Theoretical Studies of d(A:T)-Based Parallel-Stranded DNA Duplexes

Elena Cubero,[†] F. Javier Luque,*[,‡] and Modesto Orozco*[,†]

Contribution from the Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgd. Diagonal s/n, Barcelona 08028, Spain

Received May 15, 2001

Abstract: 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−Boltzmann (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.

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 has been 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 a 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]

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 with 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 indicate 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 in forming rWC H-bond pairs and (ii) the formation of a strong ionic H-bond

* Corresponding authors.
[†] Departament de Bioquímica i Biologia Molecular.
[‡] Departament de Fisicoquímica.

(1) Westhof, E.; Sundaralingam, M. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1852−1856.
(2) Rippe, K.; Jovin, T. M. *Methods Enzymol.* **1992**, *211*, 199−220.
(3) 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. *Nucl. 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 for 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. *Chem. Biol.* **1996**, *3*, 57−65.
(29) Germann, M. W.; Kalisch, G. T.; van de Sande, J. H. *Biochemistry* **1998**, *37*, 12962−12970.

10.1021/ja011200t CCC: $20.00
Published on Web 11/06/2001

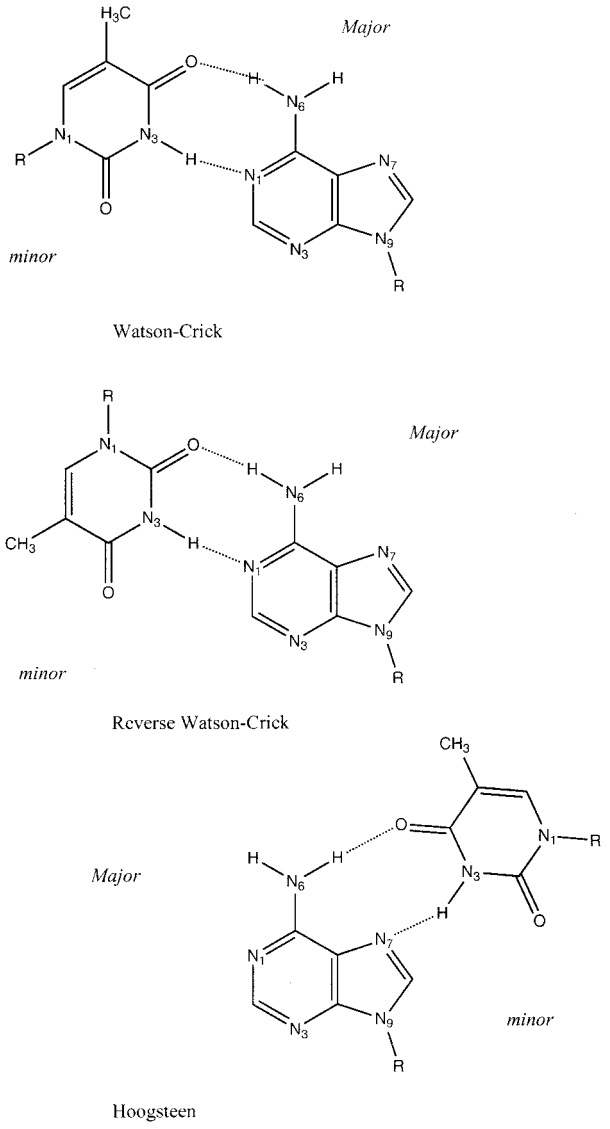

Figure 1. Schematic representation of the Watson−Crick (WC), reverse Watson−Crick (rWC), and Hoogsteen (H) pairings A:T. The nomenclature of grooves for nonstandard DNAs is based on homology with WC duplexes (rWC) and on triplex DNA (H).

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 that the WC and rWC ones.[30−32]

In this paper we present the first systematic theoretical study of 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 by using molecular dynamics (MD) simulations. 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

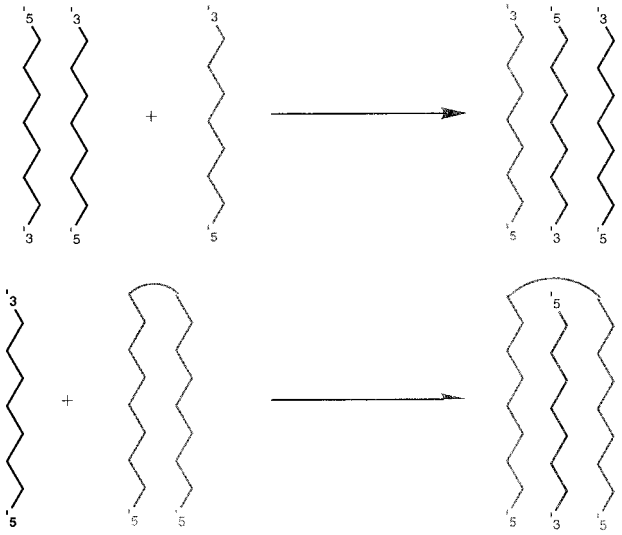

Figure 2. Strategies for the formation of triplexes: (top) single-stranded DNA as a triplex-forming oligonucleotide and (bottom) parallel-stranded duplex (hairpin) as a triplex-forming oligonucleotide.

5-mer and 15-mer duplexes), which were built up by 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 by 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 by 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−Boltzmann 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^3. The solvated systems were then optimized, heated (298 K), and equilibrated for 130 ps by using our standard multistage 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). To verify the convergence in the results trajectories of selected duplexes were extended up to 3 ns. Long-range effects were introduced by using periodic boundary conditions and the particle Mesh Ewald technique (PME[38,39]). PME calculations were performed by using a grid spacing around 1 Å, a 4th order spline, and a tolerance of 5×10^{-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

(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. Comput. 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. *Proteins* **2001**, in press.
(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.

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^+) and the DNA in thousands of grid points around the DNA. The interaction energy is computed by using a classical Lenard-Jones term, and a solvent-screened molecular electrostatic potential was obtained by solving the Poisson−Boltzmann 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 by 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−Boltzmann eq 45, using the 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 by 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^2) following Honig and co-workers.[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 by 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 the 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 with Curves.[50] For all the analyses 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)$_n$ duplexes in the WC, rWC, and H helical models are stable, as noted in the root-mean-square deviations (RMSd) with respect to the average structure (RMSd$_{av}$) for each simulation, which are clearly below 2 Å for all the simulations (Table 1). As expected for a polymer, the RMSd$_{av}$ increases with the size (around 0.08 Å/base pair for the WC duplex and 0.06 Å/base pair for the rWC and H helices). Comparison of RMSd$_{av}$ 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).

Table 1. Root-Mean-Square Deviation (Å) between the Trajectories of the Nine 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)[a]

no. of residues	antiparallel Watson−Crick	parallel reverse Watson−Crick	parallel Hoogsteen
5	0.7(0.2)	0.8(0.2)	0.6(0.1)
	1.1(0.1)	*1.4(0.3)*	*0.7(0.1)*
6		0.8(0.2)	0.7(0.1)
		1.3(0.2)	*0.8(0.2)*
7	0.9(0.2)	0.8(0.2)	0.8(0.1)
	1.5(0.2)	*1.4(0.2)*	*0.8(0.2)*
8		1.1(0.2)	0.8(0.2)
		1.6(0.4)	*0.9(0.1)*
9	1.0(0.2)	1.2(0.3)	0.8(0.2)
	2.1(0.3)	*1.8(0.3)*	*0.9(0.1)*
10		1.1(0.2)	0.9(0.2)
		1.7(0.2)	*1.0(0.2)*
11	1.0(0.2)	1.2(0.4)	1.0(0.3)
	2.2(0.3)	*2.1(0.4)*	*1.2(0.3)*
12		1.3(0.3)	1.1(0.3)
		2.3(0.4)	*1.4(0.3)*
15	1.5(0.4)	1.4(0.3)	1.2(0.3)
	2.8(0.6)	*2.9(0.5)*	*1.4(0.2)*

[a] 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 (RMSd$_{can}$) for each duplex. All the RMSd$_{can}$ 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 RMSd$_{can}$ values for the H-duplex (where the canonical models were derived directly from d(A·T-T)$_n$ triplexes[33]) are 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(A: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 that is common for MD simulations in pure solvent, but is not found when the crystal environment is considered.

Helical parameters of the three duplex families do not show large differences, as can be noted from inspection of Table 2.

(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. Strut. Dyn.* **1988**, *6*, 63−91.

Table 2. Selected MD-Averaged Helical Parameters for the $d(A:T)_n$ Sequence ($n = 11$ in Roman, $n = 15$ in Italics) in the WC, rWC, and H Conformations (Standard Deviations in Parentheses)[a]

parameter	antiparallel Watson−Crick	parallel reverse Watson−Crick	parallel Hoogsteen
twist	33.8 ± 1.0	35.1 ± 0.8	32.6 ± 2.7
	33.9 ± 1.0	*34.4 ± 1.2*	*32.5 ± 1.6*
rise	3.4 ± 0.1	3.3 ± 0.1	3.4 ± 0.1
	3.4 ± 0.1	*3.4 ± 0.1*	*3.4 ± 0.1*
roll	1.4 ± 1.7	0.7 ± 2.2	−4.1 ± 1.4
	1.9 ± 1.6	*−0.5 ± 1.5*	*−4.2 ± 1.0*
X-disp	−1.2 ± 0.4	0.7 ± 0.5	−0.2 ± 0.4
	−1.3 ± 0.4	*0.9 ± 0.4*	*−0.1 ± 0.4*
phase angle	138 ± 38	148 ± 37	109 ± 45
	138 ± 31	152 ± 41	108 ± 51
minor groove width	5.3 ± 0.5	10.3 ± 0.8	3.9 ± 0.5
	5.7 ± 0.6	*10.6 ± 0.8*	*4.0 ± 0.4*
major groove width	13.7 ± 1.2	7.4 ± 0.5	19.8 ± 1.1
	14.1 ± 1.0	*7.7 ± 0.4*	*19.2 ± 1.2*

[a] The width of the grooves is noted as P−P distances minus 5.8 Å. The rest of the helical parameters were determined with 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 deg, the rest in Å.

The rise is around 3.4 Å, and X-disp and roll are small in all cases. The twist for the H duplex is smaller than that for the WC or rWC ones, which suggests that the small twist found in DNA triplexes (around 29° 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 clearly a displacement toward the East region for the H duplex, and toward 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 and, 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 exhibits 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 especially 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).

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 are 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 kcal mol^{-1} K^{-1} (15-mer duplex), which compare with entropy values of 2.181 kcal mol^{-1} K^{-1} for the H-duplex and 2.145 kcal mol^{-1} K^{-1} for the rWC one.[52] It is worth noting that the larger flexibility of the WC duplex detected in entropy calculations performed by using all the modes[52−54] is also detected in the analysis of the lowest frequencies, which are 5−9 cm^{-1} smaller than the corresponding values in the parallel stranded duplexes (see Table 3).

In summary, the essential dynamics of the three helices are similar, which suggests that the general helical structure, rather than the specific sequence pattern, determines the major conformational movements of the DNA. However, it is clear that the three helices have different flexibility, the WC anti-parallel 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 $d(A:T)_n$ sequence. The stability of a helical structure can be determined as shown in eq 1 where E

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

(51) Soliva, R.; Luque, F. J.; Alhambra, C.; Orozco, M. *J. Biomol. Strut. Dyn.* **1999**, *6*, 89−99.

(52) According to our experience (ref 55), entropy analyses performed with 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.

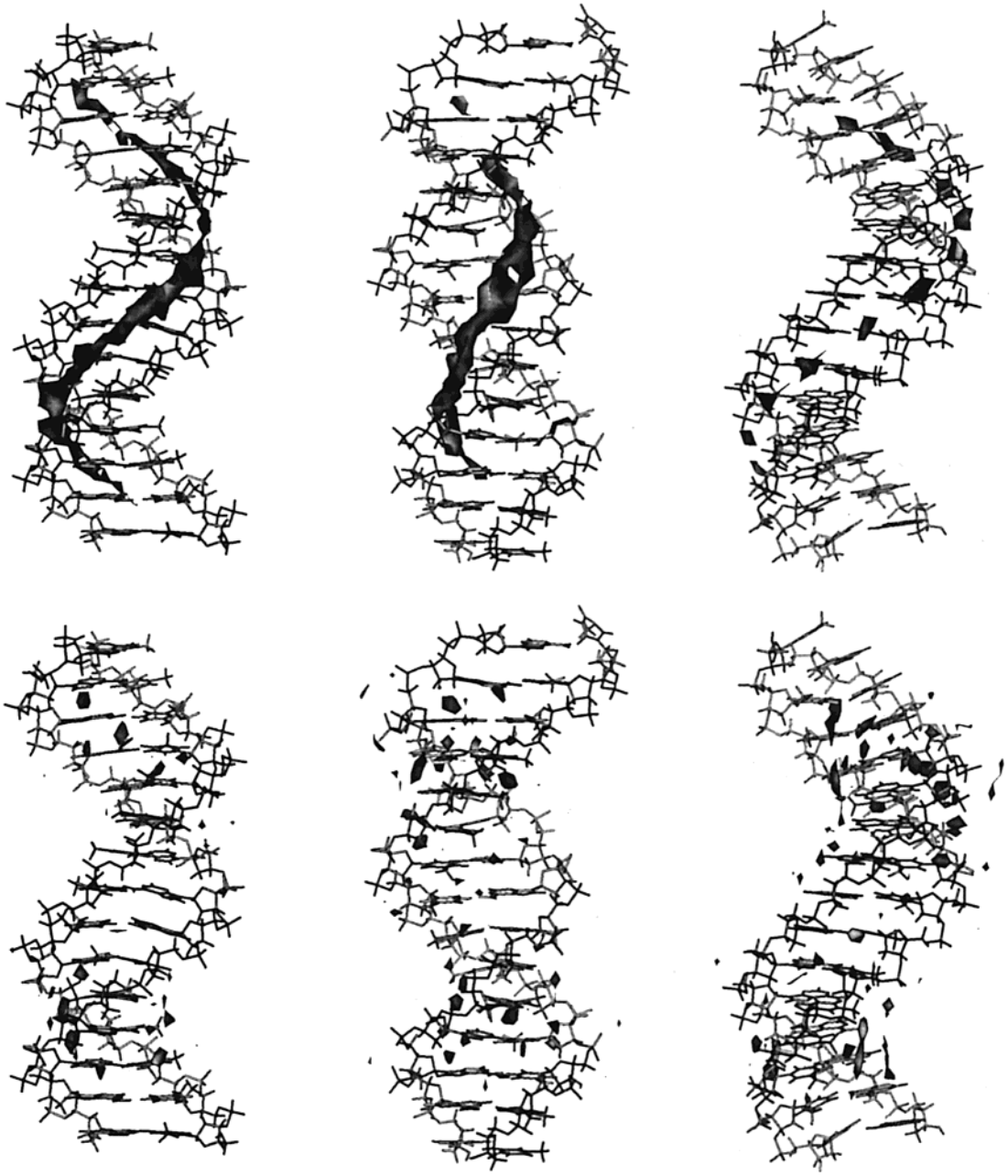

Figure 3. Top: Classical molecular interaction potential (cMIP) of the three helical models when the probe molecule is O^+ (contour level −5.0 kcal/mol). Bottom: Solvation maps for the three helical models (contour level correspond to a density of 2 g/mL, i.e., to a preferential solvation of −0.4 kcal/mol). See text for details.

is the intramolecular energy (as computed by AMBER force field) and G_{sol} is the solvation (free) energy obtained with use of the PB/SA technique (see Methods), and the MD trajectories were obtained considering explicit solvent. If the entropy corrections are ignored (ref 52; for a qualitative discussion on entropy contributions see above) the free energy difference between two helical structures can be approximated by using eq 2.

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

The use of eq 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

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[a]

helix	frequencies (cm^{-1})	% variance explained	mode description
	19	39	global bending
antiparallel	26	20	global bending
	39	9	twisting
	29	32	twisting and bending
rWC parallel	35	20	global bending
	44	10	twisting
	25	25	global bending
H parallel	32	17	twisting and bending
	46	11	twisting

[a] See text for details.

Table 4. Total Energy for the Three Families of Helical Structures Considered in This Study with Standard Errors in the Averages Displayed in Parentheses[a]

no. 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)	−2314(1.8)	−2297(1.9)
	−2321(1.2)[b]	*−2315(1.6)*	*−2298(1.1)*
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)

[a] All the values are in kcal/mol. [b] Values in italics correspond to averages obtained during the last 2 ns of unrestrained trajectories of 3 ns.

Table 5. Solvation Free Energy for the Three Families of Helical Structures Considered in This Study with Standard Errors in the Averages Displayed in Parentheses[a]

no. 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)	−3248(1.6)	−3285(1.7)
	−3206(1.2)[b]	*−3222(1.5)*	*−3282(1.0)*
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)

[a] All the values are in kcal/mol. [b] Values in italics correspond to averages obtained during the last 2 ns of unrestrained trajectories of 3 ns.

Table 6. Internal Energy for the Three Families of Helical Structures Considered in This Study with Standard Errors in the Averages Displayed in Parentheses[a,b]

no. 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)
	885(0.4)	*907(0.5)*	*984(0.4)*
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)

[a] All the values are in kcal/mol. [b] Values in italics correspond to averages obtained during the last 2 ns of unrestrained trajectories of 3 ns.

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 increase 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.

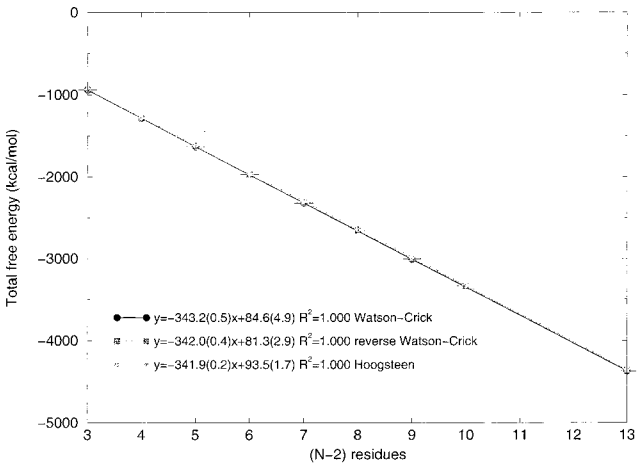

Figure 4. Dependence of the total free energy on the length of the duplexes. The regression equations (standard errors in the fitted parameters in parentheses) and the determination coefficient are displayed. The error bars in the figure correspond to the standard errors in the average energy estimates.

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 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 show 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.

The preceding results show a qualitatively correct picture of the stability of the $d(A:T)_n$ 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

(55) Harris, S.; Gavathiotis, E.; Seearle, 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.

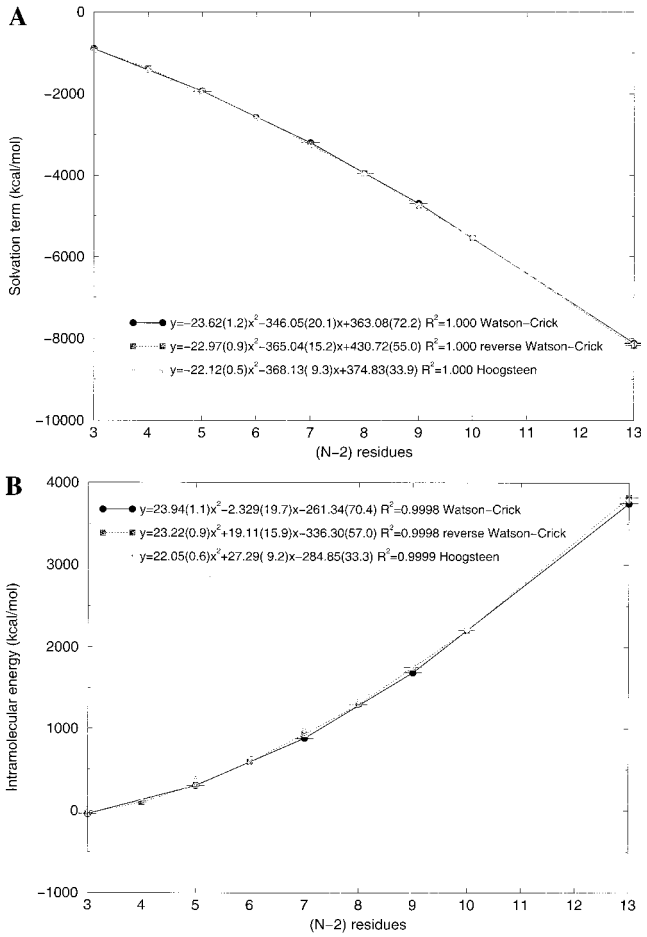

Figure 5. Dependence of the solvation (A) and intramolecular (B) free energy components on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parentheses) and the determination coefficient are displayed. The error bars in the figure correspond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

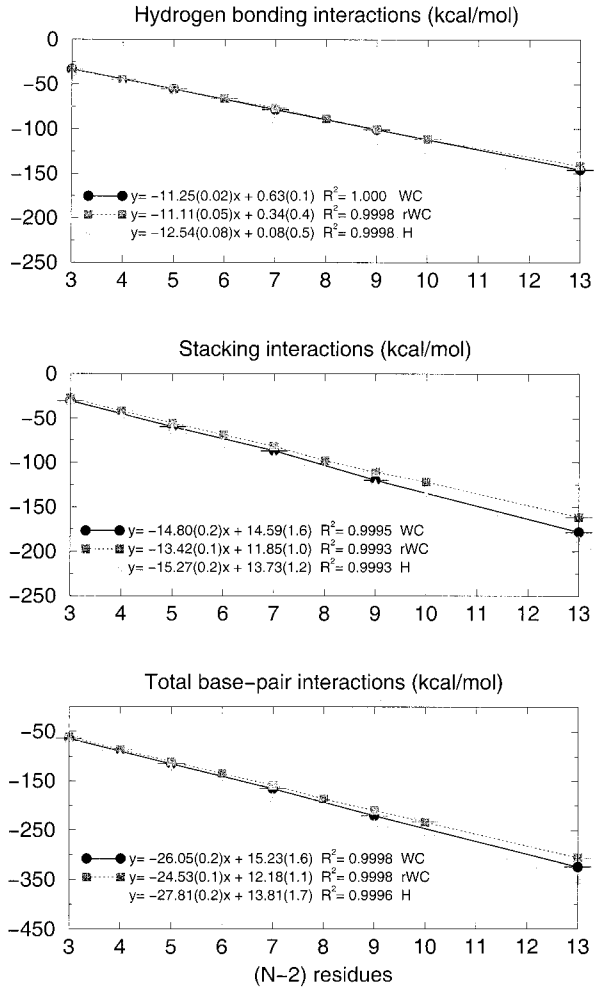

Figure 6. Dependence of the hydrogen-bonding, stacking, and total base-pair interactions on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parentheses) and the determination coefficient are displayed. The error bars in the figure correspond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

(see above) to be magnified if entropic considerations are included). It is worth noting that a portion of our results is 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 and 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 Figure 5A,B).

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 vice versa. 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 on 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 show 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 are 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 bonds, and suggest 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.

Stacking interactions are of similar intensity as H-bonds (Figure 6), which points out the importance of stacking interactions in the stabilization of duplex structures.[32] The growth of the stabilization stacking energy with the length of the helices is also linear ($r^2 > 0.999$ in all cases), demonstrating the short-range nature of stacking interactions. The H helix also has the best stacking interactions, followed by the antiparallel WC duplex, while the rWC helix displays the worst stacking (Figure 6). As expected, the dependence of the total base-pair energy (H-bond + stacking) with the length of the duplex is also linear, and suggests a surprising order of stability: H > WC > rWC (see Figure 6). It is worth noting that this ordering is opposite to that predicted from the total free energy (see above), and to that found experimentally, demonstrating that the stability of a helix is not determined only by the stability of the relative arrangement of the nucleobases.

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 repulsions 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 the equations in Figure 7 out of the fitted range, the situation might however change for very large oligonucleotides.

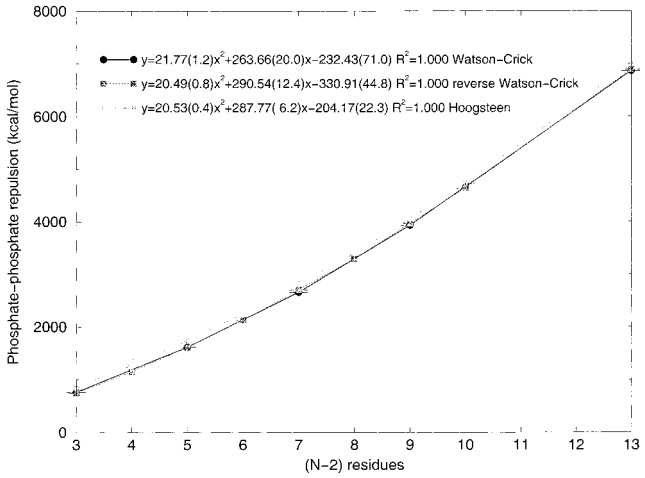

Figure 7. Dependence of the phosphate−phosphate interactions on the length of the different duplexes. The regression equations (standard errors in the fitted parameters in parentheses) and the determination coefficient are displayed. The error bars in the figure correspond to the standard errors in the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

Conclusions

(1) Double helices of DNAs based on the poly d(A:T) motif are intrinsically stable in dilute aqueous solution in both 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 helix, 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.

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

JA011200T