sande, J. H. *Biochemistry* **1998**, *37*, 12962-12970.
- 30) Ryjacek, F.; Engkvist, O.; Vacek, J.; Kratochvil, M.; Hobza, P. *J. Phys. Chem.* **2001**, *105*, 1197-1202.
- 31) Hobza, P.; Sponer, P. *Chem. Rev.* **1999**, *99*, 3247-3276.
- 32) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comp. Chem.* **1997**, *18*, 1136-1150.
- 33) Shields, G.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463-7469.
- 34) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504-1509.

- 35) Gelpí, J. L.; de la Cruz, X.; Kalko, S.; Barril, X.; Cirera, J.; Luque, F. J.; Orozco, M. *Submitted for publication* **2001**.
- 36) Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226-11233.
- 37) Shields, G.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895-5904.
- 38) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- 39) Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- 40) Ryckaert, J. P.; Ciccotti, G.; Berendsen, J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
- 41) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- 42) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845-862.
- 43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
- 44) 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.
- 45) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187-4225.
- 46) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473-486.
- 47) Bashford, D. In *Scientific Computing in Object-Oriented parallel environments*; Reynders, J. V. M., Tholburn, M. Eds.; Springer: Berlin, 1997; pp 233-240.
- 48) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978-1988.
- 49) Sherer, E.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981-5991.
- 50) Lavery, R.; Sklenar, J. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63-91.
- 51) Soliva, R.; Luque, F. J.; Alhambra, C.; Orozco, M. *J. Biomol. Struct. Dyn.* **1999**, *6*, 89-99.
- 52) According to our experience (55), entropy analysis performed using 1 ns trajectories are not expected to provide converged results, and accordingly a quantitative comparison between the entropy differences in Table 3 does not seem advisable.
- 53) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617-621.

- 54) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809-7817.
- 55) Harris, S.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc. In Press.* **2001**.
- 56) The standard errors for the total free energy are computed by adding the errors in internal and solvation terms as is they were not correlated. Accordingly, these values are an upper limit for the standard error in the averages.

#### 4.4.2 Parallel-stranded hairpins containing 8-aminopurines. Novel efficient probes for triple-helix formation

Anna Aviñó, Juan Carlos Morales, Miriam Frieden, Beatriz G. de la Torre,  
Ramón Güimil-García, **Elena Cubero**, F. Javier Luque, Modesto Orozco,  
Ferran Azorín & Ramón Eritja

*Bioorganic Med. Chem. Lett.* **2001**, *11*, 1761-1763

*(Esta página está intencionadamente en blanco)*



## Parallel-Stranded Hairpins Containing 8-Aminopurines. Novel Efficient Probes for Triple-Helix Formation

Anna Aviñó,<sup>a</sup> Juan Carlos Morales,<sup>b</sup> Miriam Frieden,<sup>a</sup> Beatriz G. de la Torre,<sup>b</sup>  
Ramon Güimil García,<sup>b</sup> Elena Cubero,<sup>c</sup> F. Javier Luque,<sup>c</sup> Modesto Orozco,<sup>c</sup>  
Ferran Azorín<sup>a</sup> and Ramon Eritja<sup>a,\*</sup>

<sup>a</sup>Cygene Spain S.L., Parc Científic de Barcelona, Baldri i Reixac, 10-12, E-08028 Barcelona, Spain

<sup>b</sup>Institut de Biologia Molecular de Barcelona, C.S.I.C., Jordi Girona 18-26, E-08034 Barcelona, Spain

<sup>c</sup>Universitat de Barcelona, Martí i Franquès 1-11, E-08028 Barcelona, Spain

Received 22 January 2001; revised 4 April 2001; accepted 24 April 2001

**Abstract**—We describe novel oligomers with a greater propensity to form triplexes than oligomers containing only natural bases. They consist of a polypyrimidine sequence linked head-to-head with a polypurine sequence carrying one or several 8-aminoadenine or 8-aminoguanines. The presence of 8-aminopurines also stabilised the parallel-stranded duplex structure. © 2001 Elsevier Science Ltd. All rights reserved.

Oligonucleotides bind in a sequence-specific manner to homopurine-homopyrimidine nucleic acid sequences to form triple helices.<sup>1</sup> This offers the possibility of designing DNA- and RNA-binding molecules, which may have several applications. A large effort has therefore been devoted to the design and preparation of modified oligonucleotides in order to enhance triple-helix binding stability.<sup>2</sup>

The synthesis of oligonucleotides containing 8-aminopurines has been described.<sup>3–7</sup> The introduction of an amino group at position 8 of adenine and guanine increases the stability of the triple helix due to the combined effect of the gain of one Hoogsteen purine-pyrimidine H-bond (Scheme 1),<sup>3,4,7</sup> and the propensity of the amino group to be integrated into the ‘spine of hydration’ located in the minor–major groove of the triplex structure.<sup>3,7</sup>

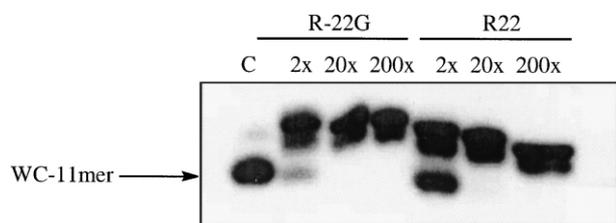
Although the preparation and binding properties of oligonucleotides containing 8-aminopurines have been described, natural oligonucleotides containing 8-aminopurines cannot bind double-stranded DNA sequences because the modified bases are purines that are in the target sequence rather than in the Hoogsteen strand

used for specific recognition. Here we describe the preparation and binding properties of oligonucleotides carrying 8-aminoadenine and 8-aminoguanine, connected head-to-head to the Hoogsteen pyrimidine strand. These modified oligonucleotides, which are a special class of parallel-stranded DNA, allow the specific recognition of single-stranded nucleic acids by binding to a polypyrimidine strand by triple-helix formation.

Oligonucleotide sequences (R-22: 5′GAA GGA GGA GA<sup>3′</sup>-(EG)<sub>6</sub>-3′TCT CCT CCT TC<sup>5′</sup>, R-22A: 5′GAA GGA<sup>N</sup> GGA<sup>N</sup> GA<sup>3′</sup>-(EG)<sub>6</sub>-3′TCT CCT CCT TC<sup>5′</sup>, R-22G: 5′GAA GG<sup>N</sup>A GG<sup>N</sup>A GA<sup>3′</sup>-(EG)<sub>6</sub>-3′TCT CCT CCT TC<sup>5′</sup> where A<sup>N</sup>, G<sup>N</sup> and (EG)<sub>6</sub> are 8-aminoadenine, 8-aminoguanine and hexaethyleneglycol) were prepared using phosphoramidite chemistry on an automatic DNA synthesiser. The parallel-stranded oligomers were prepared as follows.<sup>8,9</sup> First, the pyrimidine part was assembled using reversed C and T phosphoramidites and a reversed C-support. Then, a hexaethyleneglycol linker was added using a commercially available phosphoramidite. Finally, the purine part carrying the modified 8-aminopurines was assembled using standard phosphoramidites for the natural bases and the 8-aminopurine phosphoramidites. The phosphoramidites of 8-aminoadenine and 8-aminoguanine were prepared as described previously.<sup>3–7</sup>

\*Corresponding author. Tel.: +34-700-6145; fax: +34-93-204-5904; e-mail: recgma@cod.csic.es





**Figure 1.** Binding of parallel-stranded hairpins to polypyrimidine target WC-11mer by gel shift experiments.  $^{32}\text{P}$ -labelled DNA target and the natural (R-22) or modified (R-22G) parallel-stranded hairpins were separately annealed in a citric-phosphate buffer (pH 6) of 100 mM  $\text{Na}^+$ . Binding was started by mixing 10  $\mu\text{L}$  of a solution containing the radiolabelled DNA target (20 nM) and 10  $\mu\text{L}$  of a solution containing hairpin at different concentrations (40 nM, 400 nM and 4  $\mu\text{M}$ ). The resulting solution was incubated for 30 min at room temperature. Phytol loading buffer (5  $\mu\text{L}$ ) was added to the solution and the mixture was analysed by running the samples on 15% native polyacrylamide gels at 4°C.

**Table 2.** Energy values for the Watson–Crick and Hoogsteen interaction of adenine, guanine and their 8-amino derivatives and their complementary bases (thymine and cytosine)

	WC (thymine/cytosine)	Hoogsteen <sup>a</sup> (thymine/cytosine)
Adenine	–13.1	–14.0
8-Amino adenine	–12.8	–19.6
Guanine	–25.4	–31.9
8-Amino guanine	–24.7	–35.1

<sup>a</sup>For Hoogsteen pairs G-C and 8AG-C the cytosine is considered in its N3-H protonated form.

Binding of hairpins to target was also analysed by gel-shift experiments. The target was labelled with [ $\gamma$ - $^{32}\text{P}$ ]-ATP and polynucleotide kinase and increasing amounts of the hairpins were added to solution of the labelled target. After incubation at room temperature for 30 min, the mixtures were analysed by polyacrylamide gel electrophoresis (PAGE). The formation of the triplex was followed by the appearance of a major band with less mobility than the band corresponding to the target. This binding was specific because hairpins did not bind oligonucleotides without the target polypyrimidine sequence (data not shown).

Theoretical calculations on the stability of Hoogsteen and Watson–Crick pairing were performed using AMBER-99 force-field<sup>11</sup> and the standard B-type geometry for triplex.<sup>12</sup> Results (Table 2) demonstrate the stabilisation of parallel-stranded duplexes and triplexes by the 8-amino group, while the Watson–Crick pairing was not modified. Thus, theoretical results agree with a Hoogsteen pairing for the parallel-stranded hairpins, as

well as with the hypothesis that improved H-bonding is responsible for the 8-amino-induced stabilisation of the parallel duplex.

We described the triplex stabilisation properties of hairpins formed by a polypyrimidine part linked head-to-head with a polypurine sequence carrying several 8-aminopurines: 8-aminoadenines or 8-aminoguanines. These modified hairpins bind specifically to a pre-determined polypyrimidine target by forming a stable triplex that can be observed even at pH 7. The high degree of stabilisation obtained with the addition of several 8-aminopurines is especially relevant to the development of new applications based on triple-helix formation such as structural studies, DNA-based diagnostic tools and antigene therapy.

### Acknowledgements

The authors thank the Dirección General de Investigación Científica y Técnica (projects PB98-1222, PM99-0046 and BQU2000-0649), the Generalitat de Catalunya (project 2000-SGR-0018) and CyGene, Inc. for financial support.

### References and Notes

1. Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666.
2. Soyfer, V. N.; Potaman, V. N. *Triple Helical Nucleic Acids*; Springer: New York, 1996.
3. Güimil García, R.; Ferrer, E.; Macías, M. J.; Eritja, R.; Orozco, M. *Nucleic Acids Res.* **1999**, *27*, 1991.
4. Kawai, K.; Saito, I.; Sugiyama, H. *Tetrahedron Lett.* **1998**, *39*, 5221.
5. Rao, T. S.; Durland, R. H.; Revankar, G. R. *J. Heterocycl. Chem.* **1994**, *31*, 935.
6. Rieger, R. A.; Iden, C. R.; Gonikberg, E.; Johnson, F. *Nucleosides Nucleotides* **1999**, *18*, 73.
7. Soliva, R.; Güimil-García, R.; Blas, J. R.; Eritja, R.; Asensio, J. L.; González, C.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2000**, *28*, 4531.
8. Kandimalla, E. R.; Agrawal, S. *Biochemistry* **1996**, *35*, 15332.
9. Rippe, K.; Jovin, T. M. *Methods Enzymol.* **1992**, *211*, 199.
10. Germann, M. W.; Zhou, N.; van de Sande, J.; Vogel, H. J. *Methods Enzymol.* **1995**, *261*, 207.
11. Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
12. Soliva, R.; Loughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226.

*(Esta página está intencionadamente en blanco)*

4.4.3 Hoogsteen-based parallel-stranded duplexes of DNA. The effect of 8-amino derivatives

**Elena Cubero**, Anna Aviñó, Beatriz G. de la Torre, Miriam Frieden, Ramón Eritja,  
F. Javier Luque, Carlos González & Modesto Orozco  
Aceptado en *J. Am. Chem. Soc.* **2001**

*(Esta página está intencionadamente en blanco)*

---

# HOOGSTEEEN-BASED PARALLEL-STRANDED DUPLEXES OF DNA. THE EFFECT OF 8-AMINO DERIVATIVES

Elena Cubero<sup>1</sup>, Anna Aviñó<sup>2</sup>, Beatriz G. de la Torre<sup>3</sup>, Miriam Frieden<sup>2</sup>, Ramon Eritja<sup>3\*</sup>, F.Javier Luque<sup>4</sup> Carlos González<sup>5</sup> and Modesto Orozco<sup>1\*</sup>

The structure of parallel-stranded duplexes of DNA-containing mixture of guanines (G) and adenines (A) is studied by means of molecular dynamics (MD) simulation, as well as NMR and circular dichroism (CD) spectroscopy. Results demonstrate that the structure is based on the Hoogsteen motif rather than on the reverse Watson-Crick one. Molecular dynamics coupled to thermodynamic integration (MD/TI) calculations and melting experiments allowed us to determine the effect of 8-amino derivatives of A and G, and 8-amino-2'-deoxyinosine on the stability of parallel stranded duplexes. The large stabilization of the parallel stranded helix upon 8-amino substitution agrees with a Hoogsteen pairing, confirming MD, NMR and CD data, and suggests new methods to obtain DNA triplexes for antigene and antisense purposes.

---

<sup>1</sup> Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain

<sup>2</sup> Cygene Spain S.L., Parc Científic de Barcelona, Baldiri i Reixac 10-12, Barcelona 08028, Spain.

<sup>3</sup> Instituto de Biología Molecular de Barcelona, C.S.I.C., Jordi Girona 18-26, Barcelona 08034, Spain.

<sup>3</sup> Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda Diagonal s/n, Barcelona 08028, Spain

<sup>4</sup> Instituto de Estructura de la Materia, C.S.I.C., Serrano 119, Madrid 28006, Spain.

\* Corresponding authors.

## INTRODUCTION

The DNA can form a large range of helical structures including duplexes, triplexes and tetraplexes (1-5). The right-handed B-type duplex is the most common structure of DNA, but even now, decades after the discovery of the B-DNA (6), new double helical conformations of DNA are being described (7,8). This demonstrates that the DNA has a great flexibility and exhibits a large polymorphism depending on sequence, chemical modifications, or alterations in the DNA environment (1).

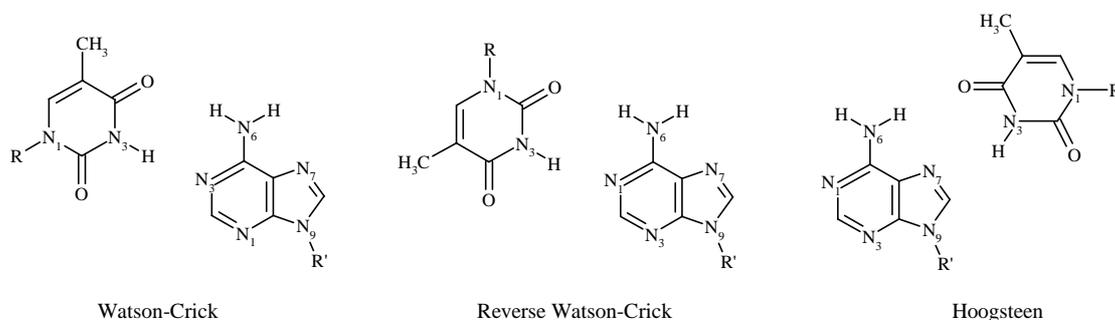
Most DNA duplexes, including the well-known B and A forms, are antiparallel (i.e., one strand runs 5'→3' and the other 3'→5'), but parallel arrangements have been found in both hairpins and linear DNAs (9-19). Sequences with propensity to form parallel DNAs have been found in specific chromosome regions (20-25), and could have an evolutionary role (26). Moreover, certain types of parallel-stranded DNA can be excellent templates for the formation of triplexes, which can be very useful for the development of antigene and antisense therapies, as well for biotechnological purposes (27-29).

Parallel DNA duplexes were firstly found in the crystal structure of a very short, mismatched DNA sequence intercalated by proflavine (30). Low resolution data of parallel stranded duplex were found for longer pieces of RNA of sequence poly d(A·U), where the position 2 of adenines was modified by addition of bulky groups (31). The first structural model of polymeric parallel stranded duplex DNA was derived by Pattabiraman, (32) who based on theoretical calculations designed a model for the parallel pairing of poly d(A·T) duplexes based on the reverse Watson-Crick motif. This model has been confirmed by low (9-15,33) and high resolution (34,35) experimental techniques on d(A·T) rich sequences.

The parallel stranded duplex model early described by Pattabiraman (32) and further refined by NMR data (12,33,34) shows a general structure not far from the canonical antiparallel B-type helix. The bases are mostly perpendicular to the helix axis, there are two equivalent grooves, sugar units present puckering in the *South* region, and the A·T pairings are reverse Watson-Crick (Figure 1). This structure - the parallel reverse-Watson-Crick (rWC) duplex- is the most stable conformation for parallel stranded helices rich in d(A·T) pairs, as demonstrated by Jovin and others using a variety of thermodynamic and spectroscopic techniques (9-15,33). The rWC double helix is less stable than comparable antiparallel helices, but can be found in hairpins and linear DNAs designed in such a way that the antiparallel d(A·T) arrangement is hindered. The presence of few d(G·C) steps in the rWC double helix might be tolerated, but destabilizes the duplex (14,33)

An alternative structure for parallel stranded duplexes based on the Hoogsteen (H) recognition mode is also possible (Figure 1). This would lead to a double helix (not yet described from a structural point of view), which might act as a template for triplex formation (27-29). Parallel stranded DNA duplexes based on the H pairing occur in duplexes where purines are modified at position 2, which prevents both Watson-Crick and reverse Watson-Crick pairings (31), or in duplexes rich in d(G·C) (or d(G·G)) pairs. These latter duplexes can exist at neutral pH, but are especially stable at low pH (16,27-

29,36-38) owing to the need to protonate the Hoogsteen cytosine (Figure 1). The stability of the duplex can be also enhanced by DNA-binding drugs such as benzopyridoindole derivatives (37). Finally, as shown by Lavelle and Fresco (16) and others (36), H-based parallel duplexes can be more stable than the canonical B-type antiparallel duplex in certain conditions.



**Figure 1.** Schematic representation of the Watson-Crick, Reverse Watson-Crick and Hoogsteen A·T pairings.

The whole of the available experimental data suggests that the structure of parallel stranded DNAs is quite flexible and can change from H to rWC motifs depending on sequence, pH, and the presence of drugs. Low pH and high content of d(G·C) pairs favor the H-based structure, while the rWC helix is favored in d(A·T) rich sequences and at neutral or basic pH. A particularly controversial issue (34,37) is the critical pH and G/A composition where the parallel duplex changes from one type of structure to the other.

In this paper we analyze the structure of parallel stranded duplexes in mixed d(A·T) and d(G·C) sequences using state-of-the-art theoretical calculations and spectroscopic techniques. We also analyze the ability of 8-amino derivatives to stabilize parallel duplexes that can be then used as templates for the formation of triple helices of DNA or DNA·RNA-DNA, which might have a large impact in biotechnological and pharmaceutical research.

## METHODS

**Molecular dynamics simulations.** We analyzed the stability of a 11-mer parallel DNA duplex with almost the same content of d(G·C) and d(A·T) pairs -d(5'-GAAGGAGGAGA-3)'·d(5'-CTTCCTCCTCT-3')- in water at room temperature when the base pairing corresponds to both rWC and H motifs. Two and three starting models were considered for H and rWC duplexes, respectively (Table 1). The two starting models for H-duplex were obtained by removing the pyrimidine Watson-Crick strand of an A- and B-type triplex (simulations H<sub>A</sub> and H<sub>B</sub>). The three starting models for rWC-duplex correspond to: i) the NMR model (35), ii) the canonical model reported by Pattabiraman (32), and iii) an equilibrated MD rWC d(A·T) duplex (see Table 1). These starting structures lead to simulations rWC<sub>1</sub>, rWC<sub>2</sub> and rWC<sub>3</sub> respectively. For comparison purposes an antiparallel B-type duplex of the same sequence was generated

using canonical structural parameters (39). In all cases the duplex was immersed in a box containing 2200-2700 water molecules and sodium ions were added to neutralize the system. Based on previous results (16,27,28,36,40-43) Hoogsteen cytosines were protonated. The hydrated duplexes were then optimized, thermalized and equilibrated for 130 ps using our standard multi-step protocol (44-46). All the systems were then subjected to 1-5 ns of unrestrained MD simulation at constant pressure (1 atm) and temperature (298K) using periodic boundary conditions and the particle-mesh Ewald method to account for long-range electrostatic effects (47,48). SHAKE (49) was used to maintain all the bonds at their equilibrium distances, which allowed us the use of a 2 fs time step for integration. AMBER-98/TIP3P (50-52) and previously developed parameters for protonated cytosines and 8-aminopurines (45,53-56) were used.

Pairing scheme	Starting structure	Length of simulation
Hoogsteen	Modeled from B-type triplex from refs 44-46	5 ns
Hoogsteen	Modeled from A-type triplex from refs 44-46	5 ns
Rev. Watson-Crick	Modeled from NMR data from ref 35	2 ns <sup>a</sup>
Rev. Watson-Crick	Modeled from theoretical model in ref 32	1 ns <sup>a</sup>
Rev. Watson-Crick	From an MD model derived from ref 32	5 ns
Watson-Crick <sup>c</sup>	From canonical model in ref 39	5 ns

**Table 1.** Summary of starting structures and simulation times used to MD analysis of parallel stranded duplexes. In all cases structures were modeled by substitution of the original sequence by the studied one prior to the optimization and equilibration process. In the case of rWC pairs the structures were modeled to show a double H-bond scheme as that suggested from FTIR data in reference 125.

<sup>a</sup> Simulation was stopped at that time since the structures of the helix was severely distorted.

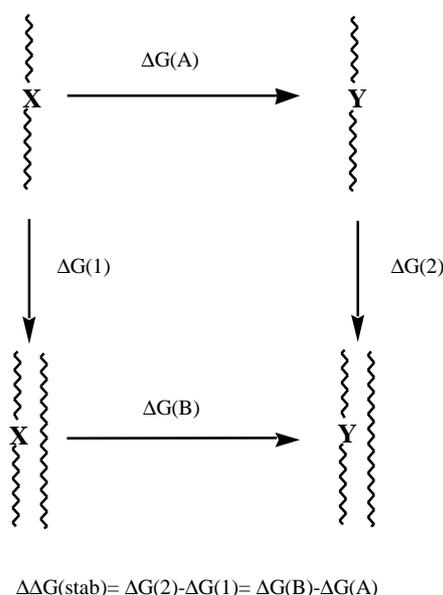
<sup>b</sup> The original rWC d(A·T) Pattabiraman structure (32) was equilibrated for 1 ns using MD, and the sequence was then modified to that of interest and then re-optimized and re-equilibrated.

<sup>c</sup> Antiparallel duplex

Geometrical analysis of the trajectories was performed using exclusively the central 9-mer duplex. For most studies the two trajectories of the H-based duplexes were averaged to obtain a better (10 ns) representation of the duplex. Analysis of DNA reactivity was carried out using our cMIP program (57). Curves (58) and *in-house* developed software was used for the structural analysis of the trajectories.

**Free energy calculations.** Thermodynamic integration technique coupled to molecular dynamics simulations (MD/TI) was used to analyze the effect of replacing 2'-deoxyadenosine, 2'-deoxyguanosine and 2'-deoxyinosine by their 8-amino derivatives on the stability of the d(5'-GAAGGAGGAGA-3')-d(5'-

CTTCCTCCTCT-3') parallel stranded duplex. For this purpose, mutations were performed between 8-amino-2'-deoxyadenosine and 2'-deoxyadenosine (8AA→A), 8-amino-2'-deoxyguanosine and 2'-deoxyguanosine (8AG→G) and 8-amino-2'-deoxyinosine and 2'-deoxyinosine (8AI→I) in both duplex and single stranded oligonucleotides. The change in stabilization free energy due to the 8AX→X mutation is determined using standard thermodynamic cycles (Figure 2).



**Figure 2.** Thermodynamic cycle used to compute stabilization of parallel stranded duplexes induced by the introduction of 8-amino derivatives.

MD/TI simulations were done considering only the H duplex due to the instability of the rWC duplex. The starting system in these calculations was defined as that obtained at the third nanosecond of the MD simulation duplex corresponding to the B trajectory of the H duplex. The 8-amino derivatives were then modeled at positions 5 (8AG and 8AI) or 6 (8AA) of the purine strand, and the resulting structures were further equilibrated for 0.5 ns to avoid any *bias* in the calculations. Two additional simulations were performed considering that the d(G·C)/d(I·C) pair at position 5 shows a wobble neutral pairing, d(G·C)<sub>w</sub>/d(I·C)<sub>w</sub>, instead of the normal protonated pair, d(G·C)<sup>+</sup>/d(I·C)<sup>+</sup>. In this case one extra sodium ion was added to the modeled system, which was then further equilibrated for 1ns. Based on our previous experience (59,60), the single strands were modeled as 5-mer oligonucleotides of sequences 5'-AGGAG-3', 5'-AGIAG-3' and 5'-GGAGG-3'.

Following our standard protocol (53-56,59,60) mutations were performed using 21 double-wide windows of 10 and 20 ps each, leading to trajectories of 420 or 820 ps. Free energy estimates were obtained using the first and second halves of each window, which allows us to have two independent estimates of the free energy change for every simulation. The values presented here correspond then to the average of 4 independent estimates, which allows us to estimate the statistical uncertainty of the

averages. All other technical details of MD/TI simulations are identical to those of MD calculations. We should note that simulations presented here correspond to more than 30 ns of unrestrained MD simulations of 11-mer H duplexes in water. This is to our knowledge one of the most extensive samplings of an anomalous DNA structure published to date.

All MD and MD/TI simulations were carried out using AMBER-5.1 computer program (61). All simulations were done in the supercomputers of the Centre de Supercomputació de Catalunya (CESCA) as well as in workstations in our laboratory.

**Preparation of oligomers containing 8-aminopurines.** Oligonucleotides were prepared on an automatic DNA synthesizer using standard and reversed 2-cyanoethyl phosphoramidites and the corresponding phosphoramidites of the 8-aminopurines. The phosphoramidite of protected 8-amino-2'-deoxyinosine was dissolved in dry dichloromethane to make a 0.1 M solution. The rest of the phosphoramidites was dissolved in dry acetonitrile (0.1 M solution). The phosphoramidite of the hexaethyleneglycol linker was obtained from commercial sources. Complementary oligonucleotides containing natural bases were also prepared using commercially available chemicals and following standard protocols. After the assembly of the sequences, oligonucleotide-supports were treated with 32% aqueous ammonia at 55 °C for 16 h except for oligonucleotides having 8-aminoguanine. In this case a 0.1 M 2-mercaptoethanol solution in 32% aqueous ammonia was used and the treatment was extended to 24 h at 55 °C. Ammonia solutions were concentrated to dryness and the products were purified by reverse-phase HPLC. Oligonucleotides were synthesized on 0.2 µmol scale and with the last DMT group at the 5' end (DMT on protocol) to help reverse-phase purification. All purified products presented a major peak, which was collected. Yield (OD units at 260 nm after HPLC purification, 0.2 µmol) was between 6-10 OD. HPLC conditions: HPLC solutions are as follows. Solvent A: 5% ACN in 100 mM triethylammonium acetate (pH 6.5) and solvent B: 70% ACN in 100 mM triethylammonium acetate pH 6.5. Columns: PRP-1 (Hamilton), 250 x 10 mm. Flow rate: 3 ml/min. A 30 min linear gradient from 10-80% B (DMT on), or a 30 min linear gradient from 0-50% B (DMT off).

**Melting experiments.** Melting experiments were performed as follows: Solutions of the hairpins and duplexes were dissolved in 1 M NaCl, 100 mM phosphate/citric acid buffer. The solutions were heated to 90 °C, allowed to cool slowly to room temperature and then samples were kept in the refrigerator overnight. UV absorption spectra and melting experiments (absorbance vs temperature) were recorded in 1 cm path-length cells using a spectrophotometer, which has a temperature controller with a programmed temperature increase of 0.5 °C/min. Melts were run on duplex concentration of 3-4 µM at 260 nm.

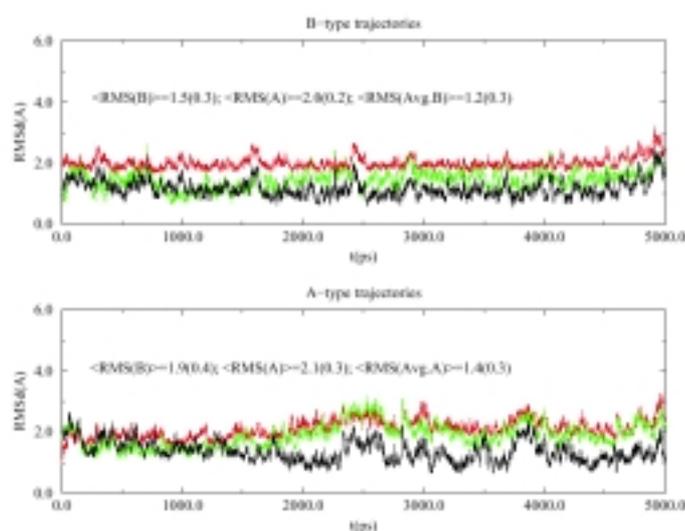
**Circular Dichroism.** Oligonucleotides were dissolved in 100 mM phosphate buffer pH 6.0, 50 mM sodium chloride and 10 mM magnesium chloride. The equimolar concentration of each strand was 4-5 µM. The solutions were heated at 90 °C, allowed to come slowly to room temperature and stored at 4 °C until CD measurement was carried out. The CD spectra were recorded on a Jasco J-720 spectropolarimeter attached to a Neslab RP-100 circulating water bath in 1 cm path-length quartz

cylindrical cells. Spectra were recorded at room temperature using a 10 nm/min scan speed, a spectral band width of 1 nm and a time constant of 4 s. CD melting curves were recorded at 280 nm using a heating rate of 20 °C/h and a scan speed of 100 nm/min. All the spectra were subtracted with the buffer blank, normalized to facilitate comparisons and noise-reduced using Microcal Origin 5.0 software.

**NMR spectroscopy.** A sample of the oligonucleotide d(3'-AGNGGNGGAAG-5'-(EG)<sub>6</sub>-5'-CTTCCTCCTCT-3') (N=8-amino-A) for NMR experiments was prepared in 250 µl of 9:1 H<sub>2</sub>O/D<sub>2</sub>O, 25mM sodium phosphate buffer, and 100 mM NaCl. The pH was adjusted by adding small amounts of concentrated HCl. The final oligonucleotide concentration was around 1 mM. Spectra were acquired in a Bruker AMX spectrometer operating at 600 MHz, and processed with the UXNMR software. Water suppression was performed by using a jump-and-return pulse sequence with a null excitation in the water signal (62). All experiments were acquired at 5 °C.

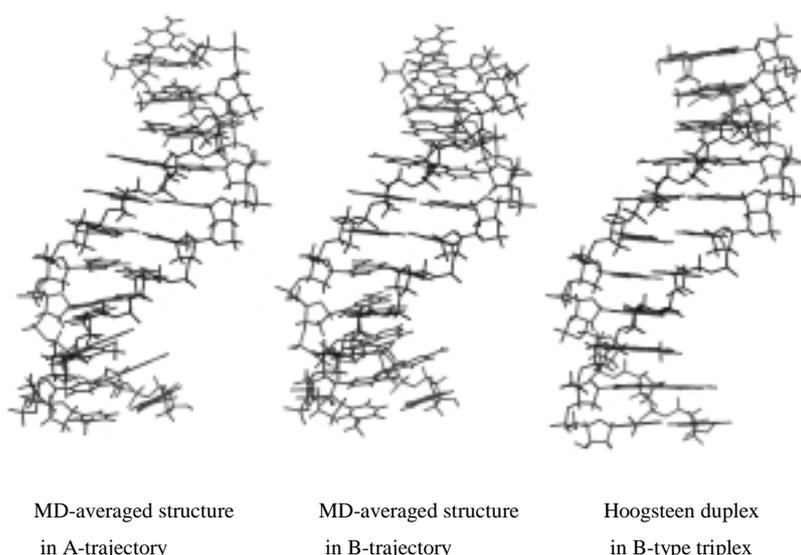
## RESULTS AND DISCUSSION

**Molecular dynamics simulations.** MD simulations of H duplexes show stable trajectories along the 5 ns simulation time (Figures 3 and 4), as noted in the average root mean square deviation (RMSd) between the trajectories and the respective MD-averaged conformations (1.4 and 1.6 Å for simulations H<sub>A</sub> and H<sub>B</sub>, respectively). The only noticeable distortions are a slight bend at the d(G·C) end and the existence of partial fraying events at the d(A·T) end. Note that similar features occur in the control antiparallel helix. It is also worth noting that the existence of two consecutive protonated pairs d(G·C)<sup>+</sup> does not introduce large structural alterations in the helix, thus confirming recent MD simulations of triple helices (54) and in agreement with NMR data (63-70).



**Figure 3.** RMS deviation (in Å) between the B- (TOP) and A- (BOTTOM) trajectories and different reference structures: Black: the MD-averaged structure of the trajectory, Green: the conformation of the Hoogsteen strands of a B-type triplex, and Red: the conformation of the Hoogsteen strands of an A-type triplex.

The two simulations, which started from different H-based duplex models, are reasonably converged and sample similar regions of the conformational space. This is noted in the RMSd between each trajectory and the MD-averaged conformation of the other: 1.9 Å (B trajectory with respect to the average structure in simulation H<sub>A</sub>) and 2.1 Å (A trajectory with respect to the average structure in simulation H<sub>B</sub>). Both trajectories sample conformational regions close to the conformation of Hoogsteen strands in a triplex DNA (Figure 4). Therefore, MD simulations suggest that the structure of the Hoogsteen strands of a triplex is not largely distorted when the pyrimidine Watson-Crick strand is removed. Thus, the RMSd between the two trajectories and the starting model in simulation H<sub>B</sub> (taken directly from a B-type triplex DNA, ref 45) is 1.4 and 1.8 Å in simulations H<sub>B</sub> and H<sub>A</sub> respectively. The RMSd is slightly larger with respect to the Hoogsteen strands of the starting model in simulation H<sub>A</sub> (an A-type triplex): 2.0 Å(-H<sub>A</sub>) and 2.1 Å(-H<sub>B</sub>).



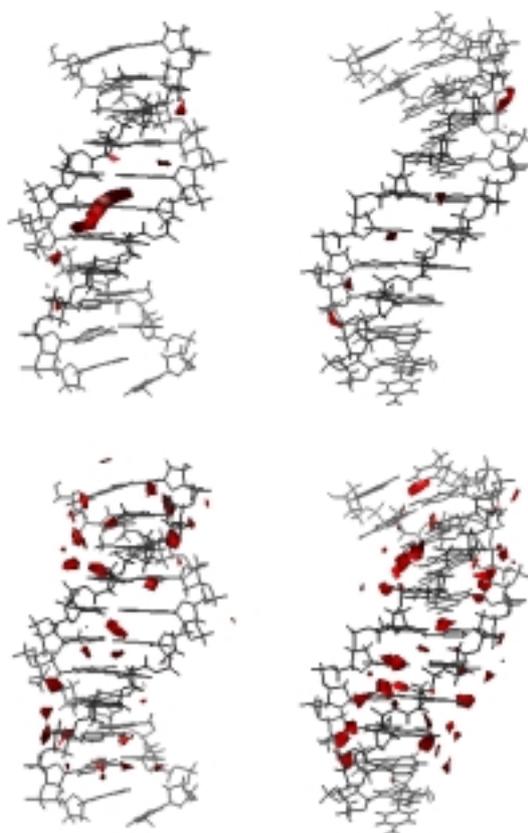
**Figure 4.** MD-averaged structures of the Hoogsteen duplexes obtained in the A- and B- trajectories. The conformation of the Hoogsteen duplex in a B-type triplex is displayed for comparison.

In contrast to these results, the simulations of rWC duplexes starting from the high resolution NMR (35) or the canonical (32) model (simulations rWC<sub>1</sub> and rWC<sub>2</sub>) diverge very quickly (Figure 5), despite of all the efforts made to reinforce the equilibration of the system and the pairing between bases. The structures are heavily distorted in less than 1 ns (Figures 5 and 6), and the helix was completely lost when the simulation was stopped (after 1 or 2 ns). The third simulation (rWC<sub>3</sub>), which started from a model derived from a previously 1 ns equilibrated trajectory of a d(A·T) rWC duplex, was stable for a longer period, but the helix was also largely distorted after the 5 ns simulation time (Figures 5 and 6). Analysis of the trajectories suggest that the amino repulsion between G and C is the main factor that causes the helix destabilization, despite our efforts to reduce the amino repulsion by promoting a wobble d(G·C) pairing (33).



based conformation seems stable during all the simulation time. Therefore, the results support the existence of H-based motifs for parallel stranded duplexes in DNAs with similar population of d(A·T) and d(G·C) pairs, and that the rWC helix is not stable when there is a high content of d(G·C) pairs.

The stability of the H-based simulations allows us to analyze the structure of a H-based parallel stranded duplex for the first time. As noted above, the helix is similar to the structure of Hoogsteen strands in a DNA triplex. The average twist is  $31^\circ$ , and the rise is  $3.4 \text{ \AA}$ . The bases are quite perpendicular to the helix axis. The sugars are in the *South* and *South-East* regions, having an average phase angle of 124 degrees, as found experimentally for rWC parallel stranded duplexes (12,34,35) and triplexes (44,45,53-56,62-69). There is a narrow groove (denoted “minor” in the following) corresponding to the minor part of the major groove in DNA triplexes (44), and a wide groove (denoted here “major”) corresponding to both the minor groove and the major part of the major groove of a DNA triplex (44) (Figure 4). The shortest P-P average distance along the two grooves is around  $9(\pm 0.6)$  and  $25(\pm 2) \text{ \AA}$  for the “minor” and “major” grooves. There are then major differences with rWC duplexes, where two equivalent grooves were found (32,35).



**Figure 7.** Classical Molecular Interaction Potentials (cMIP; TOP) and solvation maps (BOTTOM) for the canonical antiparallel duplex (LEFT) and Hoogsteen parallel stranded duplex (RIGHT). CMIP contours correspond to interaction energy of  $-5.5 \text{ kcal/mol}$  ( $O^-$  was used as probe). Solvation contours correspond to a density of  $2 \text{ g/ml}$ . For parallel duplexes cMIP and solvation maps were determined averaging over the A- and B- trajectories simultaneously.

The Molecular Interaction Potential maps (cMIP; Figure 7) allowed us to trace the regions where the DNA has a strong propensity to interact with small cationic probes (44-46,71). As expected from our previous studies on DNA triplexes (44,45), the “minor” groove is the most reactive region. It is worth noting that the ability of the H duplex to interact with cationic probes is not different to that of a B-type antiparallel duplex with the same sequence, despite the fact that all Hoogsteen cytosines are protonated in the H duplex. It is clear that the short P-P distance in H duplexes creates a strong negative potential in the vicinities of the Hoogsteen cytosines, thus screening their positive charge.

The H duplex is very well hydrated, as shown in the solvation contours represented in Figure 7. The largest apparent density of water is found in the minor groove, which is wide enough to allow the insertion of a chain of ordered waters. There are also regions of large (more than 2 g/ml) water density in the vicinities of the phosphate groups in the major groove. Interestingly, the apparent water densities around the H duplex and the reference antiparallel helix are very similar, thus confirming the findings obtained from cMIP calculations.

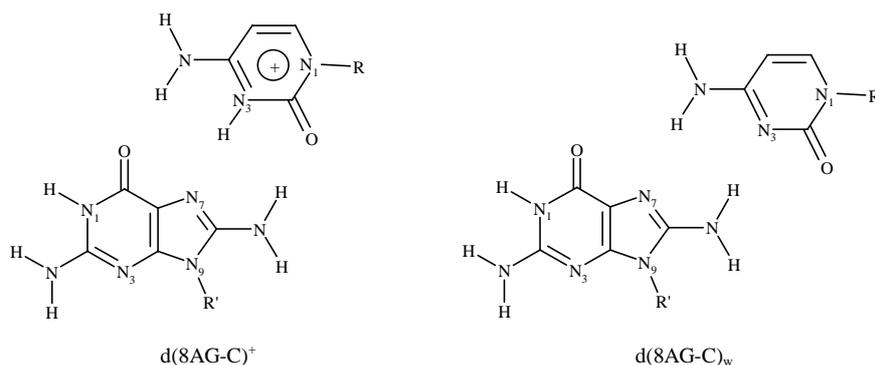
In summary, the antiparallel H duplex is a new structure shares many characteristics with DNA triplexes, but that also exhibits a series of unique reactive characteristics derived mainly from the existence of two very different grooves.

**Free energy calculations.** In a series of previous papers (53-56,72) we reported the design, synthesis and evaluation of a series of 8-amino derivatives of purine bases. These molecules strongly stabilize the DNA triplex (53-56,72), which was related, among others factors, to an extra H-bond between the 8-amino group of the purine and the carbonyl group of Hoogsteen cytosines or thymines (53,54,55). We also found (42) that the 8-amino group promotes a strong destabilization of the Watson-Crick pairing, at least for d(G-C) and d(I-C) pairs. Accordingly, we could expect that the presence of 8-amino groups should destabilize the rWC duplex and increase the stability of the H duplex. It is worth noting that the stability of the H duplex is crucial for the use of parallel stranded duplexes as templates for triplex formation. MD/TI calculations were performed only in the H duplex because the instability of the rWC duplex precludes any TI calculation. As found in previous simulations for related systems (53-56,59,60), the mutation profiles are smooth, without any apparent discontinuity, which could signal the existence of hysteresis. The standard errors in free energy estimates are 0.2-0.3 kcal/mol, thus indicating a good convergence in the results (Table 2).

The H duplex is stabilized by around 2.7 kcal/mol by the A $\rightarrow$ 8AA mutation (Table 2), a value similar to that found previously (55) using less rigorous simulation protocols for poly d(A·T·T) triplex. The mutation G $\rightarrow$ 8AG in a d(G·C)<sup>+</sup> motif increases the stability of the H duplex by around 1 kcal/mol (Table 2), while the I $\rightarrow$ 8AI mutation in the d(I·C)<sup>+</sup> motif increases the stability by around 1.4 kcal/mol (Table 2). These two latter values also agree with previous estimates in DNA triplexes (53,54). Keeping in mind the similarity between the triplex and the H duplex, this agreement gives confidence to our simulations.

Mutation	Complementary pyrimidine	$\Delta\Delta G_{\text{stab}}$ (kcal/mol)
G→8AG	C <sup>+</sup>	-1.4±0.2
G→8AG	C	-3.1±0.3
I→8AI	C <sup>+</sup>	-0.9±0.3
I→8AI	C	-3.2±0.3
A→8AA	T	-2.7±0.3

**Table 2.** MD/TI estimates of stabilization ( $\Delta\Delta G_{\text{stab}}$  and standard errors in kcal/mol) of parallel-stranded duplexes induced by 8-amino derivatives. For d(G-C) and d(I-C) motifs the simulation was performed considering two ionization states of the Hoogsteen cytosine. Calculations were carried out always using the sequence d(GAAGXAGGAG); where X is the base which is mutated.



**Figure 8.** Representation of protonated and wobble Hoogsteen 8AG-C dimers.

The preceding results clearly point out a strong stabilization of the H duplex upon introduction of 8-amino purines, and suggest that these molecules can help stabilize hairpins based on the parallel H duplex. We were, however, concerned by the fact that the G→8AG mutation stabilizes less the H duplex than the A→8AA mutation, since this finding, which agrees with previous calculations in triplexes (53,54), do not agree with melting experiments on H-hairpins (see below). This suggests that when 8AG (or 8AI) is present, the Hoogsteen recognition might not necessarily be the  $d(8AG-C)^+$  motif, but can be a wobble pair  $d(8AG-C)_w$  (see Figure 8 and discussion in ref. 56). Because the  $d(G/I-C)^+ \rightarrow d(8AG/8AI-C)_w$  mutation is technically very difficult owing to the annihilation of a net charge, we investigated by means of indirect evidences the potential role of  $d(8AG-C)_w$  and  $d(8AI-C)_w$  motifs by doing the mutations G→8AG and I→8AI in the presence of a neutral cytosine in the complementary Hoogsteen position (the rest of Hoogsteen cytosines was protonated). The results (see Table 2) suggest that the presence of 8-amino derivatives strongly stabilizes (3.1 and 3.2 kcal/mol for I and G respectively) the wobble pairing.

Note that this free energy difference is 0.5 kcal/mol larger than that found in the A $\rightarrow$ 8AA mutation and more than 2 kcal/mol larger than the stabilization due to the same mutation when the Hoogsteen cytosine is protonated. According to these results, it can be hypothesized that the presence of 8AG and 8AI favors the existence of neutral Hoogsteen motifs instead of the protonated ones (see below).

**Structure of the oligonucleotide derivatives.** In order to check MD and MD/TI-derived hypothesis, several parallel-stranded DNA hairpins carrying 8-aminoadenine (8AA=A<sup>N</sup>), 8-aminoguanine (8AG=G<sup>N</sup>) and 8-aminohypoxanthine (8AI=I<sup>N</sup>) were prepared. The sequences of the oligonucleotides are shown in Figure 9 (73).

R-22	5' GAAGGAGGAGA <sup>3'</sup> -(EG) <sub>6</sub> - <sup>3'</sup> TCTCCTCCTTC <sup>5'</sup>
R-22A	5' GAAGGA <sup>N</sup> GGA <sup>N</sup> GA <sup>3'</sup> -(EG) <sub>6</sub> - <sup>3'</sup> TCTCCTCCTTC <sup>5'</sup>
R-22G	5' GAAGG <sup>N</sup> AGG <sup>N</sup> AGA <sup>3'</sup> -(EG) <sub>6</sub> - <sup>3'</sup> TCTCCTCCTTC <sup>5'</sup>
R-22I	5' GAAGI <sup>N</sup> AGI <sup>N</sup> AGA <sup>3'</sup> -(EG) <sub>6</sub> - <sup>3'</sup> TCTCCTCCTTC <sup>5'</sup>
B-22	3' AGAGGAGGAAG <sup>5'</sup> -(EG) <sub>6</sub> - <sup>5'</sup> CTTCCTCCTCT <sup>3'</sup>
B-22A	3' AGA <sup>N</sup> GGA <sup>N</sup> GGAAG <sup>5'</sup> -(EG) <sub>6</sub> - <sup>5'</sup> CTTCCTCCTCT <sup>3'</sup>
B-22G	3' AGAG <sup>N</sup> GAG <sup>N</sup> GAAG <sup>5'</sup> -(EG) <sub>6</sub> - <sup>5'</sup> CTTCCTCCTCT <sup>3'</sup>
B-22AG	3' AGA <sup>N</sup> G <sup>N</sup> GA <sup>N</sup> GGAAG <sup>5'</sup> -(EG) <sub>6</sub> - <sup>5'</sup> CTTCCTCCTCT <sup>3'</sup>
B-AT	3' AAAAAAAAAA <sup>5'</sup> -(EG) <sub>6</sub> - <sup>5'</sup> TTTTTTTTTTT <sup>3'</sup>
B-22Acontrol	3' AGA <sup>N</sup> GGA <sup>N</sup> GGAAG <sup>5'</sup> -(EG) <sub>6</sub> - <sup>5'</sup> CCCCTTTTTTT <sup>3'</sup>
D1	5' GAAGGAGGAGA <sup>3'</sup> · 5' TCTCCTCCTTC <sup>3'</sup>
D2	5' GAAGGAGGAGA <sup>3'</sup> · 5' TCCTCCT <sup>3'</sup>

**Figure 9.** Sequences of parallel-stranded hairpins carrying 8-aminopurines prepared in the present work. A<sup>N</sup>: 8-aminoadenine, G<sup>N</sup>: 8-aminoguanine, I<sup>N</sup>: 8-aminohypoxanthine, (EG)<sub>6</sub>: hexaethyleneglycol linker. Two antiparallel duplexes used as control are also displayed.

The first group of oligomers are parallel-stranded hairpins connected through their 3' ends with an hexaethyleneglycol linker [(EG)<sub>6</sub>]. In the oligonucleotide R-22A two adenines are substituted by two 8-aminoadenines (A<sup>N</sup>), in the oligonucleotide R-22G two guanines are substituted by two 8-aminoguanines (G<sup>N</sup>), and in the oligonucleotide R-22I two guanines are substituted by two 8-aminohypoxanthines (I<sup>N</sup>). The oligonucleotide R-22 contains only the natural bases without modification.

The second group of oligomers (B-22, B-22A, B-22G) are similar in composition than the previous oligomers but the polypurine and the polypyrimidine parts are connected through their 5' ends with an hexaethyleneglycol linker [(EG)<sub>6</sub>]. In addition an oligomer having two 8-aminoguanines and two 8-aminoadenines was prepared (B-22AG) to test whether the stabilizing properties both 8-aminopurines are additive. A parallel-stranded hairpin that has only d(A·T) base pairs (B-AT) was prepared. Finally, a control hairpin (B-22Acontrol) with the same purine sequence than B22A but a non-complementary pyrimidine sequence was also prepared.

**Oligonucleotide synthesis.** Oligonucleotide sequences containing 8-aminopurines were prepared using phosphoramidite chemistry on an automatic DNA synthesizer. The parallel-stranded oligomers were prepared following previously described protocols (9,10,28). The preparation of 3'-3' linked hairpins (R-22, R-22A, R-22G and R-22I) was performed in three parts: First, the preparation of the pyrimidine part, using reversed C and T phosphoramidites and reversed C-support (linked to the support through the 5' end). After the assembly of the pyrimidine part, a hexaethyleneglycol linker was added using a commercially available phosphoramidite. Finally, the purine part carrying the modified 8-aminopurines was assembled using standard phosphoramidites for the natural bases and the 8-aminopurine phosphoramidites. The phosphoramidites of 8-aminoadenine, 8-aminoguanine and 8-aminohyxanthine were prepared as described previously (54,55,72,74-76). For the preparation of 5'-5' linked hairpins (B-22, B-22A, B-22G B-22AG, B-AT and B-22A control) a similar approach was used. In this case, the purine part was assembled first, followed by the hexaethyleneglycol. The pyrimidine part was the last part to be assembled using reversed phosphoramidites.

After the assembly of the sequences, supports were treated with concentrated ammonia. Oligonucleotides carrying 8-aminoguanine were treated with 0.1 M 2-mercaptoethanol in concentrated ammonia (76). After deprotection, the products were purified by reverse-phase HPLC using the DMT-on and DMT-off protocols.

**Melting experiments.** The relative stability of parallel-stranded hairpins was measured spectrophotometrically at different pHs (pH 4.6-7.0). In most cases one single transition was observed with a hyperchromicity around 15% at acidic pH and 10% at neutral pH that was assigned to the denaturation of the parallel-stranded hairpin. In Table 3 melting temperatures of the hairpins having 3'-3' linkages are shown.

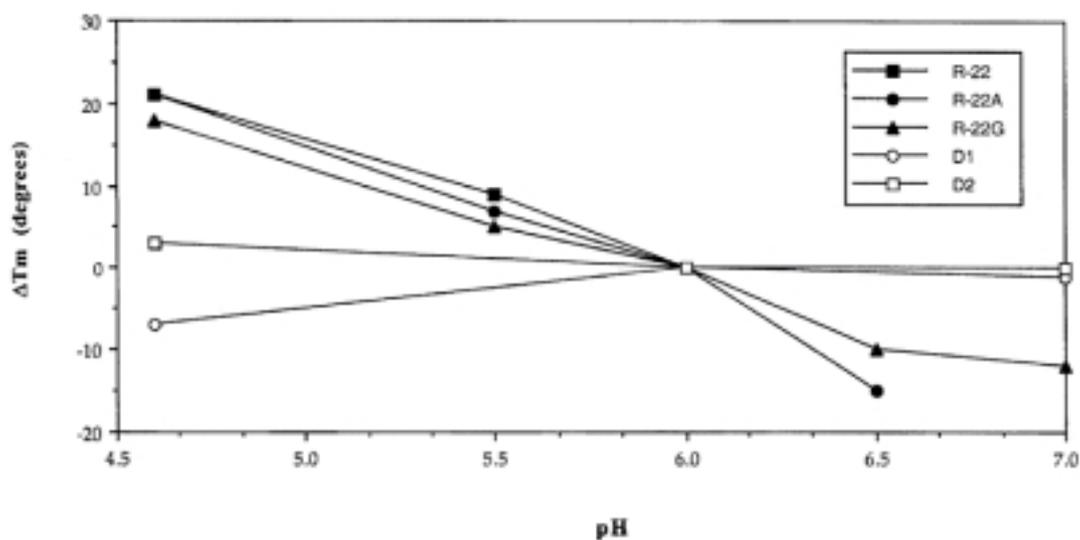
Hairpin	pH=4.6	pH 5.5	pH 6.0	pH 6.5	pH 7.0
R-22	46	34	25	--	--
R-22A	64	50	43	28	--
R-22G	68	55	50	40	39
R-22I	52	42	34	25	23

**Table 3:** Melting temperatures\* (°C) for the parallel-stranded hairpins having 3'-3' linkages.

\*1 M NaCl, 100 mM sodium phosphate/ citric acid buffer.

When the hairpin is formed by natural bases (R-22), a clear transition is observed at pH 4.6 and pH 6.0 but no transition was observed at pH higher than 6.0. Melting temperatures are pH-dependent. At lower pH melting temperatures are higher than at pH 7.0. These results are consistent with a Hoogsteen base pairing in where C has to be protonated to form the Hoogsteen base pair with G (i.e. an H-type duplex is supported). This profile of pH dependence cannot be explained for a reverse-Watson Crick

parallel duplex, and is also inconsistent with the formation of short antiparallel duplexes (like a 7 mer duplex d(-AGGAGGA)-d(-TCCTCCT-), which can be formed with the central part of oligonucleotide. To verify the later point we synthesized a 11 mer antiparallel duplexes of sequence d(GAAGGAGGAGA)-d(TCTCCTCCTTC) (D1), and d(GAAGGAGGAGA)-d(TCCTCCT) (D2) and measured their melting temperature at pH 4.5, 6.0 and 7.0. The profiles of pH dependence with the temperature found for both antiparallel duplexes are compared in Figure 10 with those found for R-22 and B-22. It is clear that  $T_m$  vs pH profiles strongly support that the antiparallel duplex is not significantly populated in our experiments.



**Figure 10.** Dependence of  $T_m$  with pH for R-22, B-22 and two antiparallel duplexes D1 and D2 (see Figure 9 for nomenclature).

The substitution of two A by two 8AA stabilises the parallel-stranded structure as seen by the higher melting temperatures at pH 4.6 and 6.0 ( $\Delta T_m$  16-18 °C) and the observation of a transition at pH 6.5. The substitution of two G by two 8AG increases even higher the melting temperatures of the hairpins. The differences on melting temperatures with respect of B-22 were between 21 and 25 °C. Also, it is possible to observe a transition at pH 7.0 and 6.5. The substitution of two G by two 8AI stabilizes the parallel-stranded structure, but this stabilization is of small intensity ( $\Delta T_m$  6-9 °C at pH 4.6-6.0). Finally, it is worth to note that the melting temperatures of hairpins having 8AG and 8AI are not decreasing so fast at neutral pH. This indicates that these hairpins are not as dependent as the other hairpins to protonation of C probably due to the extra hydrogen bond between the 8-amino group of the 8-aminopurines and the 2-keto group of C (53-55).

As noted above, in addition to the hairpins linked by 3'-3' bonds (R-22 derivatives), we prepared hairpins linked by 5'-5' bonds (B-22 derivatives). Table 4 shows the melting temperatures of these hairpins at different pHs.

Hairpin	pH 4.6	pH 5.5	pH 6.0	pH 6.5	pH 7.0
B-22	57	35	25	--	--
B-22A	61	47	38	23	--
B-22G	65	54	44	30	21
B-22AG	72	62	52	43	39

**Table 4:** Melting temperatures\* (°C) for the parallel-stranded hairpins having 5'-5' linkages.

\*1M NaCl, 100 mM sodium phosphate / citric acid buffer

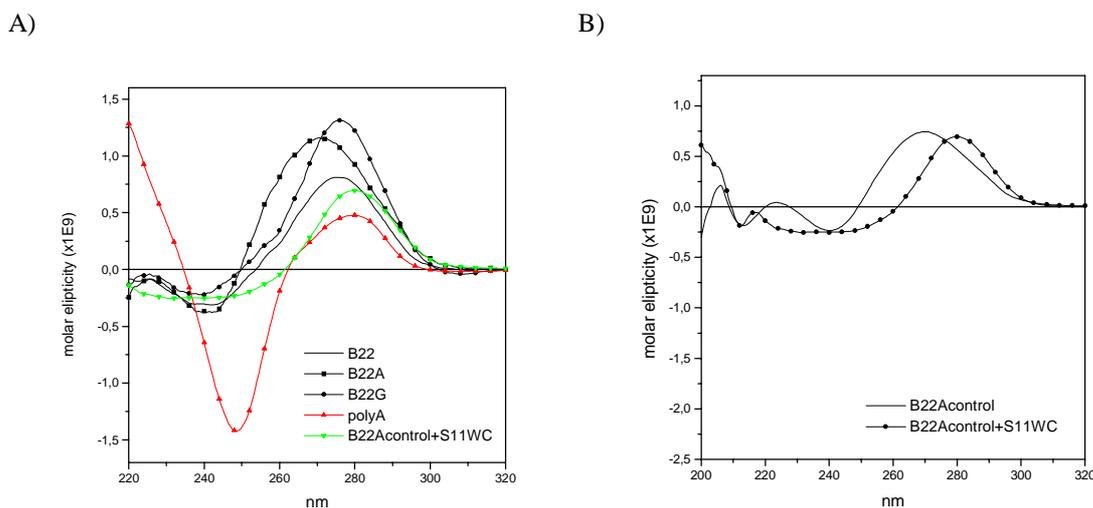
Results are similar to that described above with hairpins having 3'-3' linkages. Substitution of A or G by the corresponding 8-aminopurine derivative, induces a strong stabilization of the hairpin seen as a higher  $T_m$  at acidic pH and the observation of transitions at neutral pH that are not possible to be observed with hairpins having only natural bases. It is important to notice also that the addition of both 8AA and 8AG in the same oligonucleotide (B-22AG) has additive effects. For example at pH 6.0, the presence of two 8AA gives an increase on the  $T_m$  of 13 °C, two 8AG gives an increase of 19 °C and the addition of both two 8AA and two 8AG gives an increase of 27 °C. The low dependence of melting temperatures by the pH observed at neutral pH with hairpins having 8AG (R-22G) and 8AI (R-22I) are not observed with the hairpin B-22G, but it is observed with the hairpin having both 8-aminopurines (B-22AG). Parallel hairpins containing only A·T pairs (B-AT) had the same melting temperature ( $T_m=42^\circ\text{C}$ ) from pH=5.5 to 7.0. Control hairpin (B-22A control) had no transition at any pH.

It is worth noting the excellent agreement between MD/TI predictions derived from the assumption of an H-type parallel duplex and experimental measures. The large stabilization predicted theoretically for the amino-groups is found experimentally in increases in  $T_m$  of almost 10 degrees per substitution. Interestingly, the greater stability obtained for the G→8AG mutation compared with that obtained by the A→8AA mutation, and the smaller dependence on pH of the stability of duplexes containing 8AG suggest that neutral wobble pairing might play a key role in parallel duplexes containing d(8AG·C) pairs. Finally, the small stabilization obtained for the G→8AI mutation is the result of the balance between the stabilization of the H-duplex induced by the I→8AI mutation, and the destabilization induced by the G→I change (54).

As noted in a previous work (54), the 8-amino group destabilizes the Watson-Crick pairing for G and I, and are expected then to destabilize the reverse-Watson Crick pairing. Accordingly, the stabilization in the duplex structure found experimentally can be understood only considering that the hairpins studied here have a Hoogsteen, and not a reverse Watson-Crick secondary structure. Note also that the change in stability of the duplex induced by the G→8AG or A→8AA substitutions also argue

strongly against the existence of significant amounts of a 7-mer antiparallel duplex. Thus, the changes of 2 G (positions 5 and 8) by two 8AG lead to a decrease of 7 degrees in  $T_m$  for the two antiparallel duplexes used as controls d(GAAGGAGGAGA)-d(TCTCCTCCTC), and d(GAAGGAGGAGA)-d(TCCTCCT), while for R-22 and B-22 the same changes induced increases more than 21 degrees in  $T_m$ .

**Circular dichroism.** In order to obtain information on the structure of the hairpins, circular dichroism spectra were measured. Figure 11A shows the CD spectra of hairpins (B-22, B-22A and B-22G) the parallel-stranded hairpin with d(A·T) base pairs (B-AT). As an additional control we introduced another a modified B22 hairpin (B-22Acontrol), where the sequence of the pyrimidine strand is random, to guarantee that no parallel duplex can be formed. As noted in Figure 11B, this hairpin has not structure, but generates an antiparallel duplex if a suitable single stranded oligonucleotidic strand (S11WC) is added (B22Acontrol +S11WC).

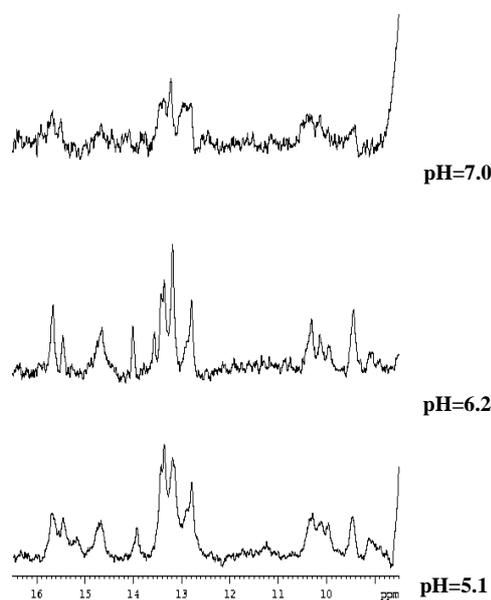


**Figure 11.** (A) CD spectra of hairpins B-22, B-22G, B-22A, B-AT (polyA), and an antiparallel duplex formed by B22Acontrol (B-22 hairpin where the sequence of the pyrimidine strand is random) and a suitable single stranded oligonucleotide (S11WC). (B) CD spectra of B-22Acontrol alone and after addition of the antiparallel complementary pyrimidine strand (0.1M Sodium phosphate pH 6.0, 50 mM NaCl, 10 mM MgCl<sub>2</sub>).

The CD spectra (see Figure 11A) of hairpins B-22, B-22A and B-22G have similar shape and they are very different from B-AT, and from the antiparallel duplex (B22Acontrol+S11WC; see also Figure 11B). The CD spectrum of this hairpin B-AT has a strong minimum at 248 nm and a smaller minimum at 206 nm and two maxima at 218 and 280 nm. This spectrum is similar to that described previously for A-T rich parallel-stranded DNA (9,10,28) that is considered a model for reverse Watson-Crick pairing. The CD spectra of B-22, B-22A and B-22G have a strong maximum between 270-290 nm and two minima: one at 242 nm and a second, more intense minima at around 212 nm. The minimum around 212 and the maximum around 280 are more intense in the hairpins containing 8-aminopurines (B-

22A and B-22G) suggesting a more stable structure. This type of spectra is characteristic of DNA triplexes (19). In summary, CD spectra demonstrate that the hairpins studied here, which contains a mixture of A(8AA)-T and G(8AG/8AI)-C steps have a Hoogsteen-type structure, and are not a reverse-Watson Crick parallel nor Watson-Crick antiparallel duplexes.

**NMR spectra.** The imino region of one-dimensional  $^1\text{H}$ -NMR spectra of the DNA hairpin d(3'-AGA<sup>N</sup>GGA<sup>N</sup>GGAAG-5'-(EG)<sub>6</sub>-5'-CTTCCTCCTCT-3') at three different pHs is shown in Figure 12. Unfortunately, the broad signals observed (due probably to the formation of Hoogsteen parallel intermolecular duplexes at the concentration of NMR experiment) prevented the acquisition of high quality two-dimensional spectra, and, therefore, the sequential assignments could not be done. However, the presence of imino signals between 14.5 and 16.0 ppm clearly indicates that some cytosines are protonated. Also, the signals around 10 ppm correspond with amino protons of cytosines forming Hoogsteen base pairs. Most probably, the resonances around 13 ppm are due to imino protons of thymines Hoogsteen. Since the chemical shifts of the exchangeable protons in reversed Watson-Crick base pairs are very similar to those observed in canonical antiparallel duplexes (34,77), this kind of base pairing can be ruled out. Finally, it is worth noting that most of the features of the exchangeable proton spectra can be still observed at neutral pH, suggesting a notable stability of the parallel duplex at neutral pH.



**Figure 12.** Exchangeable proton region of the NMR spectra of: d(3'-AGA<sup>N</sup>GGA<sup>N</sup>GGAAG-5'-(EG)<sub>6</sub>-5'-CTTCCTCCTCT-3') at T=5 °C.

Overall, NMR experiments confirm MD, MD/TI and CD results, and demonstrate that the parallel stranded duplexes studied here are stable, and that show a Hoogsteen-type H-bonding pattern similar to that of DNA triplexes. The reverse Watson-Crick model of the parallel stranded duplex, or the standard antiparallel duplex is ruled out.

## CONCLUSIONS

Very extended molecular dynamics simulations fail to provide stable helical structures for sequences containing similar number of d(A·T) and d(G·C) pairs arranged in the reverse-Watson Crick structure. On the contrary, stable trajectories are found if a Hoogsteen pairing is assumed. The structures obtained in these trajectories allowed us to characterize, for the first time, the structure of a H-type parallel duplex, whose overall conformation is close to that displayed by the Hoogsteen-strands of a DNA triplex. CD spectra support this hypothesis, which also agrees with preliminar NMR experiments. Furthermore, no evidence of reverse Watson Crick pairing is found experimentally, giving support to MD simulations.

8-amino derivatives compounds are able to largely increase the stability of DNA hairpins containing almost the same number of d(A·T) and d(G·C) duplexes, which are designed to have a parallel arrangement. This increase in stability is accurately represented by state-of-the-art MD and MD/TI calculations when a Hoogsteen-type secondary structure is assumed for the hairpins.

Overall, the combination of accurate theoretical calculations and experimental techniques have led us to the development of a new, very promising strategy for the stabilization of parallel-stranded H-type duplexes. These structures can act as templates for the formation of DNA·DNA·DNA and DNA·RNA·DNA triplexes, which might be useful for biotechnological purposes, as well as for antigene and antisense therapies.

## ACKNOWLEDGEMENTS

We thank the Direcció General de Investigació Científica y Tècnica (projects PB98-1222, BQU2000-0649 and PM99-0046), Generalitat de Catalunya (SGR/00018) and CyGene Inc., for financial support. We also thank the Centre de Supercomputació de Catalunya (CESCA) for computational facilities.

## REFERENCES

1. Saenger, W. *Principles of Nucleic Acid Structure*. Springer: New York, 1994.
2. Soyfer, V.N.; Potaman, V.N. *Triple Helical Nucleic Acids*. Springer: New York, 1996.
3. Sun, J.S.; Hélène, C. *Curr. Opin. Struct. Biol.* **1993**, 3, 345.

4. Laughlan, G.; Murchie, A.I.; Norman, M.H.; Moody, P.C.; Lilley, D.M.; Luisi, B. *Science* **1994**, *265*, 520.
5. Williamson, J.R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703.
6. Watson, J.D.; Crick, F.H.C. *Nature* **1953**, *171*, 737.
7. Ng, H.-L.; Kopka, M.L.; Dickerson, R.E. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2035.
8. Vargason, J.M.; Eichman, B.F.; Ho, P.S. *Nature Struct. Biol.* **2000**, *7*, 758.
9. Rippe, K.; Jovin, T.M. *Methods in Enzymology*, **1992**, *211*, 199.
10. van de Sande, J.H.; Ramsing, N.B.; Germann, M.W.; Elhorst, W.; Kalisch, B.W.; Kitzing, E.V.; Pon, R.T.; Clegg, R.C.; Jovin, T.M. *Science* **1988**, *241*, 551.
11. Ramsing, N.B.; Rippe, K.; Jovin, T.M. *Biochemistry* **1989**, *28*, 9528.
12. Germann, M.W.; Vogel, H.J.; Pon, R.T.; van de Sande, J.H. *Biochemistry* **1989**, *28*, 6220.
13. Rippe, K.; Ramsing, N.B.; Jovin, T.M. *Biochemistry* **1989**, *28*, 9536.
14. Rentzeperis, D.; Rippe, K.; Jovin, T.M.; Marky, L.A. *J. Am. Chem. Soc.* **1992**, *114*, 5926.
15. Germann, M.W.; Kalisch, B.W.; Pon, R.T.; van de Sande, J.H. *Biochemistry* **1990**, *29*, 9426.
16. Lavelle, L.; Fresco, J.R. *Nucl. Acids Res.* **1995**, *23*, 2692.
17. Scaria, P.V.; Shafer, R.H. *Biochemistry* **1996**, *35*, 10985.
18. Bhaumik, S.R.; Kandala, V.R.; Govil, G.; Liu, K.; Miles, H.T. *Nucl. Acids Res.* **1995**, *23*, 4116.
19. Liu, K.; Miles, T.; Frazier, J.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 11802.
20. Kan, C.H.; Zhang, R.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126.
21. Sen, D.; Gilbert, D. *Biochemistry* **1992**, *31*, 65.
22. Kremer, E.J.; Pritchard, M.; Linch, M.; Yu, S.; Holman, K.; Baker, E.; Warren, S.T.; Schlessinger, D.; Sutherland, G.R.; Richards, R.I. *Science* **1991**, *252*, 1711.
23. Tchurikov, N.A.; Ebralidze, A.K.; Georgiev, G.P. *EMBO J.* **1986**, *5*, 2341.
24. Tchurikov, N.A.; Chernov, B. K.; Golova, Y.B.; Nechipurenko, Y.D. *FEBS Lett.* **1989**, *257*, 415.
25. Tchurikov, N.A.; Shcholykina, A.K.; Borisova, O.F.; Chernov, B.K. *FEBS Lett.* **1992**, *297*, 233.
26. Veitia, R.; Ottolenghi, C. *J. Theor. Biol.* **2000**, *206*, 317.
27. Kandimalla, E.R.; Agrawal, S.; Venkataraman, G.; Sasisekharan, V. *J. Am. Chem. Soc.* **1995**, *117*, 6416.
28. Kandimalla, E.R.; Agrawal, S. *Biochemistry* **1996**, *35*, 15332.
29. Aviño, A.; Morales, J.C.; Frieden, M.; de la Torre, B.G.; Güimil-García, R.; Cubero, E.; Luque, F.J.; Orozco, M.; Azorín, F.; Eritja, R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1761.
30. Westhof, E.; Sundaralingam, M. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1852.
31. Hakoshima, T.; Fukui, T.; Ikehara, M.; Tomita, K.I. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 7309.

32. Pattabiraman, N. *Biopolymers* **1986**, *25*, 1603.
33. Mohammadi, S.; Klement, R.; Shchyolkina, A.K.; Liquier, J.; Jovin, T.M.; Taillandier, E. *Biochemistry* **1998**, *37*, 16529.
34. Zhou, N.; Germann, M.W.; van de Sande, J.H.; Pattabiraman, N.; Vogel, H.J. *Biochemistry* **1993**, *32*, 646.
35. Yang, X.L.; Sugiyama, H.; Ikeda, S.; Saito, I.; Wang, A.H. *Biophysical J.* **1998**, *75*, 1163.
36. Hashem, G.M.; Wen, J.D.; Do, Q.; Gray, D.M. *Nucleic Acids Res.* **1999**, *27*, 3371.
37. Escudé, C.; Mohammadi, S.; Sun, J.S.; Nguyen, C-H.; Bisagni, E.; Liquier, J.; Taillandier, E.; Garestier, T.; Hélène, C. *Chemistry and Biology* **1996**, *3*, 57.
38. Germann, M.W.; Kalisch, G.T.; van de Sande, J.H. *Biochemistry* **1998**, *37*, 12962.
39. Arnott, S.; Hukins, D.W.L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.
40. Lee, J.S.; Johnson, D.A.; Moogan, A.R. *Nucleic Acids Res.* **1979**, *6*, 3073.
41. Völker, J.; Klump, H.H. *Biochemistry* **1994**, *33*, 13502.
42. Wittung, P.; Nielsen, P.; Norden, B. *Biochemistry* **1997**, *36*, 7973.
43. Chollet, A.; Kawashima, E. *Nucleic Acids Res.* **1998**, *16*, 305.
44. Shields, G.; Laughton, C.A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *117*, 7463.
45. Soliva, R.; Laughton, C.A.; Luque, F.J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226.
46. Shields, G.; Laughton, C.A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895.
47. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. *J. Chem. Phys.* **1995**, *103*, 8577.
48. Darden, T.A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *103*, 8577.
49. Ryckaert, J.P.; Ciccote, G.; Berendsen, J.C. *J. Comput. Phys.* **1977**, *23*, 327.
50. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
51. Cheatham, T.E.; Cieplak, P.; Kollman, P.A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
52. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.; Klein, M.L. *J. Chem. Phys.* **1983**, *79*, 926.
53. Soliva, R.; Güimil-García, R. Blas, J.R.; Eritja, R.; Asensio, J.L.; González, C.; Luque, F.J.; Orozco, M. *Nucleic Acids Res.* **2000**, *28*, 4531.
54. Cubero, E.; Güimil-García, R.; Luque, F.J.; Eritja, R.; Orozco, M. *Nucleic Acid. Res.* **2001**, *29*, 2522
55. Güimil-García, R.; Ferrer, E.; Macías, M.J.; Eritja, R.; Orozco, M. *Nucleic Acids Res.* **1999**, *27*, 1991.
56. Soliva, R.; Luque, F.J.; Orozco, M. *Nucleic Acids Res.* **1999**, *27*, 2248.

57. Gelpí, J.L.; Luque, F.J.; Orozco, M. *CMIP Computer program*; University of Barcelona, Barcelona, 2001.
58. Lavery, R.; Sklena, J. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.
59. Cubero, E.; Laughton, C.A.; Luque, F.J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891.
60. Hernández, B.; Soliva, R.; Luque, F.J.; Orozco, M. *Nucleic Acids Res.* **2000**, *28*, 4873.
61. 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.
62. Plateau, P.; Güeron M. *J. Am. Chem. Soc.* **1982**, *104*, 7310.
63. Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859.
64. Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 667.
65. Radhakrishnan, I.; Patel, D.J. *Biochemistry* **1994**, *33*, 1405.
66. Radhakrishnan, I.; Patel, D.J. *Structure* **1994**, *2*, 395.
67. Wang, E.; Koshlap, K.M.; Gillespie, P.; Dervan, P.B.; Feigon, J. *J. Mol. Biol.* **1996**, *257*, 1052.
68. Asensio, J.L.; Dhesai, J.; Bergquist, S.; Brown, T.; Lane, A.N. *J. Mol. Biol.* **1998**, *275*, 811.
69. Asensio, J.L.; Brown, T.; Lane, A.N. *Structure* **1999**, *7*, 1.
70. Asensio, J.L.; Brown, A.T.; Lane, A.N. *Nucleic Acids Res.* **1998**, *26*, 3677.
71. Gelpí, J.L.; Kalko, S.; de la Cruz, X.; Cirera, J.; Luque, F.J.; Orozco, M. *Proteins in Press* 2001.
72. Güimil-García, R.; Bachi, A.; Eritja, R.; Luque, F.J.; Orozco, M. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3011.
73. When the nucleotides 8-amino 2'-deoxyadenosine, 8-amino 2'-deoxyguanosine, and 8-amino 2'-deoxyinosine appear in an oligonucleotide the standard abbreviations (8AA, 8AG and 8AI) are changed to A<sup>N</sup>, G<sup>N</sup> and I<sup>N</sup>.
74. Kawai, K.; Saito, I.; Sugiyama, H. *Tetrahedron Lett.* **1998**, *39*, 5221.
75. Rao, T.S.; Durland, R.H.; Revankar, G.R. *J. Heterocyclic Chem.* **1994**, *31*, 935.
76. Rieger, R.A.; Iden, C.R.; Gonikberg, E.; Johnson, F. *Nucleosides Nucleotides* **1999**, *18*, 73.
77. Germann, M.W.; Zhou, N.; van de Sande, J.H.; Vogel, H.J. *Methods in Enzymology* **1995**, *261*, 207.

## 4.5 Referencias bibliográficas

- 1) Müller-Dethlefs, K. & Hobza, P. (2000) *Chem. Rev.*, **100**, 143.
- 2) Ma, J.C. & Dougherty, D.A. (1997) *Chem. Rev.*, **97**, 1303.
- 3) Gokel, G.W., De Wall, S.L. & Meadows, E.S. (2000) *Eur. J. Org. Chem.*, 2967.
- 4) McFail-Isom, L., Shui, X. & Williams, L.D. (1998) *Biochemistry*, **37**, 17105.
- 5) Luque, F.J. & Orozco, M. (1998) *J. Comp. Chem.*, **19**, 866.
- 6) Bader, R.F.W. (1990) En *Atoms in Molecules – A Quantum Theory*. Oxford University Press, Oxford.
- 7) Hobza, P. & Havlas, Z. (2000) *Chem. Rev.*, **100**, 4253.
- 8) Crabtree, R.H. (1998) *Science*, **282**, 2000.
- 9) Kim, K.S., Tarakeshwar, P. & Lee, J.Y. (2000) *Chem. Rev.*, **100**, 4145.
- 10) Reed, A.E., Curtiss, L.A. & Weinhold, F. (1988) *Chem. Rev.*, **88**, 899.
- 11) Kool, E.T. (1997) *Chem. Rev.*, **97**, 1473.
- 12) Goodman, M.F. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 10493.
- 13) Kool, E.T, Morales, J.C. & Guckian, K.M. (2000) *Angew. Chem. Int. Ed.*, **39**, 990.
- 14) Loakes, D. (2001) *Nucleic Acids Res.*, **29**, 2437.

- 15) Chollet, A. & Kawashima, E. (1988) *Nucleic Acids Res.*, **16**, 305.
- 16) Todas las simulaciones de dinámica molecular (MD) y de integración termodinámica (TI) se han llevado a cabo con el paquete de programas AMBER.
- 17) Los cálculos cuánticos se han realizado al nivel de teoría del funcional de la densidad (DFT) empleando el funcional híbrido B3LYP.
- 18) Se han realizado cálculos SCRF y QM-MIPp, ambos a nivel *ab initio* HF.
- 19) Petruska, J., Goodman, M.F., Boosalis, M.S., Sowers, L.C. & Cheong, C. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6252.
- 20) Moran, S., Ren, R.X.F. & Kool, E.T. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 10506.
- 21) Moran, S., Ren, R.X.F., Rummey, S. & Kool, E.T. (1997) *J. Am. Chem. Soc.*, **119**, 2056.
- 22) Hobza, P. & Šponer, J. (1999) *Chem. Rev.*, **99**, 3247.
- 23) Mohammadi, S., Klement, R., Shchvolkina, A.K., Liquier, J., Jovin, T.M. & Taillandier, E. (1998) *Biochemistry*, **37**, 16529.
- 24) Lavelle, L. & Fresco, J.R. (1995) *Nucleic Acids Res.*, **23**, 2692.
- 25) Hashem, G.M., Wen, J.D., Do, Q. & Gray, D.M. (1999) *Nucleic Acids Res.*, **27**, 3371.
- 26) Bashford, D. (1997) En *Scientific Computing in Object-Oriented parallel environments*. Springer, Berlin.